Scientometrics (2022) 127:3695–3718
https://doi.org/10.1007/s11192-022-04406-6

Smart bibliometrics: an integrated method of science mapping and bibliometric analysis

Vilker Zucolotto Pessin[1] · **Luciana Harue Yamane**[1] · **Renato Ribeiro Siman**[1]

Received: 4 February 2022 / Accepted: 5 May 2022 / Published online: 21 May 2022

Abstract
Bibliometric techniques and science mapping are widely employed in the research environment to provide an overview of the state-of-the-art of scientific knowledge on a given topic. These techniques are essential to assist the researcher's work by guiding the compilation of the bibliography to support the theory discussion. To this objective, the Smart Bibliometrics was developed to facilitate bibliometric analysis and selection of theoretical references, embodied by a system that brings intelligence, dynamism, and agility to the scientific writing process. The innovation of this methodology is the fusion of two relevant criteria applied during the bibliometric analysis process: the application of a representative metric of classification of scientific papers and dynamic visuals strategically developed. The methodology differs for providing the user with dynamic navigation and interaction experience with the data collected, innovating the approach to reaching insights within the universe of discussions of the scientific community. In addition, as an innovation factor, the method is presented in a scalable Business Intelligence (BI) system that features blunt visuals, extensive analysis repertoire, intuitive navigation, and automated updating. The development was carried out in a cutting-edge technological platform to attend information and sharing intents by employing cloud computing resources, another feature that enables interaction among researcher groups also from different institutions. Additionally, it is not necessary to install any software. The output will be available for consultation, at any time and place, just by using one device with an internet connection.

Keywords Science mapping · Bibliometric analysis · Business intelligence · Making decision · Smart bibliometrics

✉ Renato Ribeiro Siman
renato.siman@ufes.br

Vilker Zucolotto Pessin
vilker.pessin@edu.ufes.br

Luciana Harue Yamane
luciana.yamane@ufes.br

[1] Department of Environmental Engineering, Federal University of Espírito Santo, Av. Fernando Ferrari, Vitória, Espírito Santo 51429075-910, Brazil

Introduction

Science mapping, as the name suggests, seeks to achieve an overview of the "state-of-the-art" scientific knowledge in each research area. Additionally, bibliometrics is the application of mathematical and statistical indicators to measure and compare the evolution of science and technique in any research area. The bibliometric analysis embraces the performance analysis of contributions on specific subjects, complemented by an interaction analysis between the researchers and their subject of study (Donthu et al., 2021), in other words, it covers the analysis of numerical indicators and interactions detected in the body of the scientific data.

The techniques of science mapping have evolved considerably in recent years thanks to advances in information technologies tools. The first discussions on mapping scientific literature indicated the importance of applying software that could perform bibliometric analyses (Ahlgren & Jarneving, 2008; Chavalarias & Cointet, 2008; Noyons et al., 1999; Small, 1997; Soós, 2011; Soós & Kampis, 2012) combining metrics and data visualization.

The use of science mapping and bibliometric techniques have been consolidated as regular practices at the very beginning of the research when several bibliometric information are raised to assist scientific writing (Aria & Cuccurullo, 2017; Murgado-Armenteros et al., 2015; Pagani et al., 2015; Pallottino et al., 2018; Rodríguez-Bolívar et al., 2018). It is a method adopted to make a scientific drawing of the subjects that have been addressed with complex discussions, which requires management since the techniques of science mapping and bibliometrics are composed of systematic steps based on different software. It is worth mentioning that some tools don't provide free access (Aria & Cuccurullo, 2017).

Bibliometric analysis is a method applied to explore and analyze volumes of data, searching for evolutionary nuances of a specific field of knowledge, while shedding light on emerging areas (Donthu et al., 2021). The application of these bibliometric techniques seeks to raise insights about the evolution of science, directing the efforts of researchers to a systematic review of the literature focusing on discussions of the academic community about a specific point based on the organization and summary of theoretical constructions as per the main scientific documents produced.

A methodology initially developed for the selection of bibliography, known as Pro-Know-C (Knowledge Development Process—Constructivist) (Ensslin et al., 2015), aims to build a bibliometric analysis in four different stages: selection of a portfolio of relevant publications, descriptive bibliometric analysis such as the analysis of numerical metrics, systematic review of publications according to the selected portfolio, definition of new researching questions according to the previous steps applied. The development of the ProKnow-C method started in the early 2000s based on the activities of the Laboratory of Multi-criteria Methodologies for Decision Support (LabMCDA) at the Federal University of Santa Catarina, Brazil, to develop a structured process for the selection of bibliographic references scientifically recognized and its analysis on a specific subject (Ensslin et al., 2015). In managing the data, spreadsheets support receiving consecutive filters to select a set of scientific outputs related to distinct research subjects.

Procknow-C, Pagani et al. (2015) proposed a method of articles classification, known as "Methodi Ordinatio", which is expressed by a metric used to rank and classify scientific outputs. It means that in addition to the filters applied systematically to scientific outputs, this methodology develops a metric that allows the classification of results based on the variables number of citations, year of publication, and influence of journals. In the Ordinatio method, it is necessary to manage data extracted from scientific

portals to generate the In Ordinatio classification index. In this case, the user needs to know spreadsheets to develop the calculations suggested by the methodology and apply a mathematical equation of classification that allows selecting the most relevant scientific outputs of the selected sample.

However, with the addition of a representative metric to generate a classification of importance among sample elements extracted from scientific portals, it is essential to develop visuals to expand the capacity of bibliometric analysis. Neuroscience proved the importance of data visualization and the comprehension of certain phenomena since images increase comprehension abilities (Dwyer et al., 2020), justifying the importance of visual impressions in cognitive processes. Thus, Rodríguez-Bolívar et al. (2018) proposed a method of scientific mapping to analyze the evolution of specific research subjects, combining different bibliometric tools able to identify subject fields and show their progress by employing different visualization tools in the research planning.

Before starting the research, it is required to choose a data analysis method to manage workflows that employ software to organize the data extracted from scientific portals that provide visual information, such as the method known as Bibliometrix (Aria & Cuccurullo, 2017). In this methodology, the data extracted from scientific portals are processed in the R software to feed the bibliometric system and generate information and visualizations. The process requires downloading and installing the R software and users' ability to interact with a not-so-intuitive programming language, which considerably hinders researchers' work. The execution algorithm elaborated in the R language, besides being complex, can present some problems during the process, being necessary configuring the environment and thorough execution of commands. Any script error may result in processing problems, and the steps are complex execution.

By considering the dynamic research environment and the need to raise more responsive bibliometric information, the Business Intelligence (BI) tools are widely applicable and powerful to assist researchers in guiding better bibliography choices. The evolution of the information systems grows exponentially, and the BI tools follow this pattern by offering technological solutions in line with the concept of "big data" that lead to positive decisions based on data, information, knowledge, and intelligence (Shollo & Galliers, 2016).

BI embraces the concept of business intelligence oriented to the processes of collecting, organizing, analyzing, and monitoring information, decision-making elements in any business centered on data and knowledge analyses(Božič & Dimovski, 2019; Chen et al., 2018). According to López-Robles et al. (2019), the definition of BI could be comprehended as the collection, analysis, interpretation, and dissemination of high-value information about strategic areas, transmitted to decision-makers at the proper time. Making choices based on information and knowledge is a prerequisite for the success of any enterprise, whether in the business or in the scientific environment where it is possible to use, from the beginning of the research, appropriate references aligned with a relevant gap.

Based on the methods of the study, a BI system was built on the method called Smart Bibliometrics to group the methodologies applied (Aria & Cuccurullo, 2017; Ensslin et al., 2015; Pagani et al., 2015) in a system that associates an assertive classification metric of scientific documents with strategically developed visualizations. This solution combines science mapping and bibliometric techniques, covering automated data manipulation and elaboration of advanced visual analyses to drive the selection of superior references through the Smart Bibliometrics. This new methodology aims to facilitate the work by automating various routines, not requiring the user to understand spreadsheets management, R programming, or even downloading any software, besides being freely accessible.

