Data Visualisation for Research Impact: A Reflection of an Exploration

Eleanor Colla

University of New England, Armidale, NSW, Australia

Abstract:

In this column, the author reflects on the process of investigating and implementing the use of data visualisation software into the workflows of an academic librarian working in a research impact team at a research-intensive Australian institution.

Author Bio:

Eleanor Colla is a Researcher Services Librarian at the University of New England. Her library-related interests include academic impact, data visualisation, research services and support, and knowledge management

Keywords: *academic libraries, Australia, academic impact, bibliometrics, data visualization*

The Setting

In recent years, there have been increasing calls and mandates for institutions and researchers to provide data which demonstrate the impact of their research (Australian Research Council, 2015). Researchers must now demonstrate the ways in which their research influences their field, how it shapes collaboration with others, and how it impacts Australian society at large (Australian Research Council, 2015). In turn, institutions must now demonstrate that they encourage their faculty to participate in interdisciplinary research and academic-corporate collaborations, the output of which has practical impacts outside of academia (Australian Research Council, 2017). In response to this demand, the use of bibliometrics is increasingly sought-after to provide quantitative, demonstrable evidence to support claims. Concurrent to this trend is the ever-continuous expansion of automated software technologies, big data, and data visualisation. Within librarianship particularly, a more hands-on approach to technology and software, as seen through events such as library carpentry and hacky hours, is occurring. It is within this context that I entered into academic librarianship and began working for a research-intensive Australian institution that was invested in creating and sustaining research relationships, while tracking how research is disseminated, received, and reused, and looking at the types of impact this research is having.

The Reasons

In my role, I provided metrics-based reports for use in grant applications, promotion dossiers, school reviews, and national and international delegations. This involved compiling bibliometrics from traditional, non-traditional, and alternative sources to assist in creating an evidence-based narrative that reflected the impact of researchers' works. Whilst talking to researchers, I became increasingly aware of the changing landscape in Australian tertiary funding requirements, and, after attending a

number of library-focused software-building events, I found myself wanting to present information in a more interactive way whilst extracting different meanings and narratives from the data.

The Actions

My educational background is not based in mathematics or computer science, so a great deal of time went into learning a new vocabulary of terms and concepts. I compared software programmes, wondering which I should use, or whether I should find/create/run a code through a dataset. If the latter, should it be R, or Python, or another programming language altogether? I contacted developers, read forums, and looked into what other institutions were using and recommending. I completed the majority of this research over a number of months, finding spare moments at work or on the weekend. It was a long, confusing, and often frustrating process. I eventually began investing more time in software programmes that showed relationships within a chosen network, which integrated easily with the data I was sourcing—mainly from Scopus and Web of Science.

Ultimately, I integrated VOSviewer, an open source programme developed at Leiden University for constructing and visualising bibliographic networks. This allowed me to show a variety of relationships between researcher networks (citation, co-citation, and collaboration), as well as keyword and subject headings, journal sources, and more into my workflow. It is possible in a single visualisation, for example, to show institutions (label of the node), their contribution to the dataset (size of the node), their connection to other institutions (through their placement within the visualisation and their links to other nodes), and their geographic location (colour of the node).

The Results

By presenting the same datasets in a different way, I was able to highlight information that revealed which individuals or institutions were more likely to collaborate, how often certain words were

used and how often they were used with other words, when researchers had cited each other, and what journals were shown in a dataset (see Figure 1).

When these data were first included in reports and outputs, the reader required an explanation of the visualisations, how they were created, and how this information differed from what was being presented in tables and graphs. On the whole, readers reacted well to a visual representation of the data, engaging with it more readily than a textual representation alone. Throughout this process I noticed a version of "digital divide" beginning to form. Some team members were supportive, though they often didn't have the time or motivation to invest in the research, whilst others did not see the benefit of investigating and integrating new types of data visualisation. As I gained proficiency in extracting and visualising bibliometric data, regularly using them in reports and research consultations, my colleagues began to take a more active interest in what they could do themselves. In an effort to help close this digital divide, I eventually created and ran a series of training sessions that allowed people to look into this process more closely, allowing them to integrate it into their own workflows.

