Applications of Bibliometrics

This chapter is intended to inspire ideas for the practical use of the tools discussed in the previous chapter and to help practitioners connect information needs with the types of bibliometric analyses that might help respond to these needs. It is appropriate here to remind ourselves of the very important challenge of using metrics responsibly, as discussed in chapter 1. If you are unsure where to start, the guidance available from the SCOPE Framework (INORMS Research Evaluation Group 2020) challenges us to think first about the “value of the entity” that is being evaluated or measured before using any bibliometric analysis. Using the SCOPE Framework as a starting point reminds us to use bibliometric analysis only if it matches our values and to avoid the streetlight effect (Freedman 2010) of counting things only because they can be counted. The SCOPE Framework also presents a very helpful decision matrix to help identify the amount of risk or impact involved depending on the motivation for using the analysis and what entity level is being evaluated (figure 3.1).

Somewhat aligned with this risk matrix, the applications of bibliometrics that will be discussed here are broken up into four types:

- collections development,
- ranking validations,
- strategic planning and unit-level analyses, and
- research groups and individual analyses.

Of course, there are likely infinite ways to organize the different analysis types. However, these groupings will also loosely align with the various bibliometric services at academic libraries or institutions. In presenting the applications, an attempt has been made to generalize the data or these details so that they can be adjusted to unique situations; however, in some cases there are references to more descriptive materials. It is recommended that you pursue these for further guidance and detail. The step-by-step instructions cannot be captured here as these tools are constantly changing and evolving. The idea is to get a sense of the possible.

There are already some very excellent sources that define and describe bibliometric indicators. Measuring Research: What Everyone Needs to Know by Cassidy R. Sugimoto and Vincent Larivière (2018) is a particularly succinct, yet thorough, recent review of the main bibliometric indicators, their limitations, and considerations in interpreting the data. Somewhat more dated, yet still very relevant, is Measure Academic Research: How to Undertake a Bibliometric Study by Ana Andrés (2009), which divides the indicators into several groups: descriptive indicators, author productivity, collaborations, author citations, and journal citations. These sources and those provided in the reference list are excellent resources for understanding the breadth of indicators that could be used in any bibliometric study. This report is taking a more practical approach to the use of these indicators and attempts to describe the use of bibliometrics for specific service-oriented applications that a practitioner may encounter.

With a focus on the practitioner, this report also has to acknowledge that the full spectrum of bibliometric methodologies cannot possibly be covered.1 Certainly, there are highly skilled expert-level practitioners and teams that have exceptional experience with a variety of complex analyses using tools and skills that go beyond the reach of this report, such as building and using data science methods with machine language algorithms or complex relational databases. Instead, this report focuses on the use of the bibliometric tools discussed in chapter 2, attempting to give the entry-level to mid-level practitioner some guidance on the various applications of these tools. However, keep in mind that most bibliometric practitioners will be required to develop some comfort with downloading
and analysis external to the bibliometric tools. At the very least, developing a proficiency in using Excel pivot tables is certainly a good start.

### Collections Development

#### Scenario 1: Looking at Potential Impact of a Transformative Agreement on Authors

A university library’s collection development department wanted to know how many of its authors from its institution have published with a particular publisher in recent years to inform on the potential impact of a recent transformative agreement. Due to challenges in extracting complete publisher and journal data from some of the main bibliometric tools, the data source Crossref was used to create an API data pull using the affiliation name and publisher via its member code. With this analysis, the library was able to determine the count of affiliated articles published each year in journals by the publisher of interest, and the analysis would aid in year-over-year costing predictions if needed (figure 3.2).

Tools used: Crossref REST API and supporting documentation on GitHub.
Transformative agreements are shifting the way publishers and libraries do business. The not-for-profit Jisc represents the UK higher-ed sector as a consortium and negotiates deals on behalf of numerous member academic institutions. It has an interest in monitoring the impact of transformative agreements and recently presented at the Bibliometrics and Research Impact Community Conference on the methods used in its investigation (Harris 2022). In the presentation slides, there are several valuable resources it has created that may be of interest to the practitioner.