Essentially, the proposed method is convenient for offering a free access solution and cloud processing capabilities, joining representative metrics of strategic theoretical and visual reference selection, intuitive and user-friendly interface, allowing the experience of data searching and interaction, which offers an overview of scientific research on a specific theme. The Smart Bibliometrics application aims to assist the automation of diverse routine data collection by building a simple and user-friendly system capable of indicating any possible research gaps in the scientific area, as well as to reverse intelligence in scientific writing focused on the evolution of knowledge in the most varied fields of expertise.

Evolution of bibliometric analysis methods and technologies

In 1934, Paul Otlet was the first author to define “bibliométrie” as a tool to measure all aspects related to book publication and reading, establishing the foundation for the measurement analysis of publications (Otlet, 1934). In 1960, the advent of information systems and the construction of databases made it possible to consolidate bibliometrics as an academic practice. Eugene Garfield is the pioneer in indexing citations in sciences and literature of academic journals, producing the first Science Citation Index in 1964 (Clarivate, 2017).

In 1986, concerned about applying a method that would allow a relative comparison between publications indicators, Vinkler was one of the first researchers who struggled to value and categorize scientific productions based on combined numerical metrics (Vinkler, 1986). For this purpose, he applied bibliometric methods, and without his studies, he applied the variables impact factor and number of citations of articles to measure research publications (Vinkler, 1986). Based on these constructions, in 1993, Archie Cochrane created a model of organization that aimed at systematic literature reviews of topics related to the health area, when there was not yet a systematic elimination of non-relevant publications (Pagani et al., 2015).

HistCite (2003)

Eugene Garfield was the founder of the Institute for Scientific Information (ISI) in 1993 and launched in 2003 an open-source system known as HistCite to conduct bibliometric surveys and publication visualizations using the ISI Web of Science (Garfield, 2009) as a database. According to Garfield (2009), the software generates chronological information through bibliography maps that result in research by subject, author, and journal, highlighting the most cited scientific productions.

In the course of scientometrics history, Garfield’s work is a turning point with the creation of the first index of literature citations, the Science Citation Index, which in the future would become the "Web of Science" (Garfield, 2004). In his early work, it is possible to recognize the effects of merging bibliometrics with data visualization analyses by combining variables such as title, authors, dates, addresses, among others, to measure studies in the most varied areas of research and highlight the importance of these scientific databases.

CiteSpace (2004)

CiteSpace is an open-sourced system that provides diverse geospatial analyses for an improved understanding and interpretation of relationship patterns between scientific productions, simplifying the process of locating clusters and potentializing the indication of highly cited papers (Chen, 2005). CiteSpace was developed in Java language, it has an updated CiteSpaceII version, and it performs analysis of hybrid co-citation networks, allowing to raise points on emerging topics (Chen, 2013).

BibExcel (2009)

Bibexcel is a bibliometric tool developed by Olle Persson, which performs bibliometric analysis of easy interaction with different software, displaying data in tabular format (Persson et al., 2009). Users may experience more flexibility in managing both data and information, as in the analysis and manipulation, since using other data sources with the Web of Science, even when dealing with data besides bibliographic records. This system generates data files that can be manipulated in Excel or any program that accepts tabulated data processed in other applications or downloads. Its latest version is available for download as a courtesy of its creator (Persson, 2017).

Proknow-C (2010)

This study raised the matter of selecting articles and choosing a bibliography by comparing sample elements extracted from scientific portals. The Proknow-C methodology is founded on four main steps: bibliographic portfolio selection, bibliometric analysis, systemic analysis, and finally, the definition of research and objectives (Ensslin et al., 2015). From the enforcement phases, it was expected that the scientist could select theoretical references by systematically applying filters in scientific data and using the variable citation number as the key metric. The Proknow-C methodology was developed in the early 2000s and implemented by the implementation of spreadsheets for data processing.

VOSviewer (2010), Gephi (2010), SciMat (2011)

Data visualization begins to be applied to enhance bibliometric analyses. Software for visualization, usually with open and free source, comes up with researchers' discretion to choose which best software could meet their search needs. Although the software for visualization emerges, the solutions do not allow the selection and classification of scientific documents having the construction of conceptual maps as a major advantage. The VOSviewer, for example, is a computer program freely available to build and visualize bibliometric maps (van Eck & Waltman, 2010). Each software has its advantages and disadvantages, for example, Pajek and UCINET. Although both have many features, their development speed is slower compared to software such as Gephi and R (Donthu et al., 2021). SciMat tool is an open-sourced software in Java developed to perform

science mapping analysis within a longitudinal framework with the strong ability to create visualizations that express relationships (Cobo et al., 2013).

Methodi Ordinatio (2015)

Literature shows that the Methodi Ordinatio emerged from constructions of Proknow-C (De Carvalho et al., 2020), following systematic steps for selecting a bibliography and standing out by creating a classification index. This strategy developed a methodology of selection, collection, classification, and systematic reading of scientific articles published in journals. Consequently, three classification criteria would generate the InOrdinatio classification index based on the variables year of publication, the number of citations, and impact factor, according to Eq. 1 (Pagani et al., 2015):

$$\text{InOrdinatio} = \left(\frac{IF}{1000}\right) + \alpha * [10 - (\text{ResearchYear} - \text{PublishYear})] + \left(\sum CI\right) \quad (1)$$

Being α is a weighting factor IF represents the Impact Factor of the journals; ResearchYear is the reference year; PublishYear is the year when the article publication; and $\sum$Ci is the sum of citations of the scientific document (Pagani et al., 2015). A spreadsheet would classify the results by ordering the elements of the InOrdinatio index.

Bibliometrix (2017)

An R language algorithm consolidates the data exported from scientific bases to generate a file fit for uploading on a web platform for processing the data into information. This methodology outperforms due to the development of visual maps and multiple analyses. Also, it presents an overview of scientific production worldwide, which employs static graphics and inhibits user interaction. The execution of this method is complex and requires systematized steps. It works with the databases Scopus, Dimensions, Pubmed, Cochrane, and Web of Science, however, consolidating the data requires users' ability to manage data in spreadsheets and knowledge in R. In addition, this system does not allow the automatized selection and classification of articles from a classification metric, having been developed especially for science mapping analysis. The previews should be uploaded, but the system updating may be time-consuming, which would require the user's knowledge to execute commands and download software. The graphical visualization is possible through a web interface, "Biblioshiny" (Massimo & Cuccurullo, 2021).

Smart bibliometrics (2022)

This methodology embodies the system named "Smart Bibliometrics" based on cloud computing and artificial intelligence resources developed in a technological platform capable to meet the needs of the 'big data' age (Chen et al., 2018). Its link is freely accessible, unnecessary to install any other software or have any previous knowledge in data management. It can be customized during its implementation in educational institutions that plan to gather researchers' teams, distributing the solution on an intranet system with restricted access. In addition, it preserves users' updating history allowing the scheduling of incremental

updates. It allows process automation through a system that combines a delimited metric for scientific documents classification with strategically developed visuals that facilitate the decision-making process from scientists' perspectives. It also provides a whole set of analyses, focusing on the relationship between specialized journals, research subjects, and authors. This system was based on Business Intelligence, aiming to reverse intelligence in decision making. It enables users' interactions while searching reports providing different visuals to generate multiple insights.

Timeline

The following timeline (Fig. 1) summarizes the evolution of methodologies and technologies that have been improved continuously to support decision-making patterns for selecting a bibliography.