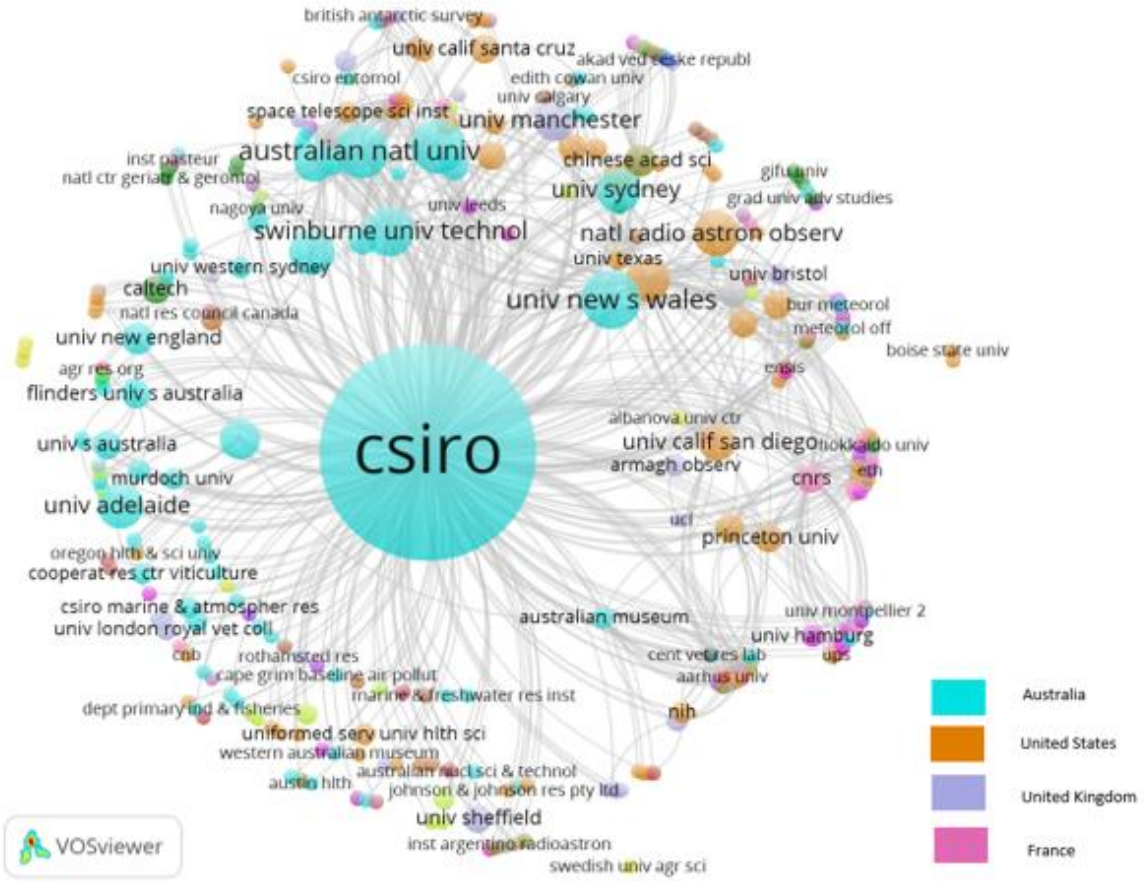
The Future

Ultimately, using data visualisation software is an addition, not a replacement, for the static tables and lists that make up our traditional reports. This is for two key reasons. Firstly, showing elements of the raw data is useful and sometimes all that is required and, secondly, VOSviewer is an inherently interactive product; as soon as a screenshot is taken, it loses this interactive appeal, granular details, and the ability to be easily updated. Throughout this process, I realised just how easy it is to manipulate data when you present them in a visualised way. By changing the repulsion/attraction setting, for example, the visualisation's entire layout may change. Similarly, by setting different minimum and maximum limiters, entire keywords, subject areas, or researchers may appear or disappear.

I found undertaking this entire process to be a worthwhile experience. I expanded my knowledge in data visualisation and interacted with the wider research community through forums, hacky hours, and through simply emailing people to see if they could offer advice. By undertaking this research, I created and delivered a training programme that integrated new elements into the team workflow, providing our clients with a new and different way of looking at data. This, in turn, created more informed and comprehensive narratives about their research in a number of ways.

Institutions represented in CSIRO publications

Web of Science indexed Articles and Reviews with 'CSIRO' in Organizations-Enhanced field, published Open Access, with 2017 publication date.



Each institution is represented by a node, the size of which is relative to the number of publications, and the colour of which represents a geographic area. The lines between nodes represent occurrence of collaborations.

Figure 1: VOSviewer settings: minimum occurrence is once, full counting, Attraction:Repulsion 4:1

References

Australian Research Council. (2015). *Research impact principles and framework*. Retrieved from http://www.arc.gov.au/research-impact-principles-and-framework

Australian Research Council. (2017). *Engagement and impact assessment*. Retrieved from http://www.arc.gov.au/engagement-and-impact-assessment