

IDENTIFYING RESEARCH TOPICS WITH BIBLIOMETRICS: GUIDELINES FOR PRACTITIONERS

Amy Trost, The Universities at Shady Grove/University of Maryland Libraries



BACKGROUND

For many researchers of bibliometrics in academia, the discipline has evolved from hierarchical & metric-focused to more descriptive and story driven (Franssen & Wouters 2019). The field is changing for practitioners, as well. While research impact assessment is still a cornerstone of promotion & tenure at many universities (Tegoning 2018), librarians and other who support bibliometric analysis are responding to customer requests for more descriptive topical and citation analyses to help with strategic planning and narrative assessments (Trost & Makar, MacDonald & Dressler).

This research explores potential use cases for a more narrative style of bibliometric analysis and outlines six guidelines for librarians and others who practice bibliometrics as a service for researchers and managers. Case studies from two separate disciplines—academic librarianship and translational science research—attempt to answer the following research questions:

- 1. Can the co-occurrence of thesaurus terms be used to map the research landscape in a specific discipline?
- 2. How is the research focus of a specific institution within the University of Maryland different from the focus of the broader collection of documents?
- 3. How can other librarians who practice bibliometrics implement these techniques into their portfolio of services?

KEYWORD ANALYSES

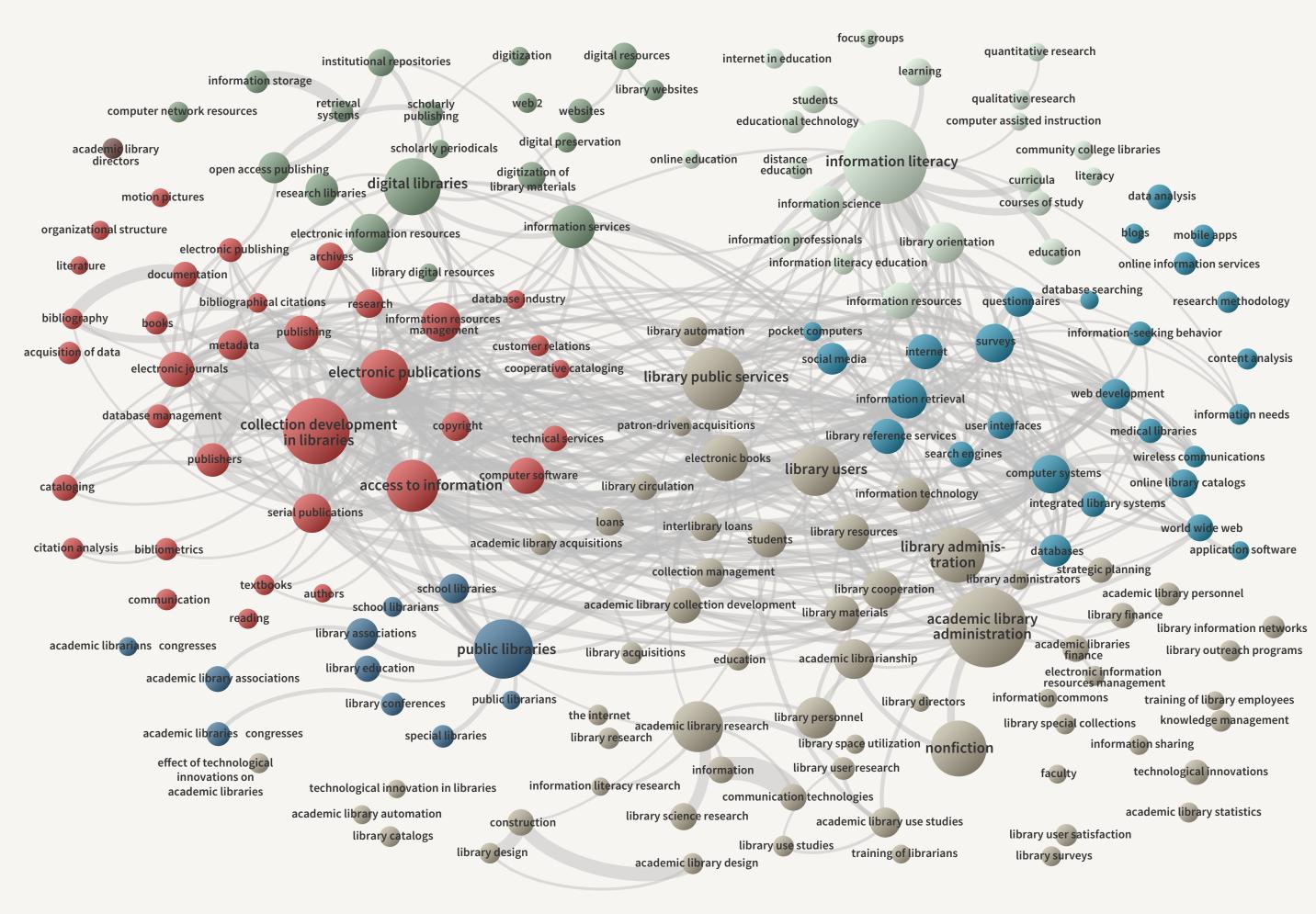
Comparing commonly assigned keywords in translational science research, 2019



This table shows the most commonly assigned subject categories for translational research in both the full result set and among the 227 records that listed the National Institutes of Health (NIH) as an author affiliation. NIH was chosen here as a proxy for the types of research activity occurring in Montgomery County, MD; every TLST student at Shady Grove must complete a professional internship and project-based research experience in the region. Roughly two-thirds of the subject categories assigned to NIH authors are broadly used in the general collection.