Scenario 2: Finding Core Journals Using Bibliometrics

Bibliometrics does not need to be complicated. When trying to understand collection development needs, bibliometrics can be used to tailor a core journal list to a particular research area, a group of researchers or author, an institution, or all of these together.

For example:

- A list of publications can be created using journal subject classifications or a tailored keyword search to define a research area and filter the results to your institution using a bibliographic data source such as Web of Science or Scopus. The resulting publication list can be uploaded into a bibliometric tool such as SciVal or InCites to analyze the journal or source list by several indicators such as total outputs, citation counts, or normalized indicators (table 3.1).
- The same publication list can be analyzed by citing and cited articles to get a list of journal titles (or other sources) related to the core publication list. This can help expand a core title list to include not just titles in which your authors are publishing but also titles they are using in the creation of their publications (table 3.2).
- The overlap between the title lists is highlighted with a gray screen in table 3.1 and table 3.2. Although this is just a simple example, it highlights that viewing the bibliometrics from different angles can help create a more robust picture of the resources that are valuable to a particular institution or research domain.

Scenario 3: Finding Core Journals Using Patent Literature

Galter Library at Northwestern University (Pastva et al. 2020) was interested in using patent literature to help identify highly cited journal publications within the health sciences research domain and to determine if these journals corresponded with usage within its existing collections (obtained from COUNTER usage statistics) and the Journal Impact Factor (obtained from InCites Journal Citation Reports). It used Dimensions as the bibliometric data source and was able to obtain NU author/inventor patent information as well as the journal article information that was cited within these patents. It found that the Journal Impact Factor did not correspond with the citation data or usage data and questioned its utility in making collection development decisions. However, the journal most cited in the patent literature did correspond with the usage data. From this analysis, Northwestern found that its existing collections aligned with the identified journals in its patent-citation analysis; however, a different set of top-cited journals, or core journals, were identified and could be used to help strengthen its collection development decisions. The visualization it used to help communicate these findings (although not reproducible here) plotted each journal along the x-axis in decreasing order of patent citation counts with secondary plots of the Journal Impact Factor and usage along the y-axis. Their work illustrates that patent citation has better alignment with usage counts versus the poor alignment with Journal Impact Factor.

Summary of tools: Dimensions (user interface and API), COUNTER, InCites, Excel, Python (and Jupyter), Tableau.

Further Reading on Bibliometrics for Collection Development

The following papers have not been referenced in this section, but they may be of interest to the reader.


The proliferation of university rankings has captured the attention of academic institutions around the globe, with administrative units contributing to the data submission and validation processes and including their ranking scores within university communications. It is now very common to see a ranking score on the splash page of a university website. However, rankings have been heavily criticized for reducing institutions to a few quantitative measures.
that mask significant nuance in the actual research and teaching missions of the institutions being evaluated (Gingras 2016). There is also a lack of consistency and transparency within and between the methodologies undertaken by the ranking bodies. Therefore, the rankings can seldom be compared from year to year and certainly cannot be compared to each other. Despite this, the participation in rankings is not slowing down. For example, since the introduction of the UN Sustainability Goals as an assessment benchmark within the Times Higher Education (THE) University Ranking in 2019, there has been considerable uptake in the institutions participating, growing from 467 in 2019 to 1,410 in 2022. This increased attention to rankings means that the bibliometric indicators are also gaining greater attention as a result. Access to the methodologies is therefore important if institutions are going to be able to understand, keep up with, and maybe even push back on how they are being ranked. It is in the institutions’ best interest to be able to understand and respond to changes in their rankings.

