Chapter 5
Applying Bibliometrics to Examine Research Output and Highlight Collaboration

Nandita S. Mani
https://orcid.org/0000-0002-0955-1066
University of North Carolina at Chapel Hill, USA

Michelle Cawley
University of North Carolina at Chapel Hill, USA

Adam Dodd
University of North Carolina at Chapel Hill, USA

Barrie E. Hayes
https://orcid.org/0000-0002-3547-1553
University of North Carolina at Chapel Hill, USA

ABSTRACT

Throughout the pandemic, research contributing to discoveries associated with COVID-19 has grown considerably. To help increase visibility of this integral body of research and illustrate the extensive organizational collaborations that help move this research forward, a library team utilized data science techniques to analyze COVID-19 research output. These analyses help to demonstrate how libraries can integrate data science expertise to showcase an institutions depth of research engagement and facilitate institutional, national, and global collaboration.

DOI: 10.4018/978-1-7998-9702-6.ch005

This chapter published as an Open Access Chapter distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

INTRODUCTION

Academic libraries have myriad opportunities to demonstrate value and contribute to the research mission of the universities of which they are a part. Bibliometrics, a component of data science (DS), offers libraries an opportunity to demonstrate expertise using DS skills, proactively identify ways in which to partner within and beyond the university, and further illustrate the breadth and depth of research. Globally, the use of bibliometrics has grown and been applied to a variety of contexts. Bibliometrics has been used in multiple ways by examining citation data over time to illustrate individual, organizational and institutional depth and breadth of collaboration, topics being investigated, areas of expertise, and identification of commonly used journals selected by authors to disseminate scholarly information in particular subject domains. Using a variety of tools from the areas of data analysis and visualization, both subscription-based resources and freely available tools are available to support bibliometrics-based inquiry. As the field of DS has evolved, academic libraries have responded by building capacity in their institutions and professional organizations. Libraries have identified new skills required to effectively address research needs which have resulted in sharing of bibliometrics toolkits, tutorials, and workshops. To illustrate how bibliometrics can be applied and assist academic libraries while remaining central to the DS ecosystem, a case study using a variety of tools and techniques is described along with opportunities academic libraries may consider when establishing or engaging with bibliometrics-based projects.

BACKGROUND

The University of North Carolina at Chapel Hill (UNC-CH) received over one billion dollars in annual research awards in 2021, including over 740 million dollars from federal funding sources. UNC Health Affairs Principal Investigators administered over 800 million dollars (75%) of UNC's total 2021 research awards, where grants administered by the School of Medicine and Gillings School of Global Public Health (research.unc.edu) were prominent. The UNC School of Medicine is in the 94th percentile among public medical schools for federal grant funding (https://www.med.unc.edu/research/), and the Gillings School of Global Public Health is ranked first among public schools of public health for NIH funding (https://sph.unc.edu/research/). Increasing competition for grant funding and emphasis on demonstrating collaboration among funded researchers provided an opportunity for libraries to envision how skills in DS and data visualization could be leveraged to promote and showcase research outputs and collaborative networks of research teams (Mani et al., 2021).

Bibliometrics is the use of quantitative and statistical analysis to investigate different facets of publication data and is particularly useful for analyzing publication sets too large to effectively review manually (Sugimoto & Lariviere, 2018). Bibliometrics is the scholarly, product-focused component of scientometrics, a "metascience" (Sugimoto & Lariviere, 2018, p. 10) that applies bibliometrics and other quantitative methods to the analysis and measurement of research. Bibliometric indicators (*metrics*) such as citation counts have long been interpreted as indicators of research impact or influence but not as direct measures of impact (Waltman & Noyons, 2018). Libraries have historically applied bibliometric methods to inform literature retrieval and collection development (Sugimoto & Lariviere, 2018). More recently, libraries are applying these methods and the latest bibliometric analysis and visualization tools (Waltman & Noyons, 2018) to assist their institutions in assessing research output, impact, and the extent of research collaborations (Gutzman et al., 2018). Bibliometric tools, methods, and indicators continue to evolve as scholarly literature, particularly research articles, grows exponentially (Moral-Muñoz et al., 2020), and research institutions, funders, and an increasing diversity of other users strive to quantitatively describe and assess scientific output, impact, and collaboration across disciplines.

Bibliographic data from citation databases such as Web of Science (WOS) and Scopus are typically used for these analyses. Google Scholar is also frequently used and has the advantages of being freely available and offering more extensive literature coverage than WOS (Clarivate Analytics) and Scopus (Elsevier). However, as Waltman and Noyons (2018) summarized, Google Scholar "lacks transparency, suffers from data quality problems, and is very difficult to use for large-scale analyses" (p. 8). Other discipline-specific bibliometric databases with rich topic metadata like PubMed (MEDLINE) are also used for a more limited range of analyses due to missing citation information and incomplete affiliation information before 2015 (Sugimoto & Lariviere, 2018). Geographically targeted bibliometric analyses may use regionally-focused databases such as the LILACS database of Latin American and Caribbean Health Science Literature or the Chinese Science Citation Database, grant, or patent data (Xing et al., 2019). Other text-based information such as course syllabi can also be described and evaluated using the methods and tools described in this chapter.

The two most widely used literature databases for bibliometrics analysis have traditionally been WOS and Scopus. Sugimoto & Lariviere (2018) discussed the characteristics and differences among these databases, Google Scholar, and others. Newer tools such as Digital Science's Dimensions product have become available recently. Since its launch in 2018, Dimensions, has been considered an alternative to WOS and Scopus and is included in recent bibliometric database content comparisons (Singh et al., 2021; Guerrero-Bote et al., 2021) and reviews of databases with features supporting bibliometrics analyses (Moral-Muñoz et al., 2020; Thelwall, 2018). Each of these databases is useful for obtaining indicators on research output, collaboration, and impact or influence. In addition to information on output, collaboration, funding agencies, and topics available in publication metadata (e.g., title, author, affiliation, grant support or funding details, topic keywords, and abstract), these databases provide citation counts and associated lists of citing references. Citation count, the total number of times a published paper is cited by other papers, is an important measure of a publication's relational connection to other publications and is typically interpreted and used as an indicator of the research output's influence and impact on the scientific community (Sugimoto & Lariviere, 2018; Waltman & Noyons, 2018). Tools such as VOSviewer (https://www.vosviewer.com) enable analyses of co-citation (when papers are cited by the same document), and bibliometric coupling (when papers share the same references) (https://www.vosviewer.com) can enable examination of additional relational connections (e.g., patterns) among networks of publications. Other comparative citation indicators may also be present in citation databases. These include indicators such as field-weighted citation index, "highly cited paper" designation and field citation ratio (Scopus, WOS, Dimensions), *h*-index (Scopus, WOS), and altmetric attention data (Scopus, Dimensions). Scopus and WOS citation databases both require a subscription to use. Dimensions also has subscription components; however, publication information in this system is freely accessible.

Products such as Dimensions offer a new frontier of research databases that has joined the bibliometric database landscape. Beyond traditional document types like articles, books, and conference papers, Dimensions uses unique identifiers and machine learning to generate links between these traditional documents and other associated research output such as preprints, datasets, supporting grants, patents, clinical trials, policy documents, and citation-based and altmetric attention indicators mentioned above (Moral-Muñoz et al., 2020; Hook et al., 2021). As Hook et al. (2021) noted, the expanded Dimensions research content, paired with data that are enhanced with automated subject categorizations and person and institution name disambiguation, can be updated daily and are accessible via an application programming interface (API), enabling analysis of emerging and dynamic events, like COVID-19, that affect research. These system design features lay the foundation for facilitating real-time bibliometrics capabilities in the future.

Bibliometric analysis results in the form of reports and visualizations generally tend to have either a descriptive or an evaluative purpose or approach. Those results that are descriptive in nature focus on articulating the status of an area of research at a point in time or identifying emerging areas of focus within a research domain. Results with an evaluative purpose are typically used to help assess research performance (i.e., output, impact, and collaboration patterns) of some research actor such as an investigator's productivity, an institutional unit's research impact, or the research return on investment for a funding mechanism (Cabezas-Clavijo & Torres-Salinas, 2021; Moral-Muñoz, et al., 2020). These descriptive- or evaluation-oriented bibliometric products may be useful to showcase the strengths of a research team or institution as part of grant applications, to support re-accreditation efforts, to recruit and retain faculty, and to identify experts, mentors, or faculty advisors, among other uses.

