Bibliometric Tools for Academia

What Are Bibliometric Tools? And How Do Bibliometric Tools Differ from Research Discovery Tools?

Bibliometric tools, at their core, integrate the data available from bibliographic data sources (as discussed in chapter 1) and make the data available in the form of bibliometric indicators. There are a variety of standard and proprietary bibliometric indicators that vary among the available tools, which will be discussed with some further detail in this chapter. We must be careful not to confuse the limited bibliographic data sources with the almost countless bibliometric tools or technologies available today. Many bibliographic data sources will provide some very limited bibliometrics, such as some basic descriptive analysis based on the number of publications, authors, and so on, while others provide slightly more robust bibliometrics, such as results sets analysis, views of author profiles that contain bibliometrics, and views of some institutional level metrics (see table 2.1). Still, these are not often considered bibliometric tools because their main function is not to provide bibliometric analyses, but rather research discovery.

Yet it is difficult to precisely define how bibliometric tools differ from the typical research discovery tool since linking bibliographic data together is at the core of both. However, bibliometric tools provide richer data sets and analysis functions that

- rely on more complex mapping of the bibliographic data, particularly the citation data;
- aggregate or summarize this data into bibliometric indicators;
- allow for in-system visualizations and saving of data sets for additional analysis outside the system; and
- use data at a scale that requires significantly more computer processing power to provide this data.

Table 2.1: Bibliographic Sources with analytical views

Product Name	Owner/Developer
Scopus	Elsevier (RELX)
Web of Science	Clarivate

Bibliometric tools can be divided by the types of analysis that they attempt to perform, with two major classifications:

1. Descriptive bibliometric analysis tools summarize data and indicators such as total publications over time, citation counts, author counts, and other more complex computations of indicators.
2. Descriptive network analyses (often referred to as knowledge mapping or knowledge graphs) compute and visualize connections between bibliographic variables such as authors, keywords, affiliations, and so on.

Most practitioners of bibliometrics will engage with additional tools to help with the analysis and visualization of the bibliometrics outside of their system of choice, such as Excel, SQL, R, Python, Jupyter, GYPHY, Pajek, Tableau, Power BI, and so on. When going beyond the descriptive bibliometrics and network analysis, some researchers and practitioners will also employ statistical analysis tools such as SPSS, Excel Analysis ToolPak, SAS, or R. These ancillary tools are beyond the scope of this report.

Current Bibliometric Tools and Their Features

This section reviews the current bibliometric tools available, including descriptive bibliometric analysis tools and bibliometric and network analysis tools. Also discussed is the recent explosion of discovery tools that employ analytical views and network analysis.

Descriptive Bibliometric Analysis Tools: The Major Commercial Players

Table 2.2: Entity type comparisons across major commercial bibliometric tools

	InCites	*SciVal*	*Dimensions*	*Lens.org*
Entity Types Available for Analysis				
Publication sets	yes	yes	yes	yes
Authors (individual)	yes	yes	yes	yes
Custom author groups	no	yes	no	no
Institutions	yes	yes	yes	yes
Geographic regions	yes	yes	yes	yes
Journal-level research categories	yes	yes	yes	yes
Article-level research categories	yes	yes	yes	yes
Source titles	yes	yes	yes	yes
Funding bodies	yes	yes	yes	yes
Patents	no	no	yes	yes

InCites, SciVal, Dimensions, and Lens.org are the current major commercial bibliometric tools available that provide web-based applications with in-system analysis in a relatively user-friendly manner that does not require any coding or data-cleaning knowledge. These systems are ideal for the bibliometric practitioner who wants a relatively broad range of descriptive statistics about research outputs and impact. InCites, SciVal, and Dimensions are all subscription-based. Lens.org currently allows free access for noncommercial use to individuals and sells subscriptions to commercial users and institutions. Dimensions has a limited free view and enhanced subscription access. These systems stand out from other systems because they are aimed at generalist users and do not require any specific technical knowledge, such as application configurations or programming languages. However, the systems still use quite sophisticated analytical functions in the background and present them in their web-based applications.

These systems are aimed at a wide breadth of users including researchers, academic institutions, publishers, funders, and research and development departments of commercial enterprises.