Smart bibliometrics breakdown

The Smart Bibliometrics aims to add a metric with a significant classification capacity of scientific documents while presenting strategically designed visualizations to support logical choices, due to selecting a high standard bibliographic portfolio. The question raised here is: how can researchers promptly and cleverly select a finer automatized bibliographic portfolio that enables creating new research for scientific advances? That said, the Smart Bibliometrics emerged targeting the automatization of the entire process, avoiding researchers to spend essential hours of study with manual program installation, management of spreadsheets, downloading, and software installation. The methodology designation—Smart—comes from the fact that this solution is supported by a business intelligence system that creates data-based classification metrics through ranking scientific works

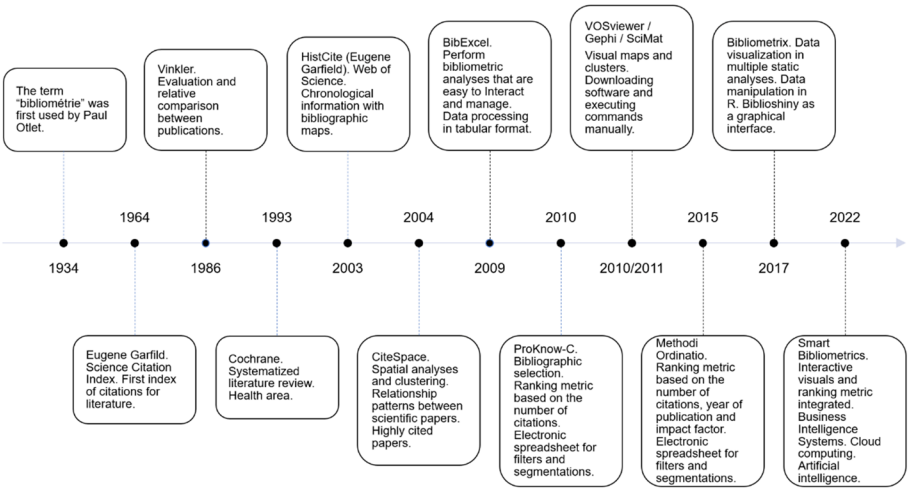

Fig. 1 Methodologies and technologies evolution for selected bibliography

according to numerical calculations and metrics via dynamic visualizations, which elicit insights and justify the origin of the term Smart in this research context.

On the classification metric, the Smart Bibliometrics relies on a mathematical equation that relates the multiple numbers of citations, year of publication, and impact factor of journals, as proposed by the "Methodi Ordinatio" (Pagani et al., 2015). The "Methodi Ordinatio" resulted in the index "InOrdinatio" for the classification of scientific productions; and the calculation approach considers citations substantially for the selected sample space, a fact that can distort analysis by favoring older publications depending on the weighting factor applied by the user.

In this new methodology, the Smart index reduces the absolute value of citations to a rate of citations per year, removing from the indicator the tendency to overvalue old scientific productions that by nature are more likely to be cited than recent publications. Longstanding publications are available for consultation for longer periods, being the reason for correcting the issue by using a new rating, which would allow any publication to the same comparison stage. By converting the number of citations to a variable rate per year, the comparison between the elements becomes more equitable, and the generated index expands its explanatory power. In addition, more recent publications bring information such as trends, gaps, and current research challenges.

Another relevant numerical variable is the impact factor of journals. This is a bibliometric metric published by the Journal Citation Reports (JCR), under the management of Clarivate Analytics, responsible for the "Web of Science" database and for keeping the impact factor (IF) of the journals updated and available for consultation. This index was considered a calculation component by the Smart Bibliometrics classification index. JCR provides quantitative tools to classify, evaluate, categorize, and compare journals using the impact factor which is a measure of the frequency that an 'average article' of a journal is cited in a given year or period (Clarivate, 2022).

To determine the Smart index, Eq. 2 calculates the rate of citations per year. Next, the index adds the impact factor of the journals, exemplified in Eq. 3.

$$\mathrm{Tx} = \frac{\sum \mathrm{Ci}}{\left[(\mathrm{Cy} - \mathrm{Yp}) + 1\right]} \tag{2}$$

$$\text{Smart Index} = \left[\left(\frac{\mathrm{Tx}}{[\mathrm{MaxTx}]}\right) * w1\right] + \left[\left(\frac{\mathrm{IF}}{[\mathrm{MaxIF}]}\right) * w2\right] \tag{3}$$

In which: ΣCi: number of times the document was cited, Cy: current year, Yp: year of publication, IF: impact factor of the publication in the Journal Citation Reports (JCR), MaxTx: maximum value for the citation rate per year of the selected sample, MaxFI: maximum value for the Impact Factor of the selected sample, $w1$: weighting for citation rate, $w2$: weighting for Impact Factor.

Due to the occurrence of extreme values, an average rate is calculated based on the distribution of 100 points to balance the equation and prevent the two components from distorting the analysis, weighting the distribution in 50% and 50% for the citation rate per year and impact factor criteria, respectively. This is a pre-established distribution as a default for the system, however, the user can change the balance of the equation by applying different weights.

In the visuals which papers ranking measure is displayed, user may interact and change weighting criteria that best match their research objectives. If the user wants to consider

the citation rate as most important, it is enough to select a higher value that will increase "$w1$" weighting. Calculations will automatically be adjusted to the new context of analysis.

In addition to a representative metric for the classification of articles, strategically developed visualizations complement the analyses for researchers to make assertive decisions. Therefore, the Smart Bibliometrics accurately relates authors and journals focusing on the research problem investigation, offering interactivity by navigating through the functions developed enabling the user to generate multiple insights and have accurate choices. Thus, considering the "Methodi Ordinatio" and the "Bibliometrix" (Aria & Cuccurullo, 2017; Massimo & Cuccurullo, 2021; Maximo & Corrado, 2021; Pagani et al., 2015), this new system considerably expands analytics by providing dynamic visuals built for interactive navigation on a cloud computing platform with processing capability to meet the requirements of the 'big data' scenario.

The flow described in Fig. 2 summarizes the steps that the researcher must take to select the bibliography.

Detailing the smart bibliometrics in stages

The methodologies of bibliometric analysis usually permeate the analysis of data of scientific portals, which requires a data download and preprocessing to generate the information (De Carvalho et al., 2020; Dervis, 2019; Donthu et al., 2021; Pagani et al., 2015). The following steps summarize the main common steps in many methodologies, with the main difference in the development of process automation, cloud computing,

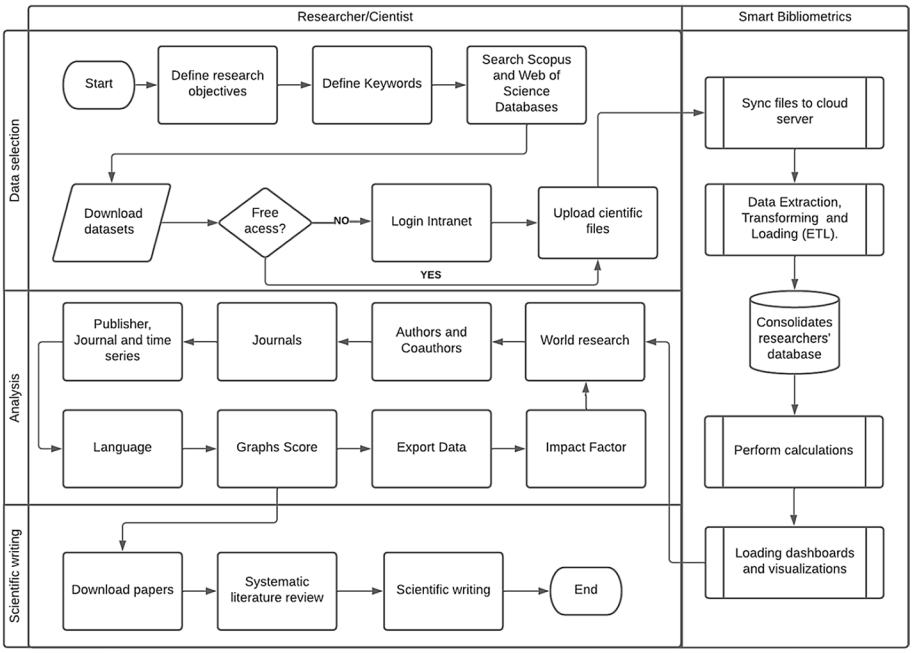

Fig. 2 Bibliometric analysis workflow

and technological resources in line with critical factors of the big data era that are the collection, analysis, and dissemination of information focused on the decision-making (López-Robles et al., 2019; Shollo & Galliers, 2016).