Bibliometrics-focused units and services that produce these products are often based in the libraries and can be found at academic institutions across the globe. Some bibliometric units and services may also support their institutions through bibliometric information resources such as online guides, educational consultations, and workshops. Examples of bibliometric service units around the world include the Department of Bibliometrics and Publication Strategies at the University of Vienna (Austria), the University of New South Wales (Australia; Drummond & Wartho, 2009), the Technical University of Munich (Germany), University of Waterloo (Canada), Oxford University Bodleian Libraries (United Kingdom), and Lund University (Sweden). The Center for Science and Technology Studies (CWTS) at Leiden University (The Netherlands) is unique in that it provides bibliometric assessment reports, including the CWTS Leiden Ranking (https://www.leidenranking.com/) for institutions around the world through its company, CWTS BV (Cabezas-Clavijo & Torres-Salinas, 2021).

This chapter uses examples drawn from a bibliometric analysis of UNC-CH authors' COVID-19 published research. The authors conducted the analyses as part of a larger and ongoing effort to demonstrate overall research impact, highlight collaborations occurring with national and international organizations, track authors who are conducting research and convey the wide range of research topics covered by publications. Evaluating COVID-19 research at an institutional level was an intentional choice by the Health Sciences Library (HSL) Director as it provided a way for the library to showcase library expertise around bibliometrics, foster future collaborations with campus partners, and provide useful data and visualizations for campus administrators. Given COVID-19 research is high profile and within a clearly defined time period (i.e., 2020–present), it serves these purposes exceptionally well.

APPROACH

When conducting bibliometric analysis, it is important to identify the context and purpose of the investigation. If bibliometrics is used to support original research, it is imperative to consider the specific research question(s) being investigated and ensure appropriate measures are examined. If the desired outcome is to use data to tell a story, one should consider what questions may resonate most strongly with the audience in question and then select appropriate data points for analysis.

The types of research questions formulated will depend on the intersection of the research context and available data. While data elements vary across data sources, common data fields available in databases such as PubMed, Scopus, and Web of Science include:

- Publication counts, collaboration patterns, *h*-index, affiliation (i.e., the name of an institution or organization), granting agencies, or country affiliation
- Number of publications or impact factor for specific journals or disciplines
- Author/index keywords to evaluate topics using term frequency, co-occurrence of terms, and change in keywords over time

Possible research questions that can be answered using citation data and bibliometrics may be descriptive or evaluative in purpose (Table 1).

Table 1. Sample research questions with associated purpose, potential stakeholders, and sample analyses that can be used to answer each research question. Purpose may be descriptive or evaluative, as indicated.

Question and Purpose	Stakeholders	Types of Analyses and Example(s)
Which journals might be a good fit for my research? **Purpose**: Identify journals publishing similar research (descriptive).	Researchers Faculty Staff Graduate students	Journal productivity analysis; journal impact analysis See Zyoud and Zyoud (2021)
Who are potential collaborators on my campus or at other institutions? **Purpose**: Review faculty colleague research outputs and impact for best alignment with one's own research focus (descriptive).	Researchers, including faculty and graduate students	Author and topic analysis; bibliographic coupling or co-citation analysis See Yu and Hayes (2018)
How are researchers on my campus collaborating with other institutions? **Purpose**: Understand collaboration patterns (descriptive); evaluate researchers based on the breadth and depth of their collaborations (evaluative).	Administrators Faculty Grant agencies	Collaboration analysis See Wu et al. (2021)
What topics are researchers publishing on? **Purpose**: Understand topics of research focus (descriptive); assess whether researcher topic foci align with department and institution research priorities (evaluative).	Administrators Faculty	Topic analysis See Saheb et al. (2021)
In what curricular areas do faculty have expertise? **Purpose**: Review faculty research outputs and impact for best alignment with PhD candidate research focus (descriptive); Identify faculty mentors for dissertation committee (descriptive).	Curriculum committees Graduate students	Author and topic analysis See Mani et al. (2022)
What is the state of knowledge for a particular research domain? **Purpose**: Identify research topics covered by a body of literature (descriptive).	Researchers Faculty Staff Graduate students	Author and topic analysis; journal productivity analysis; collaboration analysis See Sweileh (2020) and Kahwa et al. (2022)
What is the status of research and development including patent status for a particular technology or process? **Purpose**: Understand the status of a research area where patents would be an outcome (descriptive).	Researchers, faculty Staff Graduate students Administrators Inventors Entrepreneurs	Patent analysis See Xing et al. (2019)

Continued on following page

Table 1. Continued

Question and Purpose	Stakeholders	Types of Analyses and Example(s)
What effect do NIH early career training awards have on trainees' future research output, impact, and grant funding? **Purpose**: Evaluate the effect of NIH training award on trainee post-training publication and grant activity (i.e., performance; evaluative).	Researchers Faculty Staff Graduate students Administrators Clinical and Translational Science Awards (CTSA) programs	Grant analysis See Qua and Pelfrey (2020)

After research questions have been formulated, next steps include:

- Identifying appropriate data sources (e.g., Scopus, PubMed, and Web of Science)
- Constructing a comprehensive search strategy (consultation with domain experts and librarians or other information professionals with bibliographic database search expertise)
- Retrieving and cleaning data (often the most time-consuming step)
- Performing analyses and producing visualizations (using tools such as Excel, Tableau, and VOSviewer)
- Reviewing initial analyses and re-examining research questions as necessary

These steps are often iterative, and the project scope, goals, types of analyses, and deliverables will vary based on time constraints, the skillset of team members, and the output or deliverables desired.

Some databases have built-in aggregation and analysis tools that may serve to provide summary information for a given result set, such as the "Analyze Search Results" feature in Scopus. With this interface, data can be quickly identified, such as the top number of publications by year, authors, organizations, and countries or territories. Data sources that allow raw data extraction can be viewed and analyzed using common tools such as Microsoft Excel or Google Sheets to perform a similar base analysis.

For more in-depth investigation, such as network analysis and text analysis, additional skills and tools are required. For example, the VOSviewer tool (https://www.vosviewer.com), created and supported by Leiden University, can assist with building and visualizing bibliometric networks (van Eck & Waltman, 2009). Using such tools requires a deeper understanding of bibliographic data and its inherent structure and can provide a rich and visually engaging insight into data.

Prior to performing analyses, regardless of which tools are used, it is essential to develop a comprehensive and robust search strategy to identify an appropriate result set for data screening and cleaning. The value of analyses is directly correlated with the quality of input data. Constructing a strong search query is critical to retrieving a relevant data set and may decrease the amount of screening in some instances. Inevitably, data sets will contain false positives and may be incomplete or require normalization. Data screening and cleaning steps in the process are time-consuming but essential to effectively answer research questions.

The size and complexity of projects will determine the size of the team needed to conduct the analysis and what skills and expertise are required. For less complex projects (i.e., projects with a narrow scope and straightforward research questions), an individual or small team may be sufficient. Large, complex projects may require the formulation of a team. No matter the size and complexity of a project, general skillsets are necessary, such as: expert searching and knowledge of data sources, proficiency in data analysis and visualization techniques, and some expertise with statistics. Data visualization skills are a core component of the analysis phase to identify visual patterns in data and to understand and calculate some measures. Data must be prepared for analysis by performing data transformations such as merging, cleaning, and reformatting. Data management skills, including organizing, documenting, formatting, and versioning data files, are crucial for efficiency throughout a project. Creation, organization, maintenance, and preservation of detailed project documentation such as client objectives, desired deliverables, data description, and handling processes, among other aspects, are important as well, particularly for larger projects.

Bibliometric professionals and others constructing and interpreting bibliometric indicators must also understand the strengths and weaknesses of available citation databases and how various bibliometric metrics are calculated or created. This knowledge is best accompanied by an awareness of principles of responsible metrics and various international recommendations and calls-to-action pressing for ethical and responsible use of citation-based metrics as part of research assessment. Cox et al. (2019) describe a competency model including basic to advanced skills important for bibliometrics analysis work. The website responsible metrics (available from: https://responsiblemetrics.org/) provides a forum for debating the responsible use of metrics in higher education and research. Szomszor et al. (2021) discussed issues with the interpretation of bibliometric information, particularly when it is used for evaluative research management. Larger projects may require additional skills, such as project management, application development, and domain expertise. For a full list of competencies, tools, resources, and skills, see Mani et al. (2021).

OTHER DATA TYPES

While the focus of this chapter is on analyzing publication data, other data, including patent, grant, curricula, and social media or altmetric data, can be analyzed using similar tools and approaches. Investigating other artifacts, products, and outputs related to the research process can provide additional insight into the research lifecycle (Xing et al., 2019). In general, other data types are less standardized, more distributed, and disparate than bibliographic data, so they may require more work to collect and often present unique challenges. Given the nature of these other data types, librarians' skills around expert searching are crucial to discovering a comprehensive dataset for analysis. An overview of some of the considerations for these other data types is discussed below.