SHARED FEATURES

- are web-based, intuitively structured applications with no downloading or local installations required
- allow the creation and analysis of aggregated bibliometric data that is based on the selection or creation of the various entity types available in the system (See table 2.2 for the types of entities available in these tools.)
- present data in tables or charts with download options (All allow CSV, among other options.)
- allow the saving of data in reports or dashboards within the web application
- offer user guides, tutorials, technical services, and substantial ongoing system development road maps

BASIC DATA STRUCTURES IN THESE SYSTEMS

InCites, Dimension, and Lens.org all work on a filtering basis, similar to the experience of searching within a research database (see figures 2.1, 2.2, and 2.3). This means that the system starts with the entire data universe available in the system and then allows the use of filters to narrow down the data set. For example, a user can specify an institution or institutions they plan to analyze and then filter by a subject category. This allows for significant flexibility within the system for analysis. The entity types listed in table 2.2 are used to view the created data set from the perspective of the selected entity type. Using this example, the data set for the specified institution and subject category could be analyzed by author, source titles, funding bodies, and so on.

SciVal begins analyses by creating and saving custom data sets that are then added to an entity staging area where they can be selected under different views that allow analyses such as benchmarking and trend analysis (see figure 2.4). Although SciVal allows for only a limited amount of filtering and customization of entities on the fly, its advantage is that multiple custom or preloaded data sets can be selected and benchmarked together.

Entity Types—The Key to Understanding the Features of These Major Bibliometric Tools

Table 2.2 presents an overview of the main entity types available for analysis in each of these major commercial bibliometric tools. Entity types are not the same as the content types (see table 2.3 on p. 17 for content types) as they integrate data from the various data sources associated with the content types to allow for analysis. For example, in InCites, the author entity type (called Researchers in InCites) can be analyzed in a variety of ways, including looking at the number of documents an author has published that have been cited by patents. However, because patents are not an entity type in InCites, patents themselves cannot be analyzed, and the data is therefore limited to the single patent-citation metric. On the other hand, both Dimensions and Lens.org have patent entity types that have a more robust set of data analyses and indicators associated with patents.

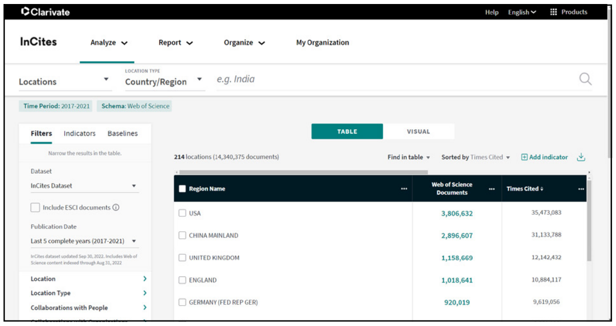

Figure 2.1
InCites interface with characteristic filtering on left-hand side of the screen. The Locations module is selected here; other options are Researchers, Organizations, Research Areas, Publication Sources, and Funding Agencies. No filters applied.

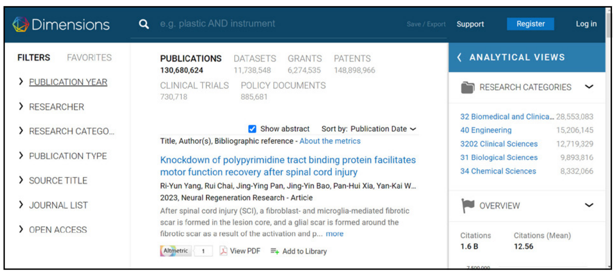

Figure 2.2
Dimensions interface with characteristic filtering on left-hand side of the screen. Publications is selected as the entity type; other entity types for analysis are Datasets, Grants, Patents, Clinical Trials, and Policy Documents. No filters applied.

PUBLICATION SETS

Publication sets are the core type of data set for any bibliometric system. They are highly customizable sets of documents and can be achieved through two main methods: (1) defined by search queries or other bibliographic metadata filters or (2) imported documents via direct connection from the data source or using persistent identifiers (e.g., DOIs). The specific steps to create, save, and analyze publication sets vary across these main bibliometric tools; however, they all use both methods to achieve these goals. In both SciVal and InCites, static publication sets can be created within the system using filters, by beginning with other data sets or from an export from the data source (i.e., Scopus for SciVal or WOS for InCites). Dimensions and Lens.org (and in a limited way SciVal using its Research Areas builder) achieve a similar goal by allowing researchers to save advanced Boolean search queries within the system. In this way these three bibliometric tools act as discovery tools as well. The advantage to this method is that the publication sets can be more dynamic, updating any time they are selected to run in the system, exactly like a saved search option available within many research databases. If you do not like this feature and need