Step 1—Definition of the research objectives. The process begins with the definition of the research objectives, that is, what is intended to be achieved. In this step, the researchers need to define the main theory options to achieve the results, which will require a systematized review of the literature.
Step 2—Download scientific data and uploading on the platform. For each objective, it applies the definition of search terms to generate data files in scientific databases for bibliometric analysis. The data exported from these databases will serve as raw material to feed the Smart Bibliometrics system. Next, the generated data files are embedded in an intranet sandbox allowing the system to upload the files to the online servers.
Step 3—Cloud computing. After loading this data, it develops a robust cloud database that sets the processes of extracting, transforming, and loading the data, processing it into bibliometric information. Consequently, it generates several automatic analyses and visualizations to facilitate the choice of proper references. An algorithm was programmed to important transformations: deletion of duplicate records, blank lines removal, standardization author's name, among other changes to generate information. After completing the data processing, the system is available to show several automated analyses.
Step 4—Analysis. The user must browse the system functions and validate the following information: scientific productions worldwide, authors and co-authors, journals and editors, history of publications, language, graphs of relationship between authors and journals, table for downloading data, and classification of journals by impact factor. Other methods present similar analyses but with some limitations. This methodology extends the analysis by allowing users to gather insights as they navigate among functions, enabling the definition and selection of the bibliography to support the study. After performing these analyses, the researcher holds sufficient resources to define the journals to publish the study and to opt for a journal that concentrates on that subject. The visualizations form a loop analysis, allowing users to systematically apply new filters to the reports as they navigate through those visualizations, bringing the bibliometric data closer to the research objectives.
4.1—Systematic review of the literature. The visual graph was developed to select the authors who published in the preferred journal, listing the scientific documents ordered by the Smart index. In this visualization, a strategically ordered matrix was arranged for the previous reading of all publications abstracts in that journal, detailing the year information, title, authors, and the download process in the database. The function allows the automatic download of scientific production in pdf format to compound the bibliography and scientific writing.
Step 5—Data export for control and management. The data export function allows downloading the data in tabular format to facilitate the registration of complementary information and management of the documents to support the research.

By considering the distinct methodologies, the table presents a summary comparing the two main methodologies that grounded the study, i.e., the "Methodi Ordinatio" and the "Bibliometrix", for aggregating the analysis of classification metrics and graphical visualizations (Table 1).

Table 1 Comparison between the smart bibliometrics, bibliometrix, and methodi ordinatio methodologies

Item	Methodi ordinatio	Bibliometrix	Smart bibliometrics
Objective	Bibliometry. Ranking among scientific publications for selection of theoretical framework	Science mapping Information on the panorama of scientific production with graphic exhibitions	Integrate Bibliometric analysis and Science Mapping. Interactive graphic visuals with ranking and classification of scientific productions. Strategic information over the panorama of scientific production focusing on the link between research issues, journals, and authors
Processing	Manual. Limited capacity of processing capabilities. It depends on the user's knowledge of manipulating spreadsheets	Manual. Download and installation of R. Limited processing capacity. Requires knowledge of R script	Automated. Cloud computing. Tool in tune with the concept of 'big data'. Powerful processing capability. It is independent of the Operational System
Interactivity	Manual classification. Table	Static. Manual selection of visuals	User interaction with the data. Application of filters and segmentation as the user navigates through the functions. Insights. Automated analysis with Artificial intelligence algorithms
Access	Local	Site with web processing	Cloud computing. Access to any location at any time on internet-connected devices
Update	Restart the process. History is preserved in a spreadsheet but will require manual updating	Restart the process. Lost history	Cloud servers are scheduled for automatic updates. History preserved. Incremental update

Over time the methodologies were refined, as new technologies have been implemented to automate the process of bibliographic analysis. If the Ordinatio Method applies bibliometric techniques to categorize documents according to the InOrdinatio, the feature of the Bibliometrix is presenting visual maps following the concept of science mapping. In both methodologies, the updating routine depends on users' direct intervention, either to manipulate spreadsheets or to run scripts in R. The process of updating the data must still be considered, even by requiring restarting the process at each update, which consumes considerable time in data management routine and executing the process manually. It is worth mentioning that the interaction with the data is a relevant factor in the analyses, observing the static interaction in both methodologies.

Technological resources have surprisingly evolved to facilitate the employment of automation routines joining bibliometric and science mapping as proposed in this methodology as a single Business Intelligence system. Cloud computing capabilities can revolutionize the way scientists perform analytics and update reports on accessible platforms with intuitive navigation. The interaction experience with the data will allow us to reach valuable insights and bring improved insights to select a bibliography from automated analyses.

Access to smart bibliometrics system

The system was developed for free or restricted access, depending on the implementation required by the user institution. In the free version, users have access to the system through the link: https://bit.ly/3GgYUZh, as presented in the interface in Fig. 3. To update the visualizations using the searching data in both Scopus and Web of Science databases, users must generate the file in the respective databases, then upload them to the following access links, respectively: Scopus https://1drv.ms/u/s!AnaPKgl31Atlal_3Tpw6OY1alUE?e=DX8W9V and Web of Science https://1drv.ms/u/s!AnaPKgl31AtlazgNcwsrDD0olCU?e=vGnLjr. The tutorial for updating the system is available through the link: https://1drv.ms/u/s!AnaPKgl31AtlgQ5SpADgZ1ITE3kV?e=YH1vqB.

Fig. 3 Initial interface and analysis

The main difference between the open and the restricted access version refers to the solution sharing and the possibility of downloading data. These restrictions do not disqualify the distributed system of the open version, which allows the distribution to all researchers freely.

The restricted version requires the implementation of an intranet to release the system through access authentication. An intranet is a local network of computers based on internet-supported communication protocols, used by organizations to provide restricted and controlled access to systems and information, through individual access authentication. In this prototype, an intranet system supports the Smart Bibliometric system through restricted access by authenticating login and password on a page linked to the educational institution domain, ensuring access to private or authorized data. Moreover, this development enables to escalate the solution in a university and replicate it in other organizations, distributing to specific users.

Access and permissions should be required to ensure that users access their searching history and receive subject-related information only. An administrator would be responsible for managing the access complying with the principles of governance and information security. After conceding the authorizations, users need a device with navigation capacity to use the services, accessing the system through individual login in a monitored intranet environment.

Whether the user has access to an open or restricted version, to update the system would be necessary to follow a similar procedure, in this case, the open version updating. The researcher will be required to search the scientific data in the Scopus and Web of Science databases and upload the scientific files to the link above. Next, the automated data processing would support the automatic extraction, transforming, and loading processes. After the information is processed, the researcher can access the system at any time and from anywhere from a device with an internet connection. The following Table 2 summarizes the visualizations and functions developed, highlighting visualizations and analysis dimensions metrics on each report page.

Features detailing

World research

World research is a function that presents the major authors. It has a cloud of words with key terms and scientific productions worldwide, also used to identify emerging themes. It contains essential visualizations that allow knowing what has been put into practice in various regions of the planet, promoting further understanding of the main discussions on a global scale. For that, click on the system home screen icon to be redirected to the visualization, as shown in Fig. 4. The report presents information processed from databases loaded after queries on "Waste management" topics consulted in Scopus and Web of Science.

Each page report allows segmenting the data by user ID, also searching files by name, year, citation range, impact factor range, journal, document type, authors, and title. Once you target one page, the filter will sync with other visuals and the analysis of alternative pages. In the Smart Bibliometrics system, for multiple select keywords, user should press control key and select. The system is dynamic and interacts with user selection, by pressing “Ctrl” also in the slicers to segment data.