Patent Data

Patent data can be analyzed to understand the design intent of research output as patents filed with governing bodies describe and index innovative systems, devices, and processes. Patent data can be analyzed to identify emerging technology trends and provide research insights, including informing new research directions (Klongthong et al., 2021; Saheb & Saheb, 2020).

Patent information is often tied directly to published research that describes the context, purpose, and intent of the patented item or process. Patent data are open access and accessible through public patent databases, including through the U.S. Patent and Trademark Office for U.S. patents (available from https://www.uspto.gov/). Subscription data sources, such as Derwent World Patents Index and Dimensions, provide structured data to analyze. Data are often structured in a similar manner to citation data allowing similar tools and approaches to be used for patent analysis.

Jürgens & Herrero-Solana (2016) provided a thorough summary of how patent publications differ from scientific publications in terms of content, access, and indexing. One limitation of patent data is around timeliness, as patent data are generally not available until 18 months after a patent is filed. Further, Jürgens & Herrero-Solana (2016) noted that not all innovations are patented or patentable; therefore, would not be found in a patent data source.

Grant Data

Grant data can be used to analyze financial awards from government agencies and non-governmental organizations and include information on principal investigator(s) (PIs), organization(s) receiving the awards, amount of funding awarded, and the purpose of the award. The tools and approaches are similar to those used with bibliographic data. Government and non-government entities often provide a structured interface for searching and retrieving grant data. The Federal RePORTER database and the USASpending.gov website provide access to grant data, as does the Dimensions platform via subscription.

Linking grant data with associated research output offers another perspective on the research lifecycle (Druss & Marcus, 2005). Analyzing grant data can assist in answering questions around demographics of award recipients (Qua & Pelfrey, 2021; Pagel & Hudetz, 2015), funding priorities of granting agencies (Lyubarova et al., 2009; Coppersmith et al. 2018), and the success rate of research grant applications. To get an accurate picture from grant data, pulling information from multiple sources may be necessary. Data from freely available sources may not be structured consistently or may be incomplete, which will require more time to prepare data for analysis.

Curricular Data

Curricular data may include course descriptions, learning objectives, required readings, and syllabi. Textual analysis such as topic analysis can be used on these data to evaluate curricular areas of focus, expertise, and scope across a school, unit, or division. By using text from course descriptions, syllabi, and faculty publications, bibliometric analysis combined with curriculum mapping and thematic domain analysis can help illustrate what learning objectives, skills, and topics are covered and by which courses and instructors. In the case of the analysis conducted to support curricular transformation for the UNC Eshelman School of Pharmacy (ESOP), analysis assisted in the discovery of faculty mentors and advisors and helped illuminate the depth of expertise in particular curricular domains (Mani et al., 2022). Curricular data typically come directly from school administrators and faculty. Open Syllabus is a non-profit research organization that collects syllabi and uses machine learning to extract metadata from the documents (available from https://opensyllabus.org/).

Administrators can use this type of information in multiple ways. For example, analyzing curricular data can be helpful to update curricula; maximize curricular efficiency across a school, division, or unit; identify faculty expertise in a particular subject domain; ascertain mentors and advisors for students (e.g., dissertation advisors and committee members); and discover faculty expertise that could support grant-related inquiries.

Altmetrics Data

Alternative metrics or *altmetrics* are quantitative and qualitative data that include discussion on social media, such as mentions on blogs, Twitter, and in news and policy sources. Other example sources of altmetric data include Wikipedia, LinkedIn, and YouTube. Social media data can provide insights into the communication and discussion of research activity on social media platforms. The purpose of altmetrics is to understand the attention a resource is generating online beyond the story told by citation metrics. For example, the altmetric attention score is a weighted statistic designed to reflect all of the attention a particular piece of research has received online.

Tools for analyzing altmetric data include textual analysis approaches such as clustering, sentiment analysis, and social network analysis. Several platforms track and aggregate these data, including Altmetric (available from https://www.altmetric.com/), Impactstory (available from https://profiles.impactstory.org), and Plum Analytics (available from https://plumanalytics.com/). Some citation databases incorporate altmetric data into publication records, including Scopus, which incorporates PlumX Metrics data from Plum Analytics and Dimensions, which incorporates data from Altmetric. The availability, structure, and coverage (in terms of social media platforms) vary by altmetric data source.

One of the major limitations of these data is understanding how to interpret them and use them responsibly. The non-profit Metrics Toolkit (available from http://www.metrics-toolkit.org/) provides a comprehensive summary with appropriate uses and limitations for a wide range of metrics including altmetrics.

APPLICATIONS

Visualizations and descriptions in this section were selected to provide a broad array of examples that could also be used for other datasets. As noted above, the dataset for analyses was derived from bibliographic data for published COVID-19 research at the author's institution. The analyses may be valuable to administrators, faculty, staff, and students to demonstrate research output and impact, provide insight into research topics and collaboration networks, and identify potential collaborators or faculty mentors, among other uses (see Table 1).

Journal Analysis

Journal analysis is performed on a set of bibliographic records to identify the journals in which researchers most frequently publish. Output can be as simple as a list of top journals (based on the number of publications on a topic or by an author) or can be combined with other data as in a three-field plot (see below). Journal analysis can be useful to demonstrate research impact. For example, administrators can illustrate research impact and reach by determining how many articles are being published in top-tier, high-impact journals. This analysis can also help researchers identify possible journals that would be a good fit for publishing their work. Further, journal analysis can provide insight into the variety of research being published. For example, at UNC-CH, researchers have published on many aspects of COVID-19, including vaccine development, implications for teaching and learning, global health, co-morbidity with other health issues (e.g., HIV, diabetes, Crohn's disease), mental health, and telemedicine. This variety of topics is reflected in the publication sources shown in Figure 1.

Example with Context

UNC-CH researchers had 1,034 publications in approximately 570 different journals or other sources over the period analyzed (January 2020 to January 2022). A total of 12 journals had 10 or more publications during the timeframe (see Figure 1).

*Figure 1. Sources with 10 or more publications for COVID-19 research published by one or more UNC-CH researchers. *A color representation of this visualization can be found via: https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).*

Journal/Source	Number of Publications	Impact Factor
Infection Control & Hospital Epidemiology	18	3.254
Science	16	47.73
New England Journal of Medicine	14	91.24
Nature	14	49.96
PLoS ONE	13	3.24
International Journal of Environmental Research and Public Health	13	3.39
JAMA Network Open	11	8.483
Journal of the American Medical Directors Association	11	4.669
American Journal of Public Health	11	9.308
Lancet	10	79.321
Clinical Infectious Diseases	10	9.079
Cell	10	41.582

Other Considerations

Using the journal impact factor (JIF) appropriately requires knowledge as to how it is calculated and when JIF for different sources can be compared. JIF is a journal-level metric based on average citations per article for recent articles published in the journal. While there remain differing opinions (Waltman & Noyons, 2018), the predominant guidance (Sugimoto & Lariviere, 2018; Wilsdon et al., 2015; San Francisco Declaration on Research Assessment, 2013; Seglen, 1998) is to use JIF only as a journal-level indicator and not as an indicator for individual publications and authors due to its vulnerability to gaming, the limited comparability across research disciplines discussed further below, and the relatively short two-year window of citations that generate the measure (Sugimoto & Lariviere, 2018). The non-profit Metrics Toolkit (available from http://www.metrics-toolkit.org/) provides a summary of JIF, including limitations and appropriate use cases.

Comparing JIF scores for journals is only applicable if they are in the same research domain. For example, Figure 1 shows that *Nature* and *Science* are both broad-coverage journals whose JIF scores could be compared. In contrast, the JIF for Infection Control and Epidemiology will have little meaning when compared to the JIF for the New England Journal of Medicine. Staff with expertise in bibliometrics offer value to research partners by educating them on responsible use of metrics like JIF (Mani et al., 2021).