It is, however, important to understand the motivations behind rankings and the use of any indicator for that matter. As Yves Gingras (2016) discusses in his very on-point book titled Bibliometrics and Research Evaluation: Uses and Abuses, there are lessons to be learned from reviewing the impacts of national rankings such as the UK Research Evaluation Framework (REF); the bibliometrics-based university-funding formulas of Australia, Flanders, and Belgium; and the French grandes écoles (Gingras 2016). At best these rankings provide flimsy proxies for more time-consuming qualitative measures such as peer review or the nuanced pursuit of truth via academic rigor, and at worst the specific indicators chosen bias the outcomes toward preconceived notions of rank—for example, a focus on total outputs biases toward larger, better funded institutions, which can be clearly seen in the overrepresentation of privately funded American universities in many national and international rankings.

The major ranking organizations are ShanghaiRanking, also known as the Academic Ranking of World Universities (ARWU); the THE World University Ranking; the QS World University Ranking; and the Centre for Science and Technology Studies (CWTS) Leiden Ranking. For a more complete list the IREG maintains a list of international rankings (IREG Observatory, 2021). However, nationally significant rankings are not covered.

Validating the QS World University Ranking

Each year the QS requests that submitting institutions validate the data that is to be included in their ranking. The data that is shared with the institution for validation is a combination of institutionally submitted data and Scopus source data that QS extracts and analyzes. The bibliometric data includes gross number of papers, gross number of citations, net number of papers, normalized number of papers, net number of citations including self-citations, net number of citations excluding self-citations, and normalized number of citations. The details of the methodologies that QS applies to generate these values are detailed in its methods documentation. Although the methodology cannot be completely replicated, the data is based on institutionally affiliated documents within a specified five-year publication window. To obtain all the necessary bibliographic metadata for each article, the SciVal data set needs to be employed. This is because information such as the number of affiliations and the All Science Journal Classification scheme used in its methodology are not available directly from the Scopus data, and therefore the publication sets have to be pulled from SciVal. This validation set allows institutions to have some control over the data that is being used in the ranking and also provides an opportunity to learn more about the methodology, along with its strengths and weaknesses, and can inform the institutions on areas where they may see opportunities for growth or where they may prefer to remain less active.

Validating the ShanghaiRanking

Even when the ranking organization does not involve the ranked institution in a data validation process, undergoing a data validation may still be of interest to the institution. For example, an institution might be interested in understanding what influenced a recent increase or decrease in its ranking in the ShanghaiRanking’s global ranking of academic subjects. Using the methodology information provided by ShanghaiRanking, it could attempt to replicate the indicators used in the ranking, which include the number of journal publications with the Q1 Journal Impact Factor Quartile, the Category Normalized Citation Impact index value, the number of publications with international collaborations, and the number of publications that received the highest number of votes from the ShanghaiRanking Academic Excellence Survey. Keeping up to date with the institution’s bibliometric indicators may aid in trend analysis or even identify high-performing research areas previously unidentified internally.
Strategic Planning and Unit Level Analyses

Bibliometrics can be more informative to institutions when they take a multidimensional approach to the data and break out of the confines of university rankings that reduce the complex organisms of academic institutions into a single rank-ordered list. Therefore, the analyses outlined here are only a jumping off point to give the reader some ideas of how bibliometrics can be applied when doing analysis for the purposes of planning at the university or strategic level. Using bibliometrics internally gives you more control over the data, and a more detailed story can be built. For example, in Canada the mostly public universities are funded partly based on the types of programs they offer, with medical schools not only being the best funded but also benefiting from a large network of affiliated hospitals and publishing in research areas that typically have high output and high citation rate. Therefore, universities without medical schools are certainly going to seem to underperform compared to these other universities. But not all universities can have medical schools. Therefore, it is important to make sure that the benchmarking is created in a way that is either comparing schools with similar characteristics or taking the differences into account. Otherwise, the bibliometric analyses will be hiding the real story behind a charade of misleading numbers.