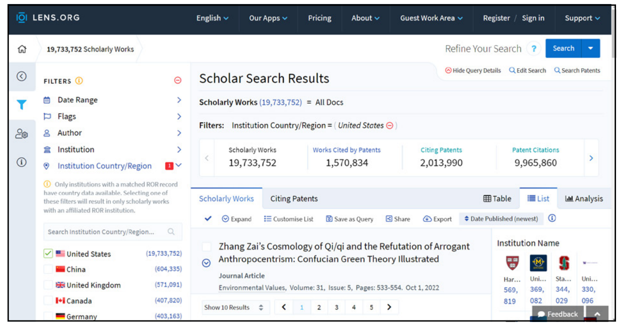

Figure 2.3
Lens.org interface with characteristic filtering on left-hand side of the screen. The Scholarly Works app is selected. There is also a Patents app available that allows similar search and analysis functions as the Scholarly Works app. A filter for United States publications is applied.

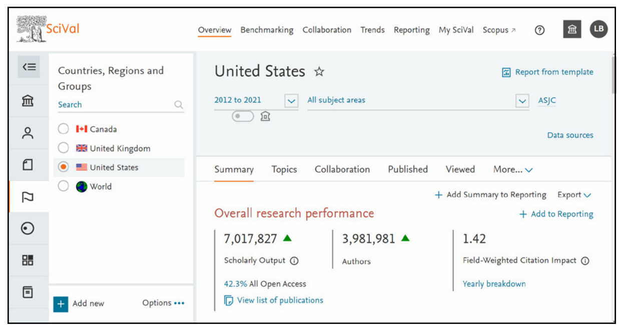

Figure 2.4
SciVal interface with characteristic entity staging area on left-hand side of the screen. The overview module is selected with the Countries, Regions and Groups entity type selected. The United States has been added to the staging area panel and selected. Data is filtered to the 2012–2021 publication years.

your publication sets to be a snapshot in time, Lens .org also provides the option for the search queries to be dynamic or static.

SciVal, InCites, and Lens.org allow publication sets to be bulk-uploaded using unique document identifiers, such as DOIs or system-assigned document numbers. These publication sets will always be static lists of publications but have the advantage of being highly customizable to the users' needs. These are usually managed under a tool that will save the files in folders within the system. It's not clear whether this same functionality is possible in Dimensions.

AUTHORS AND CUSTOM AUTHOR GROUPS

Authors, researchers, scholars—as they are variously called within these systems—present several challenges for analysis:

- author disambiguation,
- analysis of individual authors, and
- analysis of custom author groups.

Regarding disambiguation, all these systems use machine learning algorithms to help disambiguate authors and match them to documents within the systems. These algorithms usually take into account available metadata, such as name variants, existing author IDs (such as ORCID), affiliation data, research fields, journal names, common coauthors, and publication years. SciVal mints a unique identifier for each author, called the Scopus ID, and allows for merging and corrections when errors are found. InCites (as of 2022) has begun integrating its proprietary WOS Author Record into its Researchers filters within the system, and we can anticipate that researchers will be easier to disambiguate once these WOS Author Records reach full launch. Dimensions does not display a unique author ID minted by its system, but it integrates and displays ORCID and Scopus IDs and allows corrections to be requested from its customer service. Lens.org does not display a unique author ID minted by its system, but it will display an author's ORCID or Microsoft Academic ID if available.

All these systems allow individual authors to be analyzed and show up in author lists based on the selected bibliometrics within the system. They all have some type of author profile link as well that brings together the author's publications and usually lists affiliation, coauthors, and other simple bibliometrics that can be helpful to authors looking for their metrics or to verify an author's identity. InCites and SciVal link to author profiles outside of the bibliometric tool (i.e., to WOS or Scopus), while Dimensions and Lens.org contain the profiles within their systems.

Custom author groups refers to an in-system tool that allows the creation and management of author groups as a single entity that can be analyzed. Technically, a search could be conducted in any of these systems that combines all the required authors based on name or some unique author identifier (e.g., ORCID); however, SciVal is the only system that currently has an in-system tool for author and author group management that allows dynamic grouping, creation of hierarchies, and bulk editing of the group. Dimensions allows the selection of authors by using filters and stores the group in a workspace where groups can be edited or downloaded, but it does not allow hierarchies in the data structure. The custom author groups feature is of particular interest to the analyses of whole departments or specific research teams within an institution because of the customizability (see chapter 3 for an example). This type of analysis might also be achieved using a Current Research Information System, or CRIS; however, this report does not cover such systems in detail as they focus on institutional research information that displays and connects robust information about authors and the groups and research areas they belong to. In-depth bibliometric analysis is not the primary focus of these systems, although they do often display some indicators on the interface.