Author Analysis

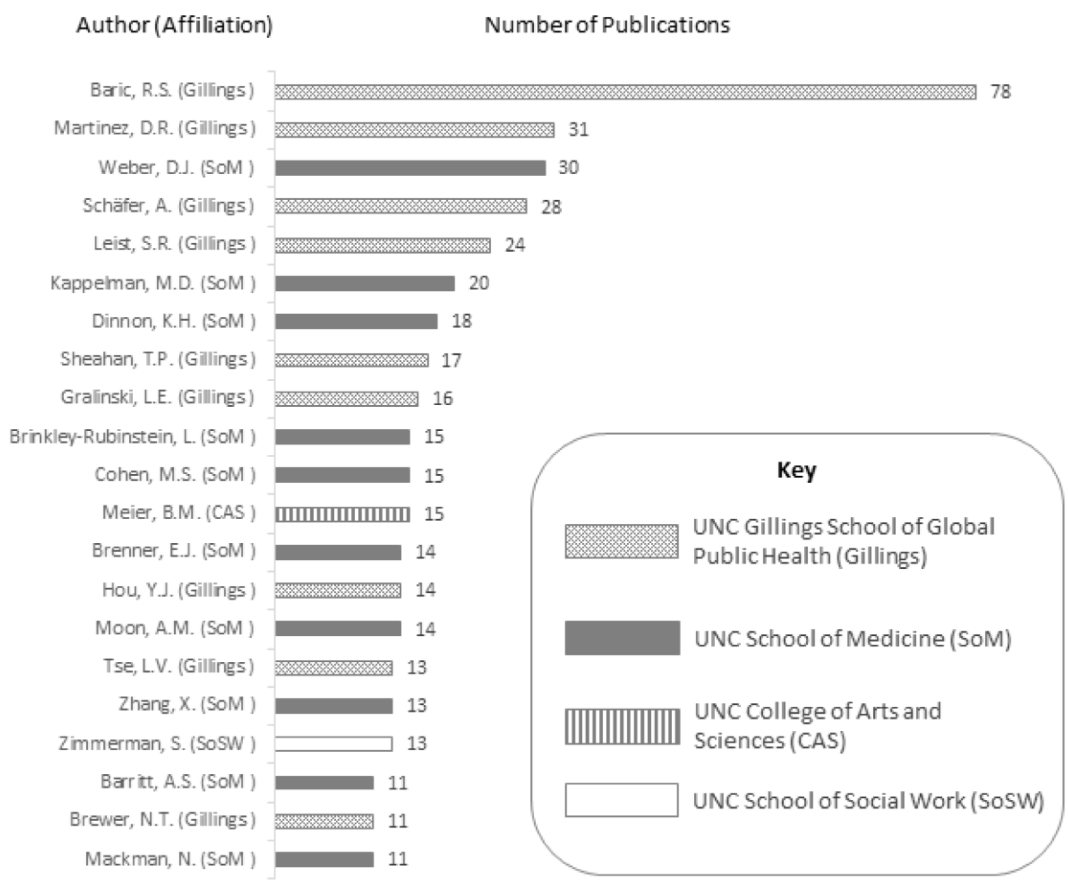
Author analysis is conducted on a set of bibliographic records to identify the authors that are publishing most frequently in the dataset being analyzed. The output from this analysis is often a straightforward list of authors with the most publications in a given dataset. Author analysis can also be combined with other data such as affiliation, unit, school, or department, or topics and sources as in a three-field plot (see below). Author analysis can be useful to demonstrate research impact for campus administrators. For instance, if a dataset of all COVID-19 research was analyzed, authors from a particular institution or group of institutions, such as all authors affiliated with one or more State University of New York (SUNY) institutions, could be ranked according to their research output relative to all authors publishing in a given research domain.

Author analysis may also be useful for enrolled or prospective students who are looking for a mentor, advisor, or dissertation committee member. Faculty may use the results of an author analysis to identify guest lecturers for courses, experts for panel discussions, or grant collaborators. Using author data, bibliometric practitioners can also identify the proportion of papers published by a single author, two authors, or three or more authors, which may be useful to understand publication patterns.

Example with Context

From January 2020 to January 2022, 21 researchers affiliated with UNC-CH, including faculty affiliated with the UNC Gillings School of Global Public Health, School of Medicine, College of Arts and Sciences, and School of Social Work, published 11 or more articles. In total, these 21 UNC-CH researchers contributed to 230 publications related to COVID-19 during the period evaluated (see Figure 2). Ralph S. Baric from the UNC-CH Gillings School of Global Public Health led the publication output of UNC-CH authors with 78 publications in the period analyzed.

Figure 2. Total COVID-19-related publications from January 2020 to January 2022 for 21 authors with 11 or more publications who reported an affiliation with UNC-CH. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).



Other Considerations

Similar to the author analysis of bibliographic data, an analysis of PIs could be performed using grant data. For example, bibliometric tools could be used to identify PIs awarded NIH grants over a 10-year period using data from the NIH RePORTER database (https://reporter.nih.gov/). Grant data tends to be more distributed across multiple sources than bibliographic data and therefore may require more effort to locate and prepare for analysis.

Unlike journal data which are typically standardized, author data require significant data preparation. Raw data must be manually reviewed to combine authors publishing with name variants, so they are treated as single authors in analyses. Authors with common names require additional time to disambiguate. Unique IDs such as ORCID IDs are helpful to minimize the effort required when disambiguating author names. However, unique IDs are only valuable to the extent they are widely used and integrated with bibliographic data by publishers and bibliographic databases. Often, publishers and databases only integrate system-independent ORCID IDs for corresponding authors, and the database's unique author ID, if there is one, is only used within that system.

The University of York (n.d.) bibliometrics guide also notes that publication counts by author will vary by many factors, including a researcher's age, career stage, and discipline (available from https://subjectguides.york.ac.uk/bibliometrics/author). This variation should be kept in mind when comparing author data.

Topic Analysis

Topic analysis can be useful to understand what research domains are covered by a set of publications. The approach is advantageous over browsing a set of articles to identify research topics, which can provide a biased picture. Clustering algorithms such as *k*-means can be used for topic analysis. Clustering is a form of unsupervised machine learning, meaning that no training data are necessary. Any text, including title and abstract text, keywords, or available full text, can be clustered. The algorithm places each record (e.g., the title and abstract for one article is a record) into a single cluster and generates a set of keywords for each cluster. It is a data-driven approach, and the algorithm-generated keywords provide descriptive information that gives insight into a dataset.

Researchers may find topic analysis useful to identify a subset of studies in a larger dataset to investigate further. This may be within a particular institution's publications or more broadly scoped to include all publications around a particular topic regardless of author affiliation. Tools for textual analysis such as clustering have the potential to be enormously helpful in locating evidence in large datasets such as the continuously expanding body of COVID-19 literature. As noted above, topic analysis can also be used on curricular data such as course descriptions, learning objectives, and syllabi.

Example with Context

The authors used the *k*-means algorithm to cluster title and abstract text in the full dataset of 1,034 publications to identify topics addressed by UNC-CH authors in 2020 publications ($N = 322$; Table 2) and publications from January 2021 or later ($N = 722$; Table 3). The *k*-means algorithm identifies a single cluster for each study and a set of keywords for each cluster that can be used to distinguish publications in each cluster. Bold formatting in Tables 2 and 3 was added by authors to indicate unique topics for each cluster. The text was clustered using *1-gram*, meaning that one-word keywords were generated by the algorithm. Additional stop words such as *copyright*, *publication*, *abstract*, *2019*, *2020*, and *2021* were excluded prior to clustering.

The keywords in Tables 2 and 3 provide insight into topics published by researchers in the articles evaluated. Splitting the dataset by date of publication offers an opportunity to understand how research focus has shifted over time. This dataset only stretches over two years, yet there is still an indication of change over time. For instance, *correctional* is a keyword appearing only in the 2020 dataset and related to the spread of COVID-19 among incarcerated individuals. *Remdesivir*, an early treatment for COVID-19, only appears in the 2020 dataset, as does the keyword *telemedicine,* which was swiftly adopted at the onset of the pandemic. The keywords *racial* and *variants* only appear in the 2021 dataset. Likewise, the keyword *recommendations* appears only in the 2021 dataset, which may be due to the time it takes to generate therapeutic and policy-based recommendations. Cluster 5 in Table 3 of the 2021 dataset is likely to contain epidemiological studies based on the keywords that appear to be related to patient risk with 95, CI, 001 related to 95th confidence intervals and 0.001 probability. Epidemiological studies take longer to complete and therefore would be less likely to be among the articles published in 2020. The two datasets also share some topics. For example, keywords related to vaccines and mental health appear in both analyses.

Table 2. Topic analysis for all UNC-CH publications related to COVID-19 from January 2020 to December 2020 (N = 322). Studies were assigned to clusters and keywords were generated using the k-means algorithm. The authors added bold formatting to indicate key topics for each cluster.