NETWORK VISUALIZATIONS

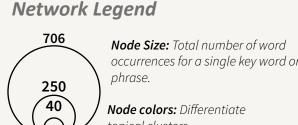
Co-occurrence of the most commonly assigned keywords related to academic librarianship, 2008-2019



database provided database-assigned keywords for 16,248 articles related to academic librarianship. Using R's text mining(tm) package, a term-document matrix was created to calculate the network shown here. Network clusters were calculated in VOSviewer, and the visualization was generated in Gephi and edited with Adobe Illustrator.

Roughly a dozen very common terms (e.g., libraries, academic librarians, etc.) were removed from the final visualization. Six major clusters reveal some surprising connections among the sub-disciplines within academic

Digital and technical services are scattered throughout all six clusters. Scholarly publishing and digitization is split between two clusters (deep green and light blue). Professional associations and conferences (dark blue) are linked with public libraries, while many of the terms associated with staffing and development are in a separate (brown) cluster. Information literacy, the largest node in the network, anchors a cluster (light green) that is strongly



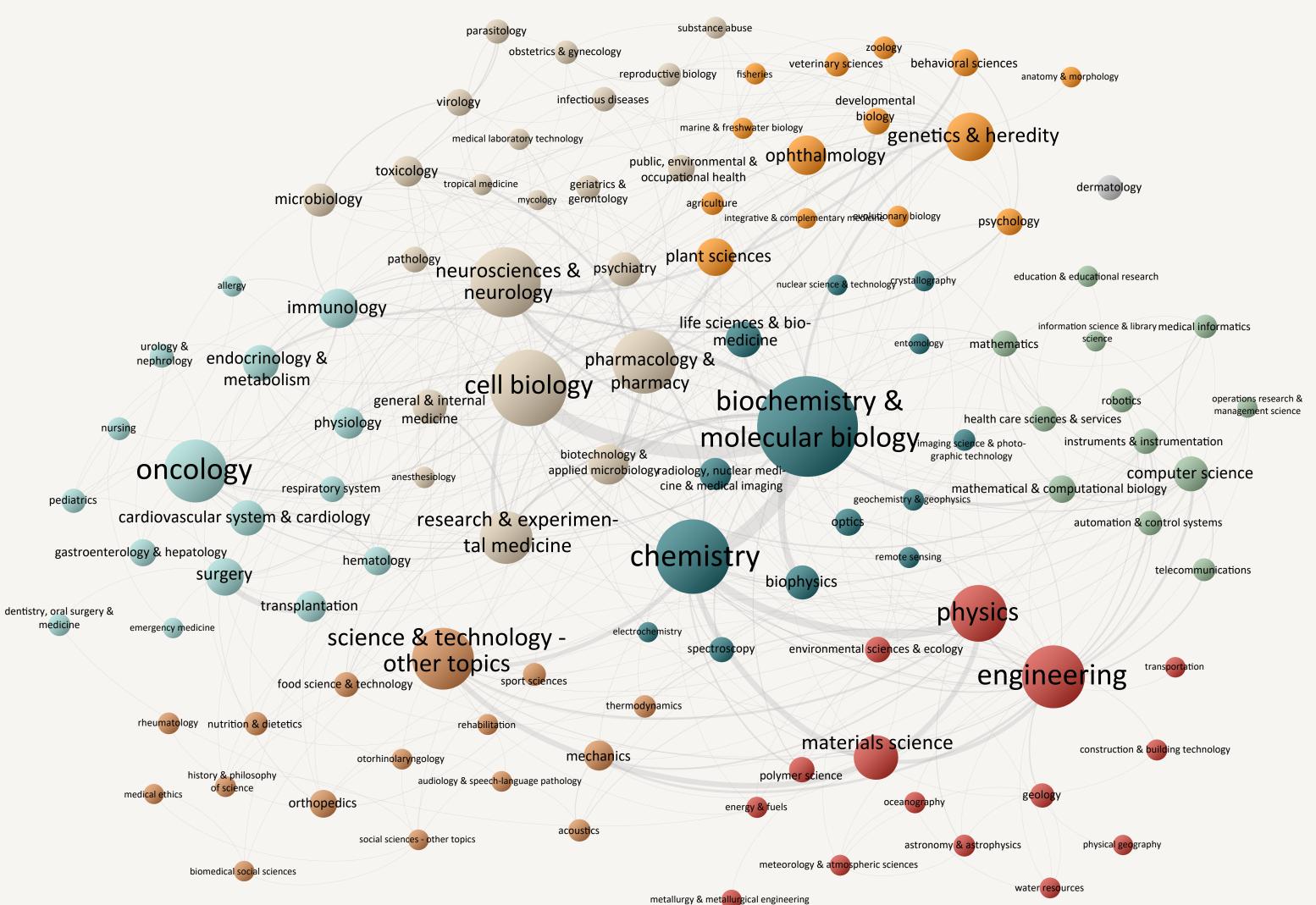


EBSCO's Library and Information Science Source (LISS)

linked to library instruction-related topics.