INSTITUTIONS

The analysis of whole institutions is ubiquitous within these systems, and over the last several years several of the bibliometric tools have improved the reliability of the institutional level data through either reviewing the accuracy of the affiliation names and hierarchies or through better integrating organizational unique identifiers such as the Research Organization Registry (ROR) and Ringgold ID. InCites (via WOS) and SciVal (via Scopus) recently underwent large reviews of their in-system institutional hierarchies to help validate and better capture affiliation data. Lens.org uses the ROR identifier to aid in disambiguating institutions. Dimensions links institutions with relational data but keeps each institution separated, outside of a hierarchy, to allow for more granular analyses.

GEOGRAPHIC REGIONS

Geographic regions are identified by the location of the affiliation. This is a standard, straightforward entity type across all these bibliometric tools. Therefore, there is not a lot of variation in how the names data is presented in these systems; most present country names and major regions like North America, Asia Pacific, and so on. A recent study by Guerrero-Bote et al. (2021) found that when aggregating data to the institution or country level, the Scopus data set maintained a larger count of documents/citations than did Dimensions, despite Dimensions having a greater total count of publications in its system. This illustrates that the completeness of the metadata fields impacts the filtering and data aggregation capacity of the system. Therefore, to aid in data validation and reporting, users should be familiar with the intricacies of their chosen tools.

Also, the types of indicators (i.e., publication counts, citations) available and the download options for visualizations, particularly geographic mapping, can vary greatly among these tools. Ideally, the data is downloadable from the data table and an image file is available from the system. This enables the user to either render a visualization external to the tool or take advantage of the images from the system. InCites appears to be the only system that allows proper image files (PNG, GIF, etc.) of visualizations to be downloaded. This is not an insurmountable challenge as the data tables can be used to create maps with external software, such as Tableau, Microsoft Power BI, R, or Leaflet, and a screen capture can create any needed image files. Users of geographic data usually want to be able to interact with the data but also want to add the maps to static reports; therefore, it is important to carefully consider how these bibliometric tools best meet a particular need. This is likely an area where we will continue to see improvements in these tools.

JOURNAL-LEVEL RESEARCH CATEGORIES

Journal-level research categories are research area classifications that have been assigned to a journal title, often referred to as a research area schema. There are many of these classifications. Some are specific to the bibliometric tool, such as the Web of Science Research Areas and the Scopus All Science Journal Classification (ASJC). However, there are many external schemata that have international or regional significance and have been mapped into the system using the existing classification structures. For example, the field of Research and Development, a classification scheme by the Organisation for Economic Co-operation and Development (OECD), is mapped to the Scopus ASJC. These types of mapping and use of classifications at the journal level allow for broad level analysis of knowledge domains. There are also critiques of these journal-level classifications. They tend to mask the true topic of a scholarly work by grouping all the works under the subject area of the journal. This is particularly problematic for multidisciplinary journals that cover a range of research fields. Journal-level research categories are available only in SciVal and InCites. Although Dimensions and Lens.org employ these journal-level schemata as research area filters in their system, they apply these at the article level. This means that individual articles are being automatically reviewed through a machine learning algorithm and then a research area is applied regardless of the journal the article is published in.

ARTICLE-LEVEL RESEARCH CATEGORIES

Article-level research categories are research area classifications that have been assigned to an individual document. These classifications usually employ machine learning algorithms that create dynamic and ever-growing thesauri or a controlled list of topics that are assigned to the article and typically involve a much broader set of categories than the journal-level classifications. They differ from author-assigned keywords because the use of thesauri is meant to help standardize terms and reduce duplication or variants of the same concept. For example, they fix spelling variations—such as *floods, flooding,* and *flood*—by assigning a single term or phrase and will group similar topics based on citation linking or phrase analysis. All the bibliometric tools being reviewed in this section employ article-level classification using machine learning algorithms to match documents to topics.

SciVal's article-level classification has two hierarchical levels of classification, with new topics constantly emerging and being re-clustered using an in-house algorithm that is based on citation relationships. This approach allows for dynamic analysis and the identification of emerging research areas but poses issues with trend analyses and reporting reproducibility.

InCites' article-level classification has three hierarchical levels of classification, with most of the new topics and re-clustering happening at the lowest level. InCites uses an algorithm based on citation relationships developed by CWTS Leiden. The same analysis issues found in SciVal apply to InCites.

Lens.org's article-level classification does not have a clear hierarchical structure in the system; however, it uses OpenAlex as the data source, which does use a hierarchy for the concepts it defines. The OpenAlex algorithm uses the title and abstract of documents. Since the hierarchies are not available in Lens.org, the specificity of terms can vary greatly, causing more productive research areas to dominate the analysis when looking at large, broadly defined data sets.