Cluster	# of Studies	Keywords
1	46	health | individuals | participants | effects | study | pandemic | social | reported | associated | perceived | communication | **mental** | learning | risk | adults | **stress** | positive | higher | impact | public
2	81	pandemic | care | patients | coronavirus | risk | **cancer** | disease | management | health | infection | **oncology** | **telemedicine** | patient | practice | providers | challenges | clinical | american | time
3	89	health | pandemic | response | data | based | **international** | public | research | time | care | systems | global | impact | **community** | **rights** | **education** | lessons | social | measures | use
4	37	prevention | action | **surgery** | pandemic | **remdesivir** | era | good | settings | income | international | health | facilities | combination | collaboration | future | **correctional** | public | services | **vaping** | water
5	69	cov | sars | infection | human | coronavirus | severe | **vaccine** | respiratory | viral | syndrome | acute | cells | disease | protein | virus | vaccines | cell | spike | antibody

Table 3. Topic analysis for all UNC-CH publications related to COVID-19 between January 2021 to January 2022 (N = 712). Studies were assigned to clusters and keywords were generated using the k-means algorithm. The authors added bold formatting to indicate key topics for each cluster.

Cluster	# of Studies	Keywords
1	170	research | health | care | challenges | clinical | **virtual** | **social** | practice | response | community | learning | **students** | education | **recommendations** | data | future | medical | public | use | **online**
2	186	health | vaccine | public | **vaccination** | disease | **social** | vaccines | people | coronavirus | global | states | **mental** | spread | united | impact | lessons | **racial** | access
3	67	patients | disease | safety | factor | mortality | associated | **ecmo** | events | levels | severity | elevated | coronavirus | **treatment** | **liver** | **venoarterial** | considered | coagulation | severe | hospital
4	134	sars | cov | infection | **respiratory** | coronavirus | acute | **antibodies** | severe | viral | virus | syndrome | antibody | disease | binding | spike | **ace2** | transmission | **variants** | viruses | infections
5	155	**patients** | 95 | CI | study | **risk** | reported | care | outcomes | compared | results | associated | data | health | methods | **001** | participants | higher | years | age

Other Considerations

VOSviewer can also be used for content analysis. The underlying approach is similar to clustering and is based on term occurrence. Results can be output as a visualization. For the bibliometric analysis for

ESOP to support curricula redesign, HSL staff analyzed faculty research output and developed an image depicting the change in keywords used over four years of publications (Mani et al., 2022). Topic analysis conducted on a dataset organized by publication year is helpful to demonstrate change in research focus over time and identify topic bursts. For example, Shen et al. (2018) conducted a bibliometric analysis on 2,704 publications related to mobile health or mHealth and noted that the research focus was shifting from acceptance to feasibility.

Collaboration Analyses

Similar to author and journal analyses, collaboration analyses are done on a set of bibliographic records using author affiliation data. The objective is to describe and visualize collaboration networks among researchers. Results may be presented in tabular form to indicate the top authors or institutions with which a given group is collaborating and as *network diagrams*, or maps indicating connections among researchers. Collaboration analyses are largely conducted to illustrate research impact and reach.

For example, UNC-CH has a mission around global engagement which includes strengthening international partnerships with academic institutions and research programs to further support international research and provide global education opportunities for students (Strategy–UNC Global available from: https://global.unc.edu/about/strategy/). Collaboration between UNC-CH researchers and those at institutions around the world provides evidence towards advancing this mission and is useful for attracting faculty and graduate students and engaging donors. Further, these data can be used to illustrate the depth and breadth of collaboration in support of research grants.

Example with Context

The 1,034 publications pertaining to COVID-19 by UNC-CH-affiliated researchers were analyzed to discern collaboration patterns. UNC-CH researchers engage in a high degree of collaboration both with U.S. institutions and international entities (see Figure 3). Collaboration analyses indicate that UNC-CH researchers co-authored five or more publications with 109 other institutions in the United States. Of these 109 institutions, UNC-CH researchers collaborated with 28 entities on 20 or more publications each—collaborations with Harvard University ranked highest with 77 publications (see Figure 4a). UNC-CH partnered with researchers affiliated with 40 international institutions on six or more publications (see Figure 4b).

In total, of the 1,034 UNC-CH authored publications analyzed, UNC-CH researchers collaborated with more than 8,000 external researchers in 102 other countries (see Figure 5). The binary map provides a powerful visualization of research breadth. Research depth can also be depicted in a map using shading. For example, international collaborations with UNC-CH authors primarily occurred with researchers in the United Kingdom ($N = 116$), Canada ($N = 106$), Australia ($N = 55$), and China ($N = 47$) (see Figure 6).

Other Considerations

Collaboration analyses are powerful visuals to show the reach and extent of collaborations. Stratifying the results by date of publication, location (e.g., domestic or international), private or public institution, or size of the institution may provide additional insights. Network diagrams are even more powerful if users can engage with underlying data. VOSviewer is used to generate network diagrams (e.g., Figure

3). The recently released VOSviewer Online version can be used to display interactive visualizations hosted on websites as well as two-dimensional images like the COVID-19 example. While a static image can be useful in a grant application or as part of accreditation documentation, an interactive dashboard allows users to explore connections more precisely.

Figure 3. Organizational collaborations between UNC-CH researchers and U.S. or international institutions with five or more shared publications. All UNC-CH units (e.g., health affairs schools, UNC Health, institutes) are combined into single data point in the center of the diagram. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).

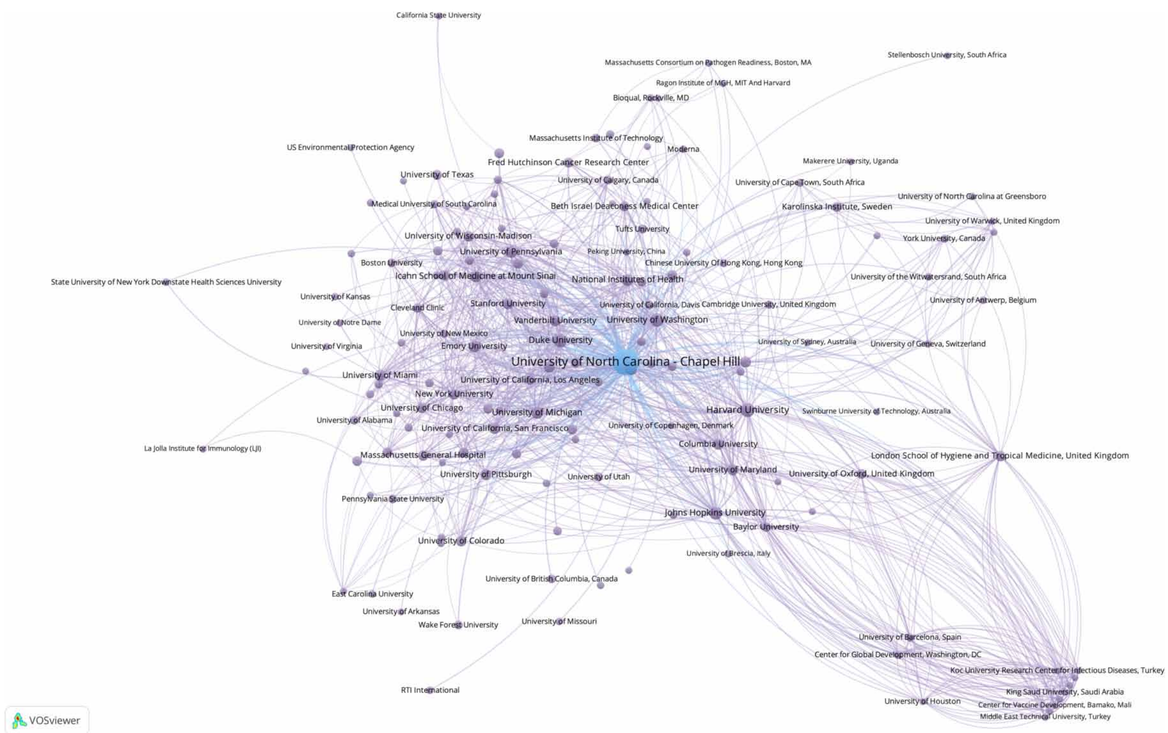


Figure 4. Collaboration between UNC-CH authors and U.S. and global institutions on COVID-19-related publications. (a) shows institutions in the U.S. with 20 or more publication collaborations; (b) shows global institutions with six or more publication collaborations. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).

(a).