Table 2 Description of metrics, visuals, dimensions, and analysis

Analysis	Metric	Visual	Dimensions
World research	Number of authors Number of Citations Number of records	Stacked Bar Chart Wordcloud Map	Authors Countries Key Words
Authors & coauthors	Number of authors Number of citations Number of Records Smart Index	List Stacked Bar Chart Matrix	Authors Research groups Paper Title
Journals	Number of Authors Number of Citations Number of Records Number of posts per author	Stacked Bar Chart Smart Narrative (Inteligência Artificial) Tornado Diagram	Journals Authors
Publisher, journals, and time series	Number of Authors Number of Citations Number of Records	Treemap Stacked Column Chart	Publisher Journal Year
Language	Number of Authors Number of Citations Number of Records	Sankey Diagram Journey Chart	Document Type Language Clusters
Smart graphs	Number of Authors Number of Citations Number of Records Smart Index	Network Navigator Matrix	Authors Title Abstract TWO Link download Journals Clusters
Export data	Number of Authors Number of Citations Number of Records Smart Index Impact Factor	Word Cloud Table	Key Words Year Title Journal Abstract Link download or ID TWO
Impact factor	Percentage of Strata Number of Authors Number of Citations Number of Impact Factor Records	Funnel Chart Table	Journal Subject Area Impact Factor

Note that the charts are interactive, and the filters are applied when selecting any visualization, reflecting on the dashboard context. This feature makes a considerable difference in obtaining relevant insights that effectively increase the analysis capability. For example, by selecting the first author of the list, we can check the main keywords, number of records, citations, number of authors and co-authors, and the geographic region of the publication. All related metrics fit the filter context, bringing intelligence to the entire process.

Author and coauthors

The report presents data on authors and co-authors using interactivity. By navigating through visualizations, users can mix crucial information to detect relevant authors, metrics of records by author, groups, citations, and details of the co-author team. That enables

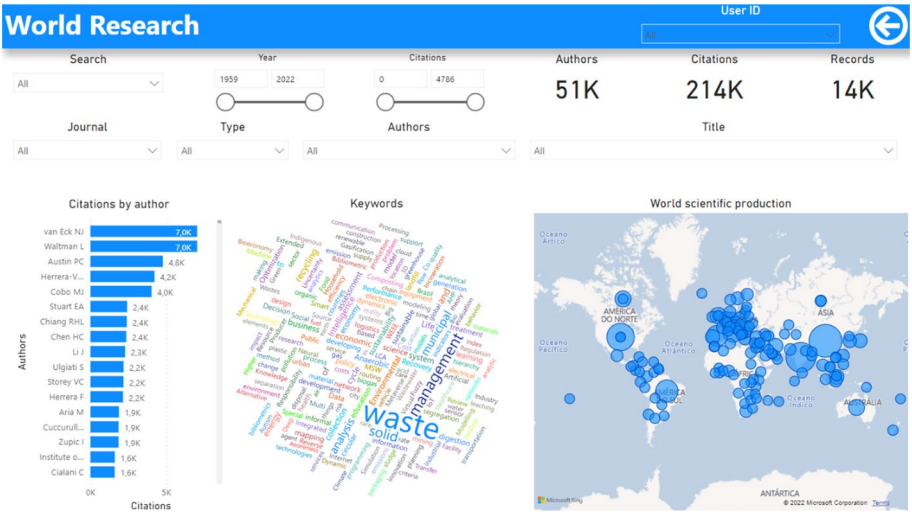

Fig. 4 World research

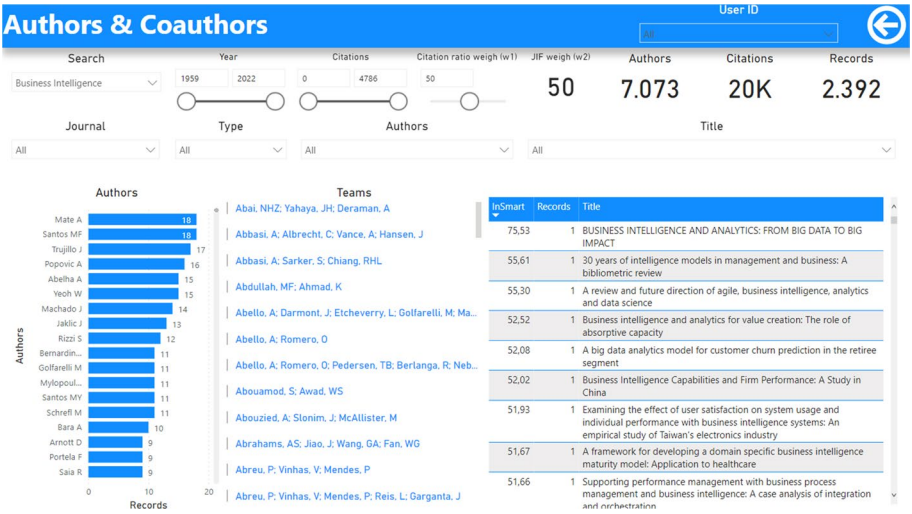

Fig. 5 Authors and coauthors

an escalation of scientific productions by using the Smart Index defined in Fig. 5, applying the filter "Business Intelligence".

The author selection is displayed throughout the report, presenting the groups involved in the various scientific productions.

Journals

The report lists specialized journals, the number of citations, authors, and records. When the researcher decides to publish the study, finding a specialized journal to submit

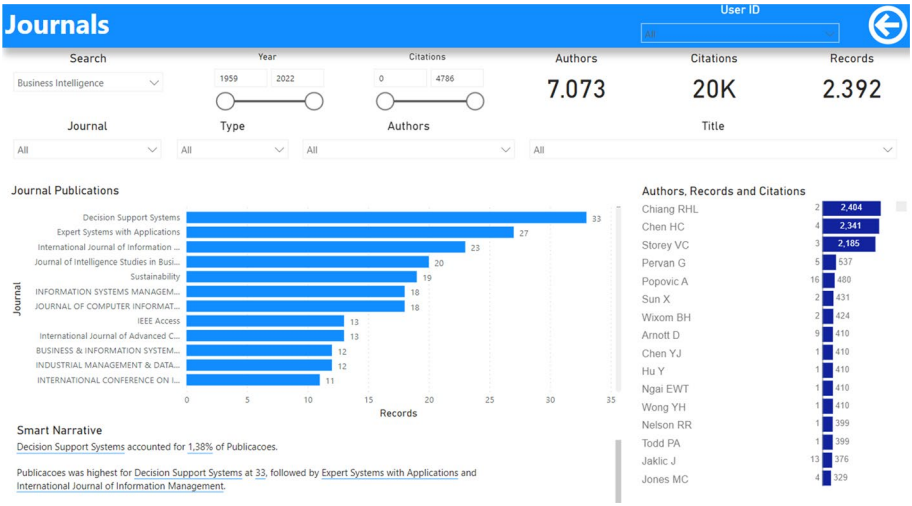

Fig. 6 Journals

the scientific article is a crucial step. For this purpose, the visualization presented in Fig. 6 privileges this analysis by correlating the journals that may have an interest in specific research subjects and linking the lead authors involved in scientific discussions on that subject in that journals' list.

This solution also features artificial intelligence capabilities for automatic analysis, such as the ability to highlight smart narrative analyses. The algorithm computes the percentage of leading journals' publications, considering the selected sample. This reinforces the reality that artificial intelligence capabilities can produce automated analysis. Intelligent writing is performed by algorithms that perform automated analysis, searching for data patterns and providing insights while the user navigates through the report's filter options.

Publisher, journal and time series

The Treemap visualization displays the top editors, in order of importance by the number of publications. The hierarchy follows the upper left-bottom vertical targeting. The column chart displays the distribution of publications over time. The history analysis demonstrates whether a particular subject has been debated by the academic community over time. An example is the highlighted trends in Business Intelligence in recent years shown in Fig. 7.

Language

English is the predominant language of international publications. However, it is possible to use other languages for research, depending on the criteria used in the datasets search. The visualization demonstrated in Fig. 8 provides an overview of the main types of documents and the corresponding publication language.