Scenario 1: Benchmarking with National Comparators

An institution was interested in benchmarking against a set of its peer institutions in its country. However, this is entering a territory similar to ranking, where it is easy to reduce an institution to a rank using only a single (or at best a few) oversimplified indicators. As discussed above, benchmarking should be handled with great care. It is ideal to present a breadth of indicators or to use multidimensional analyses and ensure that the choices of comparator institutions, data filtering, and selected indicators are reasonable and clearly communicated. Therefore, the institution decided to use the SciVal data set, with the field-weighted citation impact (FWCI) as the main indicator, evaluate the data from a five-year window, and present the data in two figures, one including all subject classifications (figure 3.3) and another
that has publications from the medical sciences and related fields filtered out of the data (figure 3.4). This allowed the institution to clearly see the effects of having a medical school on even a normalized citation index like the FWCI and that there is obviously an advantage to focusing on the medical sciences. The message is not necessarily that the university should pursue a medical school or even more research within medical science. Rather, it is clear that the university has strengths outside of these research areas. What these strengths are precisely cannot be determined from this unidimensional analysis. Therefore, it may want to investigate further.

Tools used: SciVal and Excel.

**Scenario 2: Understanding the Research Focus of an Institution in the National Context**

Following the previous analysis, the institution was interested in understanding some more detail about the research areas that make it stand out at the national level. Therefore, it used a multidimensional

---

**Figure 3.5**

FWCI relative to the national average FWCI, plotted against the percentage of national outputs from the institution, by research area, 2016–2020. Source: SciVal.
The researchers Pablo García-Sánchez and Manuel J. Cobo (2018) wanted to explore the impact of international collaborations with the researchers from universities within the Andalusian region of Spain. They wanted to know if publications with more geographically diverse authorship collaborations would see higher citation rates. They used the Dimensions API and Python code to export publications from the nine public universities of Andalusia, identified using the Global Research Identifier Database. They filtered the publications to include only articles as the publication type and the publication years 2010–2015. They were interested in looking at papers authored by only one university in Andalusia, papers where all the authors belonged to Andalusian universities, papers where all the authors were Spanish and at least one was Andalusian, and finally all the Andalusian-authored papers with coauthors from any region of the world. This provided a very interesting perspective on collaboration networks and the progressive diffusion of authorship collaboration types. The papers in the group that had the most geographically diverse authorship collaborations were much more likely to receive a high number of citations. Further details of this study, including all the figures from the analysis, are available in the full paper (García-Sánchez and Cobo 2018). On the other hand, it must be kept in mind that confounding factors could affect the citations obtained by these authorship collaborations, such as the simple effect of more authors on a single paper meaning it may receive more citations. Controlling for these variables may be needed to get a clearer picture of the real impact of authorship collaborations.

Tools used: SciVal, Excel

**Scenario 3: Exploring the Effects of International Collaboration**

The researchers Pablo García-Sánchez and Manuel J. Cobo (2018) wanted to explore the impact of international collaborations with the researchers from universities within the Andalusian region of Spain. They wanted to know if publications with more geographically diverse authorship collaborations would see higher citation rates. They used the Dimensions API and Python code to export publications from the nine public universities of Andalusia, identified using the Global Research Identifier Database. They filtered the publications to include only articles as the publication type and the publication years 2010–2015. They were interested in looking at papers authored by only one university in Andalusia, papers where all the authors belonged to Andalusian universities, papers where all the authors were Spanish and at least one was Andalusian, and finally all the Andalusian-authored papers with coauthors from any region of the world. This provided a very interesting perspective on collaboration networks and the progressive diffusion of authorship collaboration types. The papers in the group that had the most geographically diverse authorship collaborations were much more likely to receive a high number of citations. Further details of this study, including all the figures from the analysis, are available in the full paper (García-Sánchez and Cobo 2018). On the other hand, it must be kept in mind that confounding factors could affect the citations obtained by these authorship collaborations, such as the simple effect of more authors on a single paper meaning it may receive more citations. Controlling for these variables may be needed to get a clearer picture of the real impact of authorship collaborations.