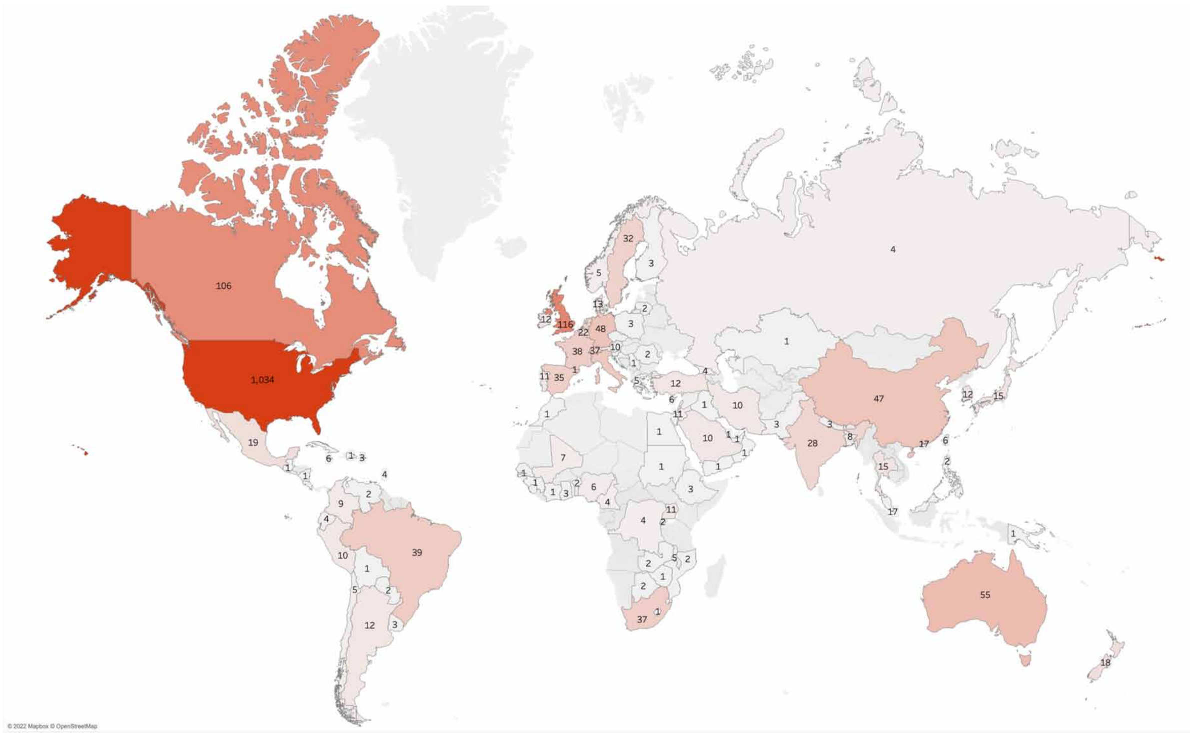
Organization	**Number of Publications**
Harvard University	77
Duke University	62
University Of Washington	62
Johns Hopkins University	56
University Of Michigan	42
Icahn School Of Medicine At Mount Sinai	40
National Institutes Of Health	39
Emory University	37
Yale University	37
University Of Pennsylvania	35
Stanford University	34
Columbia University	33
University Of Maryland	33
University Of Miami	32
Vanderbilt University	32
University Of California, San Francisco	31
Washington University, Saint Louis	31
University Of Pittsburgh	29
University Of Colorado	27
Brown University	25
Massachusetts General Hospital	25
New York University	24
University Of Wisconsin-Madison	24
Georgetown University	22
University Of California, Los Angeles	22
University Of Chicago	21
Beth Israel Deaconess Medical Center	20
Northwestern University	20

(b).

Organization	**Number of Publications**
University of Toronto, Canada	31
London School of Hygiene and Tropical Medicine, U.K.	26
University of Oxford, U.K.	21
Karolinska Institute, Sweden	19
University College London, U.K.	14
University of Calgary, Canada	12
University of British Columbia, Canada	11
Cambridge University, U.K.	10
University of Cape Town, South Africa	10
York University, Canada	10
Chinese University of Hong Kong, Hong Kong	9
University of Barcelona, Spain	9
University of Geneva, Switzerland	8
King Saud University, Saudi Arabia	7
University of Antwerp, Belgium	7
University of The Witwatersrand, South Africa	7
Koc University Research Center For Infectious Diseases, Turkey	6
University of Copenhagen, Denmark	6
University of Warwick, United Kingdom	6

Figure 5. Total reach for UNC-CH COVID-19 global research collaborations. UNC-CH authors collaborated with researchers from 102 countries, including the United States. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).

Figure 6. Variation by country in UNC-CH COVID-19 global research collaborations. UNC-CH authors collaborated with researchers from 102 countries, including the United States. Color density indicates the contribution of publications co-authored by UNC-CH authors and researchers by country. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).



Three-Field Plot

A *three-field plot* or *Sankey diagram* shows multiple strands of data in a single visualization. It works well to synthesize multiple data elements as well as show how they are related. Three-field plots may be useful to researchers as part of grant applications and to campus administrators to demonstrate research impact and breadth or to engage donors. For example, campus administrators and faculty may be interested in exploring the most prolific authors, keywords associated with their publications, and top journals or sources in a single diagram. These data would come from the author, keyword, and source fields in bibliographic records. The author and journal analyses are discussed above. The keyword analysis in this plot provides data that may complement topic analysis using clustering.

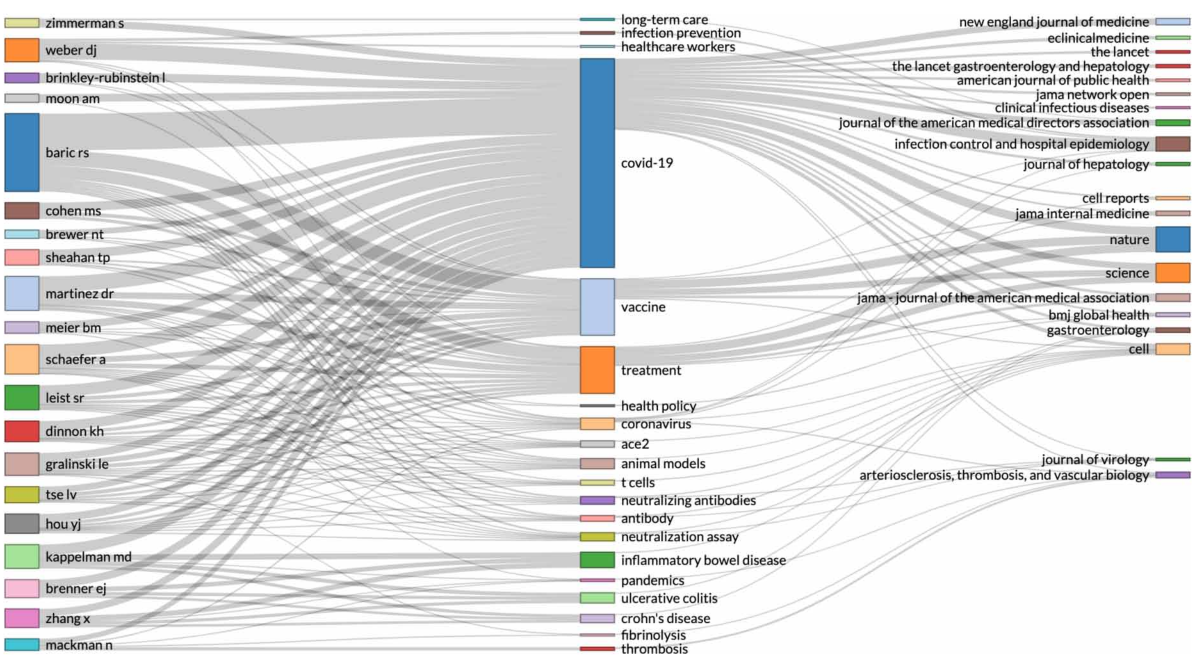
Fatehi et al. (2020) used a three-field plot to show the relationship among countries, research topics, and year of publication for articles discussing General Data Protection Regulation (GDPR) in healthcare. The GDPR was implemented in the European Union in 2018, and the plot shows what research topics the top countries are focused on by year. Das and Parabhoi (2020) used a three-field plot to depict the relationship among Universities using affiliation data, research topics using keyword data, and publication source or journal for library and information science (LIS) research published by women faculty in India. The Bibliometrix package in R offers the ability to combine data from the following fields, if available:

- Author
- Affiliations
- Countries
- Keywords
- Titles
- Abstracts
- Sources
- References
- Cited sources

Example with Context

The three-field plot was created using author, keyword, and source data from the bibliographic records for the 1,034 articles related to COVID-19 published by UNC-CH researchers. The most prolific UNC-CH researchers with publications related to COVID-19 are publishing on a variety of topics in a broad range of journals (see Figure 7). All research is related to COVID-19 as expected, and other top keywords include *vaccine* and *treatment*. UNC-CH researchers are also notably publishing on COVID-19 as it relates to gastrointestinal disorders, including inflammatory bowel disease, ulcerative colitis, and Crohn's disease. The top keywords provide insight into specific areas of expertise among UNC-CH researchers.