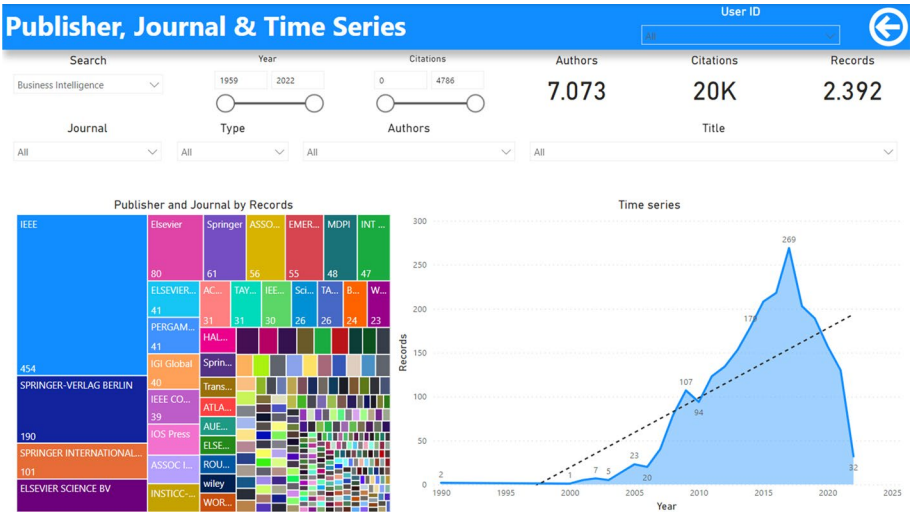

Fig. 7 Publisher, Journal and Time Series

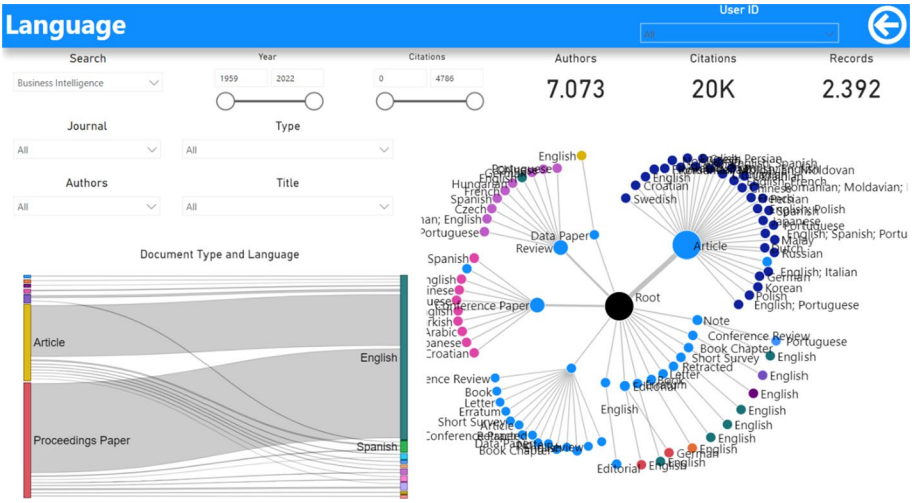

Fig. 8 Language

Detailing core analysis—smart graphs

A great advantage of the new methodology is the user navigation experience through visualizations when the segmentation is applied to the data to obtain insights and propose a new form of interaction. Compared to other solutions such as Bibliometrix, or visualization software such as VOSviewer or Gephi, the visualization is static while this methodology allows dynamic navigation by applying segmentations to the data. Neuroscience provides explanations about the influence of visualization on decision-making processes. For instance, when the interpretation of the human brain of a data set is facilitated by the

improvement caused by the visualization, leading to a decision-making action (Roselli et al., 2019).

For reasons of convenience, the Smart Grafos visuals will be detailed for the relevance of their interactive visual to guide the researcher's selection of theoretical framework and download of documents in pdf format. In the dashboard, the main information of the latest publications will be displayed to help decide whether to use them as references according to a cross-analysis of journals, authors, and scientific productions on a research subject. This analysis has a high potential to increase publication chances by directing scientific writing to research gaps oriented to the expectations of a particular journal, considering the latest theoretical constructions elaborated by the scientific community.

There are two main graphs: a matrix with the information of the scientific works ordered according to the Smart index for each publication. The higher the value, the more representative the scientific production may be. Clicking on the [+] icon expands the matrix fields according to Fig. 9 which successively displays the information following the hierarchy: author, title, summary, DOI, download link.

The visual matrix allows navigation through the main bibliometric elements, making it functional for the user interaction with the system, settling with greater flexibility the importance of the reference in the research context. The Smart Bibliometrics will always be updated to the filters applied, providing a sequential analysis according to documents that follow the parameters reported and preserving users' browsing history. A cutout in the year of publication, theme, and document type will considerably narrow the result for agile analysis.

Thus, this visualization facilitates the systematic review of the literature to analyze the primary information of articles that will form the theoretical framework in a sequential and orderly manner. A previous reading of the abstract helps to decide whether or not the article is used as a reference or read the full text. When navigating the matrix hierarchy and confirming the relevance of the scientific production to the research, the user can copy the link and paste it into the web browser and, from that point, will be automatically directed to downloading the document, since having access to the Scopus

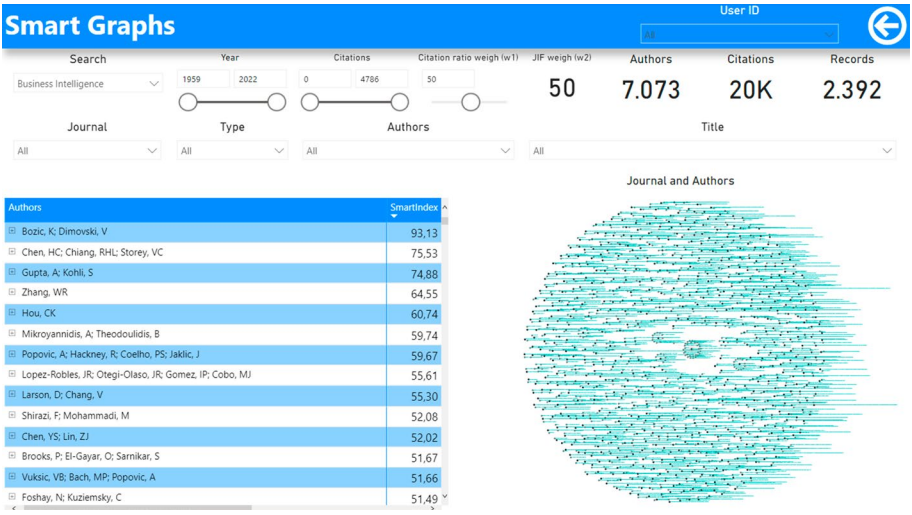

Fig. 9 Smart graphs—matrix and graphs visuals

and Web of Science databases. Right-clicking on the link in matrix enables the copy value to get the scientific file document.

The second visualization of the graph "Relationship—Authors and Journals" shows the authors' connections and the main journals. It consists of vertices and edges, creating an interconnection between journals and authors in each investigation area. Each junction represents a journal. By selecting one, it will be possible to visualize (Fig. 10) the connection among various authors and their contributions to a specific subject. The systematic study of each author's contributions significantly increases the probability of identifying a relevant research gap, guided by a targeted and focused reading.

Similarly, when selecting only one of the authors, the work will be described with the information of the respective matrix visualization, as seen in Fig. 10. The interactivity of the visualization considerably increases the power of data analysis. Authors' scientific contributions, grouped according to the target subject and journal, besides revealing discussions of the scientific community, have a high potential to bring clarity to the various publication vehicles' expectations about their specialization in a specific research field. It is an agile method considering the need for a systematized literature review for offering a logical sequence of articles' priorities, as well as linking journals, authors, and ideas to the research problem. Such analysis can direct the work development with great assertiveness, converting intelligence into writing focused on results and publication.