Dimensions' article-level classification uses two different classifications, the Fields of Research (FOR) and the UN Sustainable Development Goals. The FOR uses a three-level hierarchy; however, it is not clear if the algorithm uses citation relationships or text-based analysis to determine the documents' category assignments. The UN Sustainable Development Goals does not have a hierarchy, and documents are matched to categories based on a combination of machine learning and keyword searches.

SOURCE TITLES

Source titles include any of the publication titles included in the data set. Normally, these include journal titles, conference proceedings, and so on. This is not only a standard entity type but also an essential entity type as most of the bibliometric analyses were born out of publication source analyses. One of the most important limitations to these systems is the titles' coverage of the data source. Each pulls in bibliographic data from different sources, as discussed in chapter 1, although there is some overlap with open data sources such as Medline, Crossref, and so on. Therefore, it is essential to report the source of the data set in any analysis for transparency on the limitations of the resulting bibliometric analysis.

The ability to look at the source titles by selected indicators allows evaluation for the purposes of collection development and publication decisions. All these systems allow source titles to be analyzed by subject area, output counts, citations, and other standard metrics. However, it appears that only Lens.org links patent citations to publication titles in its available analytics.

Table 2.3: Content type comparison between the major commercial bibliometric tools. Notes are included on funding/grant award amounts, patents, and news media.

	InCites	*SciVal*	*Dimensions*	*Lens.org*
Content Types				
Scholarly publications (articles, books, conference proceedings, etc.)	yes	yes	yes	yes
Usage data	no	yes	no	no
Funding/grant award amount[a]	no	yes	yes	no
Clinical trials	no	no	yes	no
Patents	yes[b]	yes[c]	yes[d]	yes[e]
Data sets	no	no	yes	yes
Policy documents	no	no	yes	no
News media	no	yes[f]	no	no

a. Although funding body data is captured at the article level in all these tools, only SciVal and Dimensions ingest data directly from the funding bodies to enable reporting on awarded amounts at the institution and country level.
b. Data sources: draws data from Clarivate-owned Derwent Innovation, including 59 patent office sources.
c. Data sources: European Patent Office, UK Intellectual Property Office, Japan Patent Office, US Patent and Trademark Office, World Intellectual Property Organization.
d. Data sources: European Patent Office, UK Intellectual Property Office, Japan Patent Office, US Patent and Trademark Office, World Intellectual Property Organization, German Patent and Trademark Office, Canadian Intellectual Property Office, Intellectual Property India, Intellectual Property Office (UK), National Industrial Property Institute, Intellectual Property Department (Hong Kong), Russian Patent Office.
e. Data sources: European Patent Office, US Patent and Trademark Office, Intellectual Property Australia, World Intellectual Property Organization.
f. New media sourced from LexisNexis Metabase (2013–) into Elsevier's Newsflo system.

FUNDING BODIES

Funding bodies data is derived from the bibliographic metadata, not through the funding source, and does not capture details about the amount awarded within the grants. Therefore, although funding bodies are an entity type that can be analyzed within all these bibliometric tools, only SciVal and Dimensions report the actual award amounts. This is because they ingest funding amounts directly from the funding bodies and connect this data with institution- and country-level data. See the document type "Funding/grant award amount" in table 2.3 for more details.

Funding body data and award data are of high interest to bibliometric tools as their user base expands to include research administrators and other university units interested in having a clearer link between awarded grants and their research impacts.

PATENTS

Patents are a challenging entity type to capture in bibliometric tools because they do not adhere to the common bibliographic standards, making them difficult to connect to research outputs. The most comprehensive patent data is available in Dimensions and Lens.org. This patent data makes these tools stand out from the other two systems as they provide detailed patent data that is separated out from the data available for scholarly publications and can therefore be searched and analyzed using unique fields such as inventor, owner, legal status, and so on.

Although InCites and SciVal do not have patents as a separate entity type for analysis, there are patent metrics, such as patent-citation counts, available in these tools. This is accomplished by linking documents to each other through shared metadata fields. For example, a research publication might be cited in a patent article, allowing these two separate documents to be connected. This would mean the article has received a patent citation. Analysis of articles containing patent citations can be accomplished in any of these tools by creating a publication set from either search results or the presented patent-citations bibliometric indicator in the system. Once the unique publication data set is created, any of the analysis options that are standard within the bibliometric tool (i.e., research areas, years, collaborations, and publication lists) are possible.