Figure 7. Three-field plot for COVID-19-related research published by authors affiliated with UNC-CH. Diagram depicts relationships among authors with the highest research output, most used author keywords, and publication sources. A color representation of this visualization can be found at https://cdr.lib.unc.edu/concern/multimeds/k3569d63f (Mani et al., 2022).



Other Considerations

Similar to the network diagrams used for collaboration analyses, a three-field plot is particularly powerful when users can engage with the data as part of an interactive visualization. The Bibliometrix R package, which can be used through Biblioshiny, a desktop-based user interface, can be used to construct interactive visualizations as well as static images like the example above. While a static image can be useful in a grant application or as part of accreditation documentation, an interactive dashboard allows users to explore connections more deeply.

Other Visualizations and Considerations

Additional visualizations beyond the examples provided above can be created with publication type, language of publication, or internal affiliation. For example, publication type can be used to depict how many publications are research articles, reviews, conference papers or proceedings, comments, notes, or editorials. The COVID-19 dataset referenced throughout only includes articles published in English. A dataset derived from global research will have articles published in languages other than English, and providing descriptive statistics for article language could be useful when evaluating research from international institutions.

The network diagram for UNC-CH research related to COVID-19 (see Figure 3) combines all UNC entities into a single point on the graphic. Unpacking the data by specific affiliation at UNC-CH (e.g., School of Medicine, School of Social Work) reveals that more than 18 specific sub-units on campus are publishing research around COVID-19. The internal affiliation data could be used to show collaboration among units on campus in a campus map or a network diagram similar to Figure 3.

In addition to other analyses, it may be useful to campus partners to provide access to underlying data. For example, for the COVID-19 analyses described throughout, HSL staff made the citation list and search strategies used for these analyses available for download. The topic analysis included the citation data by cluster to allow interested individuals to examine a subset of literature more closely. Paring citation lists and associated analyses with search alerts is another way to add value and help interested groups stay abreast of new research.

Future iterations of the COVID-19 analyses will focus on implementing process efficiencies so that literature and visualizations can be easily updated on a semi-annual basis. For example, data wrangling requires significant time when conducting bibliometric analyses. Simplifying this step will save significant time and allow for quicker updates to analyses. For an ongoing project such as the evaluation of COVID-19 research, process efficiencies could include search alerts and standardized thesauri for keyword and author disambiguation. Realizing efficiencies will make more feasible the creation of an interactive dashboard from which users can view and interact with real-time data.

Opportunities for Academic Libraries

Academic libraries are partners in the academic and research enterprise and can cultivate, activate, and share expertise as it relates to DS, especially in the area of bibliometrics. Advancing bibliometrics as an area of library expertise can raise awareness around DS skills among library staff. Further, when librarians partner with campus administrators to demonstrate research impact, they add value by advising on the most meaningful analyses, educating groups on appropriate, contextually informed interpretation and the use of research performance indicators (i.e., output and impact; Szomszor et al., 2021), and offering solutions to track research in a timely way. Ultimately, when librarians foster partnerships and contribute their unique skills, they can effectively raise the visibility of an institution's programs and research portfolio.

For libraries establishing a bibliometrics area of expertise, the authors recommend seeking training opportunities via webinars, workshops, and conferences. Another viable option to consider is partnering with a library that has an established bibliometrics program (which supports building capacity amongst academic library staff) or one that has a similar goal of growing a new area of expertise. The latter encourages a team-based approach to participating in and providing learning opportunities and utilizing resources, which can be mutually beneficial by reducing the financial burden required to support training and aiding in building institutional partnerships. Different ways in which this could play out include cost-sharing training resources, creating shared instructional materials, hosting online workshops and integrating staff from each institution(s) as instructors and participants, and designing use cases and project teams around bibliometrics projects to support experiential learning. Having skills in data analysis, visualization, and manipulation and an understanding of research performance indicators and their appropriate use is critical when building a team that will provide expertise and outreach for bibliometrics-based projects.

Libraries have a unique opportunity to build skills for staff and students by integrating bibliometrics skills in information and library science curricula or by designing tutorials and workshops on topics related to data analysis and visualization (e.g., Excel, VOSviewer, and Tableau), data manipulation or wrangling (e.g., R, Python, and Library Carpentry), and search strategy techniques (e.g., using resources such as searching in Dimensions, PubMed, Web of Science, Scopus, and others). As collections budgets continue to be constrained, access to subscribed resources necessary to support bibliometrics-based projects can become a hindrance. Given the corpus of bibliometrics data often stems from subscribed databases, alternatives need to be considered for institutions without subscriptions to resources such as Web of Science or Scopus. In those instances, searching databases such as PubMed, preprint servers, and freely available components of resources such as Dimensions (where publication data are free to use) is recommended. From the data analysis vantage point, several tools are freely and widely available, including tutorials for Excel, VOSviewer, and the R Bibliometrics package.

Once expertise is established and projects commence, it is critical for teams to communicate project outcomes to relevant stakeholders (keeping in mind the context in which a project request was generated) and make data used for analysis accessible to groups interested in delving beyond initial visualizations provided. When possible, it is beneficial to capture and share project outcomes in the form of a project gallery (https://hsl.lib.unc.edu/hub/project-gallery/), deposit completed projects in a local institutional repository, and share results with a broader community (via reports, presentations, publications, or faculty meetings) to help demonstrate the value of this area of library expertise and how it increases awareness and visibility of institutional progress.

CONCLUSION

The case study described in this chapter demonstrates information professionals' use of bibliometrics to increase the visibility of research and collaboration networks. These analyses increase awareness of expertise, help identify emerging areas of interest, assist in locating collaborators both internal and external to an organization, demonstrate breadth and depth of research over time, support grant applications, and help showcase an institution's research profile within a particular domain. The value that bibliometrics (and other types of research analytics) can provide cannot be overstated, as it helps raise awareness of library expertise and encourages libraries to be seen as partners in the early stages of project design and development. Providing bibliometrics instruction through credit-based courses, webinars, and hands-on workshops can increase the utilization of important bibliometrics techniques and help place the library at the forefront as a research partner. In this vein, proactively reviewing and adding new data-related services is critical as libraries develop and engage with their strategic goals in support of the research and educational enterprise. Academic libraries are at the crux of enabling growth around the use of DS tools and techniques and can ensure their central place in DS ecosystems through proactive partnership, engagement, and information provision.

REFERENCES

Cabezas-Clavijo, A., & Torres-Salinas, D. (2021). Bibliometric reports for institutions: Best practices in a responsible metrics scenario. *Frontiers in Research Metrics and Analytics, 6*, 696470. doi:10.3389/frma.2021.696470

Coppersmith, D. D. L., Nada-Raja, S., & Beautrais, A. L. (2018). An examination of suicide research and funding in New Zealand 2006–16: Implications for new research and policies. *Australian Health Review*, *42*(3), 356–360. doi:10.1071/AH16189 PMID:28297631

Cox, A., Gadd, E., Petersohn, S., & Sbaffi, L. (2019). Competencies for bibliometrics. *Journal of Librarianship and Information Science*, *51*(3), 746–762. doi:10.1177/0961000617728111

Das, J. M., & Parabhoi, L. (2020). Research productivity of LIS women faculty in India: A bibliometric study during 1988–2018. *Library Philosophy and Practice, 4298*. https://digitalcommons.unl.edu/libphilprac/4298

Drummond, R., & Wartho, R. (2009, June). Rims: The Research Impact Measurement Service at the University of New South Wales. *Australian Academic and Research Libraries*, *40*(2), 76–87. doi:10.1080/00048623.2009.10721387

Druss, B. G., & Marcus, S. C. (2005). Tracking publication outcomes of National Institutes of Health grants. *The American Journal of Medicine*, *118*(6), 658–663. doi:10.1016/j.amjmed.2005.02.015 PMID:15922698

Fatehi, F., Hassandoust, F., Ko, R. K. L., & Saeed, A. (2020). General Data Protection Regulation (GDPR) in healthcare: Hot topics and research fronts. *Studies in Health Technology and Informatics*, *270*, 1118–1122. doi:10.3233/SHTI200336 PMID:32570555

Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., Mendoza, A., & de Moya-Anegón, F. (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*, 593494. doi:10.3389/frma.2020.593494 PMID:33870055

Gutzman, K. E., Bales, M. E., Belter, C. W., Chambers, T., Chan, L., Holmes, K. L., Lu, Y.-L., Palmer, L. A., Reznik-Zellen, R. C., Sarli, C. C., Suiter, A. M., & Wheeler, T. R. (2018). Research evaluation support services in biomedical libraries. *Journal of the Medical Library Association: JMLA*, *106*(1), 1–14. doi:10.5195/jmla.2018.205 PMID:29339930