Export data

The function Export Data allows exporting the data in a spreadsheet; however, this feature is only available in its restricted version, being necessary a subscription required by Microsoft. Still, the user can filter the data using the word cloud or other slicers and copy values by right-clicking (Fig. 11).

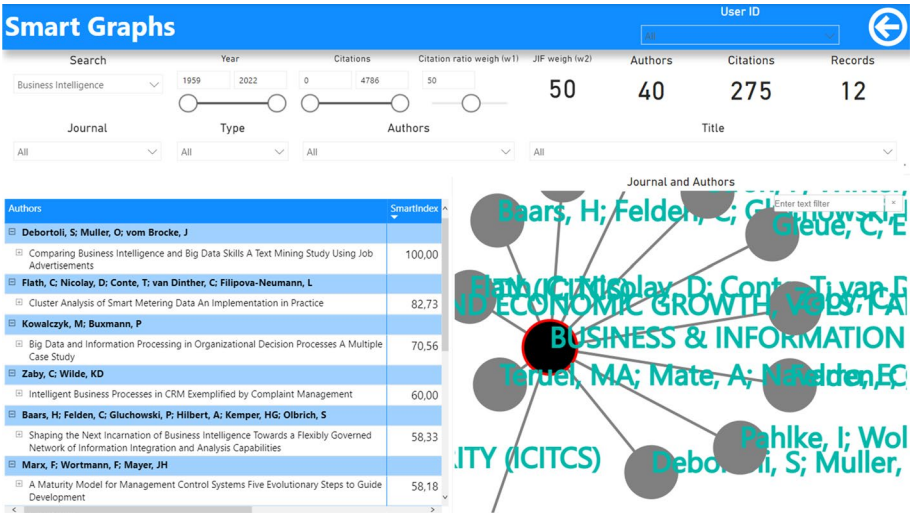

Fig. 10 Graphs of authors and journals

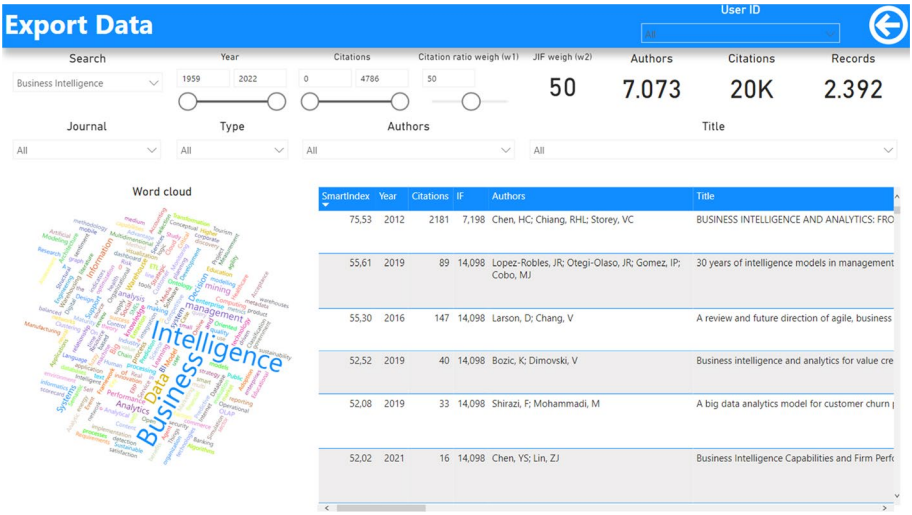

Fig. 11 Export data

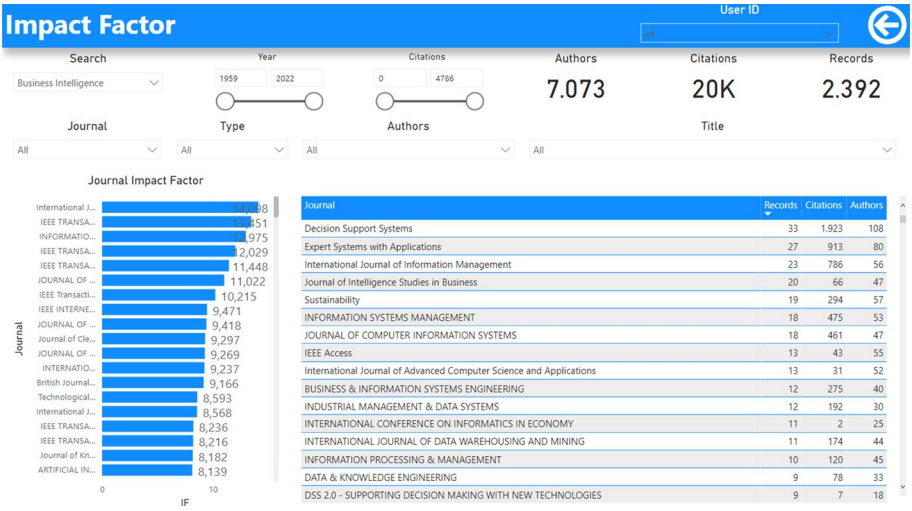

Fig. 12 Impact factor

Impact factor

The visualization was developed to complete the analysis cycle when the research has an overview of the impact factor of journals sample extracted from Scopus and Web of Science databases. As can be seen, the key metrics related to each journal are summarized, including impact factor index, records, number of citations, and authors (Fig. 12).

Finally, it is worth noting that this analysis should not exclude or restrict the selection to documents with higher Smart Index ratings. Scientists should be reasonable to not exclude those documents that, although having lesser important Smart metrics, contain essential

information to use in the research. Another fundamental point that stands out in this 'other methodologies' system is the possibility of classifying documents that do not employ the Impact Factor metric, as occurs with scientific articles. According to Pagani et al. (2015), documents like book chapters or complementary materials could be affected in the InOrdinatio analysis for present inexpressive values. In this methodology, the problem's solution is segmenting the data based on document type, allowing the classification of complementary bibliography even if it is not in article format.

Processing and new developments

The Minimum Viable Product (MVP) preliminary processing tests occurred stably and efficiently processed the data of all users who used the platform during the testing period presenting ample capacity to include new profiles. The processing occurred in a scheduled manner, the tool was able to generate information from a cloud database with the availability of controlled access by using login and access passwords. By applying this method, future validation steps act when results would be measured with methodology satisfaction and effectiveness evaluations.

There are system improvements in progress, including the possibility of processing data from other databases such as Digital Science Dimensions, PubMed, Cochrane (Dimensions & Inc., 2021; John Wiley & Sons, 2021; PubMed, 2021), also, text mining functions, visualizations with artificial intelligence features and cluster analysis.

Final considerations

The information age has significantly changed organizations' decision-making processes. If recently emotions and feelings were strong triggers to justify options, currently choices are centered on data, information, and knowledge. And in academia, these processes grow in importance. The data available on scientific portals are essential sources to reveal strategic information that may indicate new research trends. We have moved from the "Information Age" to the "Age of Knowledge", a new configuration where data volumes analysis and interpretation are the base for decisions, especially when considering the scientific research environment.

The Knowledge Age carries the debate on open science. Years ago, science practices were restricted to laboratories, today, they have broad and democratic participation with the contribution of researchers worldwide, promoting diversity of ideas. Arising queries about open science are related to the access to data as a critical characteristic for an efficient and progressive system and the importance of data sharing for the evolution of science (Hardwicke et al., 2018). That said, new systems need to promote collaborative science, materializing bibliometric theories for democratic access and sharing of scientific discovery.

The world of technological solutions and the hyper connection is currently on solid ground, connecting people and finding solutions to studies developed in new democratic practices of bibliometrics and science in general. A related issue is a political dimension that falls within the broader field of open science (Lyon, 2016). The technological transformation and the emergence of systems developed for open access, with unrestricted participation, can influence public policies in the natural path of science democratization. New

technologies impact the way we communicate and interact, then science is characterized by construction in multiple perspectives and experiences.

Distributing the system open version with free of charge advanced cloud computing capabilities, strengthens this movement of democratization of science, expanding the possibilities of overcoming research challenges and problems that would be dealt with collectively.