Bibliometric and Network Analysis Tools

Bibliometric and network analysis tools are likely to be considered the premier type of bibliometric analysis tools within the bibliometric (and scientometric) research community and with advanced level practitioners. They tend to be used for more in-depth bibliometric studies due to the additional technical training or knowledge that is required to use these applications. However, there is a spectrum among these tools—from the more user-friendly (VOSviewer, VOSviewer Online, Biblioshiny, CitNetExplorer) that do not require any programming knowledge, significant data cleaning, or training to the more advanced (Bibliometrix, CiteSpace, SciTools, SciMAT) that do require more advanced training and knowledge. This should not discourage the keen practitioner or scholar;

Table 2.4: List of the most commonly used bibliometric and network analysis tools

Product Name	Description	Owner/Developer
CitNetExplorer https://www.citnetexplorer.nl/	Bibliometric analysis with citation graphs (download)	CWTS Leiden University
Bibliometrix https://www.bibliometrix.org/home/	Bibliometric and network analysis package (download)	Bibliometrix
Biblioshiny https://www.bibliometrix.org/home/index.php/layout/biblioshiny	Bibliometric and network analysis software (online, no coding)	Bibliometrix
VOSviewer https://www.vosviewer.com/	Bibliometric network analysis software (download)	CWTS Leiden University
CiteSpace https://citespace.podia.com/	Bibliometric network analysis software (download)	Chaomei Chen
VOSviewer Online https://app.vosviewer.com/	Bibliometric network analysis software (online)	CWTS Leiden University
Sci2 Tool https://github.com/CIShell/sci2/releases/tag/v1.3.0	Bibliometric and network analysis software (download)	Indiana University and National Science Foundation
SciMAT https://sci2s.ugr.es/scimat/	Bibliometric and network analysis software (download)	University of Granada

Note: HistCite is not included here because it does not appear to be maintained. Although it can still be downloaded and used, more advanced and user-friendly options are available.

there are certainly many cases of bibliometric practitioners and other nonexperts who have been able to upskill very quickly. However, with this large spectrum of tools, nontechnical practitioners of bibliometrics can gain fairly quick entry to this class of tools, and scientometric researchers will find the advanced and statistical functions within these tools advantageous to their in-depth research questions.

Table 2.4 lists the main network analysis tools. There are some variances in the details of their functionality; however, all these tools have three main workflows:

1. **Data ingest and entity management:** The data set is imported using standard data files that have been created and exported from a bibliographic source. As discussed in chapter 1, these sources are typically Web of Science, Scopus, Crossref, and OpenAlex. The data files can be saved as raw files within the systems to allow for multiple analyses depending on the entity type. For example, authors, documents, countries, and institutions can be used as the main entity types.
2. **Select network analysis options:** There are several types of network analysis that these tools offer, including co-citation analysis (using authors, documents, or journals), keyword co-occurrence analysis, bibliographic coupling, coauthorship, and citations.
3. **Data visualization:** The mapping visualizations and access to the underlying data are the shining features of these tools. They apply advanced natural language processing, mapping algorithms, and in-system data cleaning (via thesauri, etc.) that make the clustering and mapping visualizations deceptively easy to create.

What Is Bibliometric Network Analysis?

A bibliometric network is a visual representation of the relationship between bibliographic objects. In technical terms the objects are *nodes* and the relationships are *edges*, represented by lines, and they can indicate not just the existence of a relationship but its strength as well. Bibliographic nodes, which are also referred to as entities in this report, are publications, journals, researchers, or keywords. The relationships (edges) studied can include co-citations (with authors or documents), keyword co-occurrence, bibliographic coupling, coauthorship, and citations, as were also previously mentioned (van Eck and Waltman 2014; Chen 2017).

CO-CITATION ANALYSIS

Co-citation analysis allows documents to be analyzed based on shared citing documents. This means that two documents will be linked because they have both been cited by the same document. The strength of the relationship between two documents is determined by the number of shared co-citing documents. Analysis of co-citations can be done with documents, authors, or journals as this main entity (or node).