Hook, D. W., Porter, S. J., Draux, H., & Herzog, C. T. (2021). Real-time bibliometrics: Dimensions as a resource for analyzing aspects of COVID-19. *Frontiers in Research Metrics and Analytics*, *5*, 595299. doi:10.3389/frma.2020.595299 PMID:33969256

Jürgens, B., & Herrero-Solana, V. (2017). Patent bibliometrics and its use for technology watch. *Journal of Intelligence Studies in Business*, *7*(2), 17–26. doi:10.37380/jisib.v7i2.236

Kahwa, E., Dodd, A., Conklin, J. L., Woods Giscombe, C., Leak Bryant, A., Munroe, D., Ferguson, V. H., Singh, S. G., Lynch, M., & Bolton, A. (2022). A bibliometric analysis of nursing and midwifery research in the Caribbean. *Journal of Nursing Scholarship*, *54*(2), 226–233. Advance online publication. doi:10.1111/jnu.12721 PMID:35129290

Klongthong, W., Muangsin, V., Gowanit, C., & Muangsin, N. (2021). A patent analysis to identify emergent topics and convergence fields: A case study of chitosan. *Sustainability*, *13*(16), 9077. Advance online publication. doi:10.3390u13169077

Lyubarova, R., Itagaki, B. K., & Itagaki, M. W. (2009). The impact of National Institutes of Health funding on U.S. cardiovascular disease research. *PLoS One*, *4*(7), e6425. doi:10.1371/journal.pone.0006425 PMID:19641617

ManiN. S.DoddA.HayesB.YuF. (2022). *Bibliometrics supporting curriculum transformation*. doi:10.17615/mss6-fw77

Mani, N. S., Hayes, B. E., Dodd, A., Yu, F., & Cawley, M. A. (2021). New data-related roles for librarians: Using bibliometric analysis and visualization to increase visibility of research impact. In N. S. Mani & M. Cawley (Eds.), *Handbook of Research on Knowledge and Organization Systems in Library and Information Science* (pp. 317–345). IGI Global. doi:10.4018/978-1-7998-7258-0.ch017

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

Pagel, P. S., & Hudetz, J. A. (2015). Scholarly productivity and National Institutes of Health funding of Foundation for Anesthesia Education and Research grant recipients: Insights from a bibliometric analysis. *Anesthesiology*, *123*(3), 683–691. doi:10.1097/ALN.0000000000000737 PMID:26114414

Qua, K., & Pelfrey, C. M. (2020). Using bibliometrics to evaluate translational science training: Evidence for early career success of KL2 scholars. *Journal of Clinical and Translational Science*, *5*(1), e24. doi:10.1017/cts.2020.516 PMID:33948247

Saheb, T., & Saheb, M. (2020). Understanding the development trends of big data technologies: An analysis of patents and the cited scholarly works. *Journal of Big Data*, *7*(12), 12. doi:10.118640537-020-00287-9

Saheb, T., Saheb, T., & Carpenter, D. O. (2021). Mapping research strands of ethics of artificial intelligence in healthcare: A bibliometric and content analysis. *Computers in Biology and Medicine*, *135*, 104660. doi:10.1016/j.compbiomed.2021.104660 PMID:34346319

San Francisco Declaration on Research Assessment. (2013). https://sfdora.org/read/

Seglen, P. O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, *69*(3), 224–229. doi:10.3109/17453679809000920 PMID:9703393

Shen, L., Xiong, B., Li, W., Lan, F., Evans, R., & Zhang, W. (2018). Visualizing collaboration characteristics and topic burst on international mobile health research: Bibliometric analysis. *JMIR mHealth and uHealth*, *6*(6), e135. doi:10.2196/mhealth.9581 PMID:29871851

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. doi:10.100711192-021-03948-5

Sugimoto, C. R., & Lariviere, V. (2018). *Measuring research: What everyone needs to know*. Oxford University Press. doi:10.1093/wentk/9780190640118.001.0001

Sweileh, W. M. (2020). Bibliometric analysis of scientific publications on "sustainable development goals" with emphasis on "good health and well-being" goal (2015–2019). *Globalization and Health*, *16*(1), 68. Advance online publication. doi:10.118612992-020-00602-2 PMID:32723366

Szomszor, M., Adams, J., Fry, R., Gebert, C., Pendlebury, D. A., Potter, R. W. K., & Rogers, G. (2021). Interpreting bibliometric data. *Frontiers in Research Metrics and Analytics*, *5*, 628703. doi:10.3389/frma.2020.628703 PMID:33870066

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, *12*(2), 430–435. doi:10.1016/j.joi.2018.03.006

University of York. (n.d.). *Bibliometrics: A practical guide*. Retrieved March 1, 2022, from https://subjectguides.york.ac.uk/bibliometrics/author

van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. doi:10.100711192-009-0146-3 PMID:20585380

Waltman, L., & Noyons, E. (2018). *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction*. CWTS BV.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. *Higher Education Funding Council for England*. https://re.ukri.org/sector-guidance/publications/metric-tide/ doi:10.13140/RG.2.1.4929.1363

Wu, M., Zhang, Y., Grosser, M., Tipper, S., Venter, D., Lin, H., & Lu, J. (2021). Profiling COVID-19 genetic research: A data-driven study utilizing intelligent bibliometrics. *Frontiers in Research Metrics and Analytics*, *6*, 683212. doi:10.3389/frma.2021.683212 PMID:34109284

Xing, Z., Yu, F., Du, J., Walker, J. S., Paulson, C. B., Mani, N. S., & Song, L. (2019). Conversational interfaces for health: Bibliometric analysis of grants, publications, and patents. *Journal of Medical Internet Research*, *21*(11), e14672. doi:10.2196/14672 PMID:31738171

Yu, F., & Hayes, B. (2018). Applying data analytics and visualization to assessing the research impact of the Cancer Cell Biology (CCB) Program at the University of North Carolina at Chapel Hill. *Journal of Escience Librarianship*, *7*(1), e1123. doi:10.7191/jeslib.2018.1123

Zyoud, S., & Zyoud, A. (2021). Coronavirus disease-19 in environmental fields: A bibliometric and visualization mapping analysis. *Environment Development and Sustainability, 23*. https://doi/org/doi:10.1007/s10668-020-01004-5

ADDITIONAL READING

Aguillo, I. F. (2016). Informetrics for librarians: Describing their important role in the evaluation process. *El Profesional de la Información*, *25*(1), 5–10. doi:10.3145/epi.2016.ene.01

Ball, R., & Tunger, D. (2006). Bibliometric analysis—A new business area for information professionals in libraries? *Scientometrics*, *66*(3), 561–577. doi:10.100711192-006-0041-0

Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S., & Grant, J. (2013). *Measuring research: A guide to research evaluation frameworks and tools*. RAND Corporation.

Hendrix, D. (2010). Tenure metrics: Bibliometric education and services for academic faculty. *Medical Reference Services Quarterly*, *29*(2), 183–189. doi:10.1080/02763861003723416 PMID:20432142

Hicks, D., Wouters, P., Waltman, L., de Rijcke S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–31.

KEY TERMS AND DEFINITIONS

Altmetrics (Alternative Metrics): Metrics designed to capture attention a resource (e.g., article) is receiving online such as discussions on social media platforms or in blog posts.

Bibliometrics: The quantitative analysis of publication data to determine indicators of research output, impact, and collaboration that help identify patterns, trends, and gaps in research, education, or clinical care activities.

Citation Analysis: A major method of bibliometrics that examines the quantitative data derived from the use of citations to reference and connect documents. Citation metrics are used to assess the scholarly influence or impact of publications and researchers.

Clustering: Type of machine learning that does not require training data or supervision. The *k*-means algorithm, nonnegative matrix factorization (NMF), and latent Dirichlet allocation (LDA) are examples of clustering algorithms. This is also referred to as *unsupervised machine learning*.

Data Management: The practice of describing and organizing research data to increase understanding and enable further analysis of the data. Well-managed data enables data to be discovered, accessed, and reused to replicate results or repurpose for new research.

Data Visualization: The discovery of data insights and communication of findings through techniques that employ the innate ability to distinguish visual patterns.

Responsible Metrics: The use of quantitative, citation-based measures or indicators (e.g., citation counts, *h*-index, and journal impact factor) in the evaluation of education and research in ways that are accurate, appropriate, transparent, and ethical.

Scientometrics: The application of quantitative methods such as bibliometrics and citation analysis to the analysis of scientific research.

VOSviewer: A software tool for constructing and visualizing different kinds of bibliometric networks. The tool has been available as a desktop application since 2021 as an interactive, online application. It is developed and supported by the Centre for Science and Technology Studies (CWTS) at Leiden University.