Regular updating has become a challenge for any professional in the age of knowledge, especially in the academy. Producing relevant scientific material requires considerable dedication and time. Therefore, the developmental method has a high potential to optimize scientific production and facilitate researchers' studies on bibliometric analysis considering the volume of data currently available.

Specialized journals have become increasingly demanding. In this way, a work based on consistent bibliometric analysis is fundamental in raising the chances of publication success. Having a broad overview of the "state-of-the-art" subject at the beginning of the research is a major attribute that determines a work approval for publication in a renowned scientific journal.

Finally, science mapping and bibliometric processes, through the Smart Bibliometrics, can potentially automate manual and routine processes, connect people and ideas, providing more agile analysis in choosing relevant scientific productions in an innovative, simplified, and accessible manner. With this system, scientists will have a powerful tool that provides strategic information, increasing the chances of success in publications by identifying significant research gaps and contribute to an open science. The expected result is a solution that simplifies researchers' work, expands assertiveness in scientific production, and contributes to knowledge development.

Acknowledgements We appreciate anonymous reviewers' contributions and recommendations. All suggestions were very helpful, and we believe that by addressing the comments, our revised manuscript was significantly improved. Also, we would like to acknowledge Roberta Ribeiro Souza, who helped us to revise the English Language.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahlgren, P., & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping. *Scientometrics, 76*(2), 273–290. https://doi.org/10.1007/s11192-007-1935-1

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

Božič, K., & Dimovski, V. (2019). Business intelligence and analytics for value creation: The role of absorptive capacity. *International Journal of Information Management, 46*, 93–103. https://doi.org/10.1016/j.ijinfomgt.2018.11.020

Chavalarias, D., & Cointet, J. P. (2008). Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study. *Scientometrics, 75*(1), 37–50. https://doi.org/10.1007/s11192-007-1825-6

Chen, C. (2005). The centrality of pivotal points in the evolution of scientific networks. *International CoNference on Intelligent User Interfaces Proceedings IUI*. https://doi.org/10.1145/1040830.1040859

Chen, C. (2013). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 64*, 1852–1863.
Chen, H., Chiang, R., & Storey, V. (2018). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.
Clarivate. (2017). *In Memoriam: Dr Eugene Garfield*.
Clarivate. (2022). *Journal Citation Reports*. https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/
Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2013). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology, 64*, 1852–1863.
De Carvalho, G. D. G., Sokulski, C. C., Da Silva, W. V., De Carvalho, H. G., De Moura, R. V., De Francisco, A. C., & Da Veiga, C. P. (2020). Bibliometrics and systematic reviews: A comparison between the Proknow-C and the Methodi Ordinatio. *Journal of Informetrics, 14*, 3. https://doi.org/10.1016/j.joi.2020.101043
Dervis, H. (2019). Bibliometric analysis using bibliometrix an R package. *Journal of Scientometric Research, 8*(3), 156–160. https://doi.org/10.5530/JSCIRES.8.3.32
Dimensions, & Inc. (2021). *Linked research data from idea to impact*. https://www.dimensions.ai/
Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*(March), 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
Dwyer, T., Cordeil, M., Czauderna, T., Delir Haghighi, P., Ens, B., Goodwin, S., Jenny, B., Marriott, K., & Wybrow, M. (2020). The data visualisation and immersive analytics research lab at Monash University. *Visual Informatics, 4*(4), 41–49. https://doi.org/10.1016/j.visinf.2020.11.001
Ensslin, L., Dutra, A., Ensslin, S. R., Chaves, L. C., & Dezem, V. (2015). Research process for selecting a theoretical framework and bibliometric analysis of a theme: Illustration for the management of customer service in a bank. *Modern Economy, 06*(06), 782–796. https://doi.org/10.4236/me.2015.66074
Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science, 30*(2), 119–145. https://doi.org/10.1177/0165551504042802
Garfield, E. (2009). From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics, 3*(3), 173–179. https://doi.org/10.1016/j.joi.2009.03.009
Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Mohr, A. H., Clayton, E., Yoon, E. J., Tessler, M. H., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science, 5*, 8. https://doi.org/10.1098/rsos.180448
John Wiley & Sons, I. (2021). *Cochrane Library*. Potomki
López-Robles, J. R., Otegi-Olaso, J. R., Porto Gómez, I., & Cobo, M. J. (2019). 30 years of intelligence models in management and business: A bibliometric review. *International Journal of Information Management, 48*, 22–38. https://doi.org/10.1016/j.ijinfomgt.2019.01.013
Lyon, L. (2016). Transparency: The emerging third dimension of open science and open data. *LIBER Quarterly, 25*(4), 153–171. https://doi.org/10.18352/lq.10113
Massimo, A., & Cuccurullo, C. (2021). *Biblioshiny: the shiny interface for bibliometrix*. https://bibliometrix.org/About.html
Maximo, A., & Corrado, C. (2021). *biblioshiny: The shiny interface for bibliometrix*.
Murgado-Armenteros, E. M., Gutiérrez-Salcedo, M., Torres-Ruiz, F. J., & Cobo, M. J. (2015). Analysing the conceptual evolution of qualitative marketing research through science mapping analysis. *Scientometrics, 102*(1), 519–557. https://doi.org/10.1007/s11192-014-1443-z
Noyons, E. C., Moed, H., & van Raan, A. F. (1999). Integrating research performance analysis and science mapping. *Scientometrics, 46*(3), 591–604. https://doi.org/10.1016/S0138-9130(00)86853-7
Otlet, P. (1934). *Traité de documentation: Le livre sur le livre. Théorie et pratique* (Editiones Mundaneum (ed.)).
Pagani, R. N., Kovaleski, J. L., & Resende, L. M. (2015). Methodi Ordinatio: A proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. *Scientometrics, 105*(3), 2109–2135. https://doi.org/10.1007/s11192-015-1744-x
Pallottino, F., Biocca, M., Nardi, P., Figorilli, S., Menesatti, P., & Costa, C. (2018). Science mapping approach to analyze the research evolution on precision agriculture: World, EU and Italian situation. *Precision Agriculture, 19*(6), 1011–1026. https://doi.org/10.1007/s11119-018-9569-2
Persson, O. (2017). *BibExcel: a tool-box developed by Olle Persson*. https://homepage.univie.ac.at/juan.gorraiz/bibexcel/

Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. *International Society for Scientometrics and Informetrics., 95*, 1–10.

PubMed. (2021). *National Library of Medicine*. https://pubmed.ncbi.nlm.nih.gov/

Rodríguez-Bolívar, M. P., Alcaide-Muñoz, L., & Cobo, M. J. (2018). Analyzing the scientific evolution and impact of e-Participation research in JCR journals using science mapping. *International Journal of Information Management, 40*, 111–119. https://doi.org/10.1016/j.ijinfomgt.2017.12.011

Roselli, L. R. P., de Almeida, A. T., & Frej, E. A. (2019). Decision neuroscience for improving data visualization of decision support in the FITradeoff method. *Operational Research, 19*(4), 933–953. https://doi.org/10.1007/s12351-018-00445-1

Shollo, A., & Galliers, R. D. (2016). Towards an understanding of the role of business intelligence systems in organisational knowing. *Information Systems Journal, 26*(4), 339–367. https://doi.org/10.1111/isj.12071

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics, 38*(2), 275–293. https://doi.org/10.1007/BF02457414

Soós, S. (2011). The functional anatomy of science mapping: Katy Börner: Atlas of science: Visualizing what we know. *Scientometrics, 89*(2), 723–726. https://doi.org/10.1007/s11192-011-0480-0

Soós, S., & Kampis, G. (2012). Beyond the basemap of science: Mapping multiple structures in research portfolios: Evidence from Hungary. *Scientometrics, 93*(3), 869–891. https://doi.org/10.1007/s11192-012-0713-x

van Eck, N. J., & 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

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics, 10*(3–4), 157–177. https://doi.org/10.1007/BF02026039