BIBLIOGRAPHIC COUPLING

Bibliographic coupling also allows documents to be analyzed based on having shared citations in their

Table 2.5: List of additional bibliometric and network analysis tools currently available that are not detailed in table 2.4, as they are either not frequently updated, regionally specific, or less well-known

Product Name	Description	Owner/Developer
BibExcel https://homepage.univie.ac.at/juan.gorraiz/bibexcel/	Bibliometric analysis package (Excel)	Olle Persson
Scimeter https://scimeter.org/	Bibliometric analysis software (limited, arXiv.org source)	Frankfurt Institute for Advanced Studies
ScientoPy https://www.scientopy.com/en/	Bibliometric analysis software (limited, graphs)	University of Cauca
CRExplorer https://andreas-thor.github.io/cre/	Bibliometric analysis software (limited, historical citation analysis)	Andreas Thor, University of Applied Sciences for Telecommunications, Leipzig
RPYS i/o http://www.leydesdorff.net/comins/rpys/index.html	Bibliometric analysis software (limited, historical citation analysis)	Virginia Tech Applied Research Corporation
VIPER https://www.openaire.eu/viper-the-visual-project-explorer	Bibliometric and network analysis software (limited use)	OpenAire
Metaknowledge https://uwaterloo.ca/networks-lab/projects/metaknowledge	Bibliometric and network analysis software (limited use)	University of Waterloo
Scholarometer https://scholarometer.indiana.edu/	Bibliometric network analysis software (limited)	Center for Complex Networks and Systems Research, Indiana University Bloomington
Social Science Research Network (SSRN) https://www.ssrn.com/index.cfm/en/	Bibliometric ranking data	Elsevier (bought from Social Science Electronic Publishing Inc. in 2016)
Scimago Viz Tools https://www.scimagojr.com/viztools.php	Bibliometric visualization tool	Scimago

reference lists. This means that two documents will be linked because they have both cited the same document. Again, the strength of the relation between two documents is determined by the number of similar citations within their reference lists. Bibliographic coupling can be done with documents, journals, authors, institutions, or countries as the main entity (or node).

KEYWORD CO-OCCURRENCE

Keyword co-occurrence allows documents to be analyzed based on having shared keywords within their text, usually the title, abstracts, and listed indexed and author keywords. The strength of the relationship between two documents is determined by the number of shared keywords. The nodes presented in these analyses are the keywords themselves, and this approach is a popular analysis for looking at the clustering of research domains within a group of documents.

COAUTHORSHIP ANALYSIS

Coauthorship analysis allows documents to be analyzed based on having shared authors. Authors who frequently publish together therefore have stronger relationships. Coauthorship analysis can be done with individual authors, institutions, or countries. The relationships for institutions and countries are determined by the authorship; however, at these levels the data is aggregated to the institution or country level based on the affiliation information in the document's bibliographic information.

CITATION ANALYSIS

Citation analysis is one of the simplest analyses. It allows documents to be analyzed based on the number of times they cite one another. Although simple, this analysis tends to yield fewer relationships because of the direct relatedness needed between the documents (van Eck and Waltman 2014).

Other Bibliometric and Network Analysis Tools

Table 2.5 lists other bibliometric and network analysis tools that are currently available. These tools appear to be more limited in their scope of features, functionality, or adoption; however, they are still worthy of mention as many have been developed by researchers and research institutes that study and perform bibliometric network analyses as their field of research.

An Explosion of Discovery Tools

The landscape of bibliometric tools can be very confusing. This confusion is exacerbated by the recent

Table 2.6: Discovery tools using a variety of network analysis functions to aid users in research discovery

Product Name	Type of Tool
CiteSeerX https://citeseerx.ist.psu.edu/	Discovery
Scinapse https://www.scinapse.io/	Discovery and analytic consultancy
Open Research Knowledge Graph https://orkg.org/	Discovery and workflow management
Scilit https://app.scilit.net/	Discovery with analytical views
Google Scholar https://scholar.google.ca/	Discovery with analytics views
Academia.edu https://www.academia.edu/	Discovery with analytics views and author level impact
ResearchRabbit https://www.researchrabbit.ai/	Discovery with citation graphs
Connected Papers https://www.connectedpapers.com/	Discovery with citation graphs
Litmaps https://app.litmaps.com/	Discovery with citation graphs
Inciteful https://inciteful.xyz/	Discovery with citation graphs
PURE suggest https://fabian-beck.github.io/pure-suggest/	Discovery with citation graphs
CitationGecko https://www.citationgecko.com/	Discovery with citation graphs (no longer maintained)
CoCites https://www.cocites.com/	Discovery with citation graphs
Scite https://scite.ai/	Discovery with citation influence/contextualization and citation graphs
Semantic Scholar https://www.semanticscholar.org/	Discovery with citation maps and citation influence/contextualization
Open Knowledge Maps https://openknowledgemaps.org/	Discovery with keyword graphs
Yewno https://www.yewno.com/	Discovery with knowledge graphs
Iris.ai https://iris.ai/	Discovery with workflow management
Elicit https://elicit.org/	Discovery with workflow management
Scholarcy https://www.scholarcy.com/	Discovery with workflow management