Tools used: SciVal, Excel

**Interdisciplinarity Analyses**

The areas of interdisciplinarity analyses using bibliometrics are too complex to illustrate in a short example here. Therefore, we will explore the variety of methodologies outlined by Larivière and Gingras in their book chapter “Measuring Interdisciplinarity” (Larivière and Gingras 2014). They outline that interdisciplinarity has been measured in the following ways:

- identifying the authors’ disciplinary affiliations using cited and citing references by either
  - measuring the percentage of citations made by sources outside the discipline or
  - measuring the percentage of papers from a group of researchers that publish papers outside their “main” discipline
- identifying articles that are published in journals classified in more than one research area, using standardized classifications such as the ASJC or Web of Science categories
- measuring the movement of a researcher from one discipline to another throughout their career

**Research Group and Individual Analyses**

Analyses at the research group or individual level require access to author-level data from a bibliometric data source. Although it is possible to create publication sets based on author name or author ID searches within any bibliometric tools or their associated data source, it is only SciVal that currently allows the creation and management of author groups and hierarchies within its system. This provides a great advantage to the system as groups can be created regardless of the accuracy of their affiliation information. For example, a researcher may forget a credit in a paper to a department where they are working as an adjunct faculty member. They may just fill in their main institution and department as their affiliation. However, the other institution may still like to count that paper in its analysis. It can do so by including the researcher in a group in SciVal. With this example in mind, this section illustrates two examples where the SciVal author tool provides an advantage for the analysis. However, similar analysis may be possible with the
other bibliometric tools with a bit of creativity or by working through a few more steps, such as creating a search string of author IDs.

**Scenario 1: Investigating the Impact of Mobility on Early-Career Researchers**

The researchers Maxim Kotsemit, Ekaterina Dyachenko, and Alena Nefedova were interested in looking at the impact of mobility on young, early career researchers at the National Research University Higher School of Economics. Using the researchers’ curricula vitae, they were able to select researchers based on their age (< 39 years) and sort them according to whether or not the researcher had past international educational opportunities that lasted at least three months. With this set of mobile and nonmobile researchers, they uploaded the researchers’ Scopus author IDs into SciVal and organized them into their respective groups. This enabled them to analyze the two groups of researchers based on a number of bibliometric indicators, including number of publications, number of publications per researcher in each group, average number of citations per publication, and the field-weighted citation impact, among others. This study found that there was a positive correlation between mobility and a number of indicators such as number of papers, the prestige of the journal (based on the CiteScore), and citations (Kotsemit, Dyachenko, and Nefedova 2021).

**Scenario 2: Exploring the Impact of Coauthorship on Citations**

The researchers Nicola Cucari, Ilaria Tutore, Raffaella Montera, and Sofia Profita wanted to further analyze a list of top-cited authors in the field of corporate social responsibility that they discovered through a topic analysis in SciVal (Cucari et al. 2022). They were interested in understanding more about the collaboration activities of these authors based on the assumption that authors with more international collaborations do not always have higher publication output or citations. They used the author identifier ORCID to create publication sets from the Scopus database that could be uploaded in the VOSviewer system for network analysis. The resulting visualization of the coauthorship analysis by countries illustrated the relative productivity of each country in the field of corporate social responsibility and how strongly connected each country was based on the number of coauthored papers. The strength of the connection was visualized by the closeness of the nodes (countries) and the thickness of the edges (number of coauthored papers). Their analysis showed that countries like the United States which have high productivity also have many coauthorship links; however, there are also countries like Australia with a good share of coauthorship links that are less productive. This may support the authors’ assumption.

Tools used: SciVal, Scopus, VOSviewer

**Notes**

1. Only one example of a network analysis is used in the case studies presented in this report as many examples of scholarly research detail network analyses and comparatively few present case studies on the major commercial bibliometric tools. This report attempts to fill this gap. The paper by Makar and Trost (2018) provides some additional cases studies of network analyses that are directed at the practitioner.

2. Note that this contradicts the earlier example from García-Sánchez and Cobo (2018), who found that authors with more collaboration among external institutions had higher citations.

**References**


The Current and Evolving Landscape of Bibliometric Tools and Technologies

Laura Bredahl