explosion in the development of discovery tools that use bibliometric networks analysis as a method of research discovery (table 2.6). Many of these tools use a single seed or set of seed documents to present relevant research to the user. The idea is that the papers linked to these seed papers are highly relevant based on the co-citation, bibliographic coupling, or similar network mapping that they employ. The user can then navigate through the presented papers and select those that are of interest. Although these tools are very fascinating and are gaining popularity within the academic community, they are not useful for bibliometric analysis as the data is not presented for analysis but rather discovery, and therefore the systems do not often have adequate explanatory documentation for the user to understand and report the details of the methodology of analysis. Despite this, there is interest and evidence of these systems being used to supplement traditional search methods for systematic reviews, and they may become a standard method for reviews in the future.

There are also discovery tools that are beginning to contextualize the types of citations that are contained within research papers. They not only identify the existence of a citation but also make some assessment of the value of the citation to the original document. Scite and Semantic Scholar are two such research discovery tools that approach this challenge in different ways. Scite reports citations as "supporting," simply "mentioning," or "contrasting." Semantic Scholar reports the intent of the citations as either "background," "methods," or "results" and also indicates the velocity, acceleration, and whether the paper has influential citations. All these added features rely on the full text of the papers being available. Therefore, the data sources may be more limited than with the traditional bibliographic databases; this fact is a reason why these and similar systems are advocates for open access publishing options.

Choosing Tools That Are Right for Your Organization

Sugimoto and Larivière (2018) outline five key issues when considering bibliometric analysis that can also be applied when considering the tools that best fit the job at hand.

1. **Time:** Consider the constraints of publication and citation windows. First, quality scholarship does not happen quickly, and the accumulation of citations is known to take several years. Second, there are disciplinary differences in the life cycle of scholarship that need to be acknowledged and actions taken to mitigate unfair comparisons in any analysis.
2. **Data quality:** Data quality can be improved through data cleaning. Several bibliometric tools use unique identifiers for author names, institution names, funding bodies, and so on. Work can be done prior to analysis to improve the accuracy of this data. For example, ORCID is being adopted by many institutions as an open and nonproprietary author identification number. If the system uses ORCID, these numbers can be used to verify the identity of authors prior to analysis.
3. **Normalization:** When planning to benchmark across disciplines, the use of metrics that use normalization can help mitigate differences in data set size, disciplinary norms, and publication types.
4. **Coverage:** All bibliometrics data sets have limitations and constraints. It is important to be familiar with the content coverage of a data set. Research areas that are not adequately covered in the data set cannot be adequately analyzed.
5. **Alignment:** No analysis should be done without a clear objective. These are usually identified through clear research questions and the analysis, and selected metrics then must align with these questions. As discussed in chapter 1, the *SCOPE Framework* developed by the INORMS Research Evaluation Group (2020) provides more guidance on how bibliometric practitioners can align their analyses with the goal of using metrics responsibly.

When selecting bibliometric tools, consider these five factors. The tools you or your institution chooses will depend on your usage and what data you wish to analyze.

References

Chen, Chaomei. 2017. "Science Mapping: A Systematic Review of the Literature." *Journal of Data and Information Science* 2, no. 2: 1–40. https://doi.org/10.1515/jdis-2017-0006.

Guerrero-Bote, Vicente P., Zaida Chinchilla-Rodríguez, Abraham Mendoza, and Félix de Moya-Anegón. 2021. "Comparative Analysis of the Bibliographic Data Sources Dimensions and Scopus: An Approach at the Country and Institutional Levels." *Frontiers in Research Metrics and Analytics* 5 (January). https://doi.org/10.3389/frma.2020.593494.

INORMS Research Evaluation Group. 2020. *The SCOPE Framework: A Five-Stage Process for Evaluating Research Responsibly.* INORMS Research Evaluation Group. https://inorms.net/wp-content/uploads/2021/11/21655-scope-guide-v9-1636013361_cc-by.pdf.

Sugimoto, Cassidy R, and Vincent Larivière. 2018. *Measuring Research: What Everyone Needs to Know.* Oxford: Oxford University Press.

van Eck, Nees Jan, and Ludo Waltman. 2014. "Visualizing Bibliometric Networks." In *Measuring Scholarly Impact: Methods and Practice*, edited by Ying Ding, Ronald Rousseau, and Dietmar Wolfram, 285–320. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-10377-8_13.