Research area co-occurrence in translational science, 2019



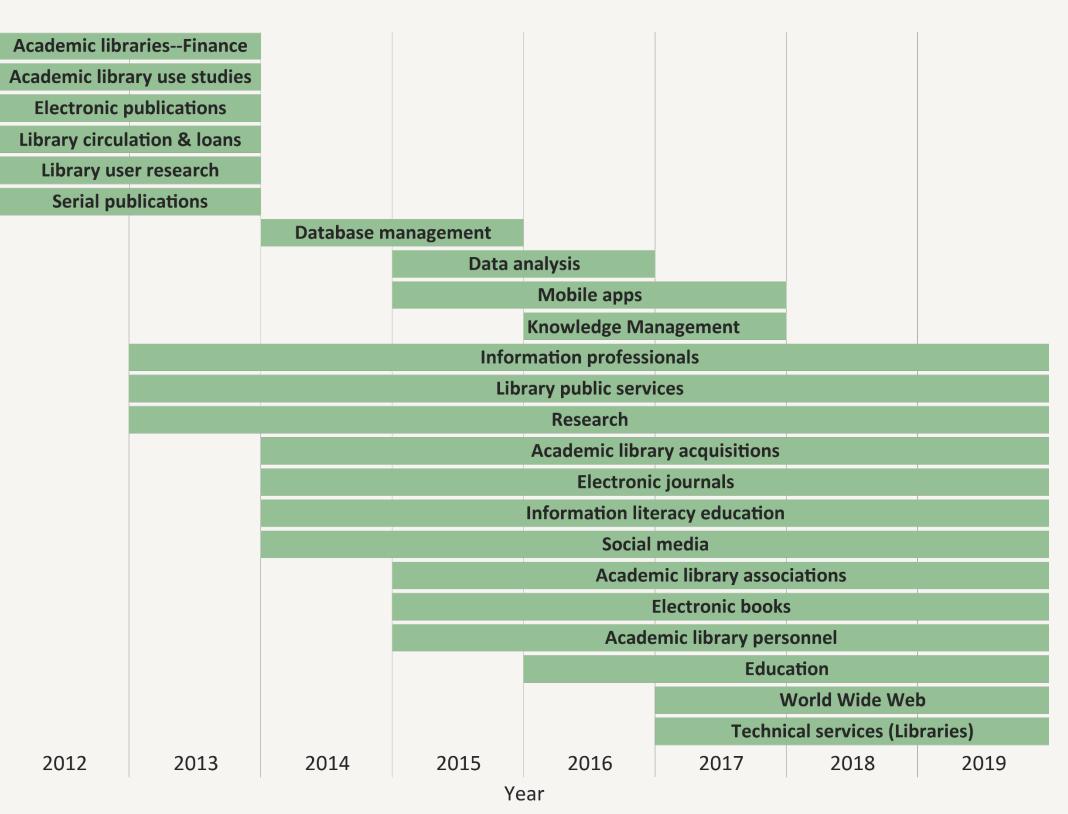
Translational research is a form of applied research that aims to transform basic scientific research into clinically useful results. A degree program in Translational Life Science & Technology (TLST) was launched at the Universities at Shady Grove in late 2018, the first of its kind in the United

this emerging field and how they relate to one another, a search in Web of Science for all records containing the term "translational" as a title, keyword, or abstract terms was conducted for 2019. 117 subject categories for 9,698 results were analyzed and visualized using the same tools and software that were outlined in the previous example. A similar analysis with author-designated keywords was conducted as well.

In order to understand the research areas in

Seven clusters illustrate the disciplinary structure of translational research. Most records connect pure and applied science, with significant groups of papers connected to applied disciplines such as engineering, materials science, and computer science. Dozens of medical research areas are also represented.

Keyword burst detection in academic librarianship, 2012-2019



Burst detection algorithms identify periods of increased activity over time. The Sci2 tool uses Kleinberg's burst detection algorithm to identify any keywords in the academic library paper set that occurred with more frequency than would be statistically expected for a given year. The algorithm identified 156 instances of increased keyword usage from 2012-2019. The twenty-three most frequently used terms are pictured at left. Along with the expected technology-related terms, there are a few surprises on this list, including terms related to library associations and public services.

PRELIMINARY GUIDELINES FOR PRACTITIONERS



Learn the basics of network analysis.

A conceptual understanding of network structures and behaviors (centrality, clustering, etc.) is essential.



Be flexible in developing a narrative.

The focus of a visualization will often shift after the data are analyzed.



Budget time for data cleaning.

Surveys of data scientists show that data cleaning can take anywhere from 15% (Mooney 2018) to 80% (Crowdflower 2016) of their time. Master a tool or language that allows you to convert data between formats, as many tools and libraries only accept specific formats (ISI, xml, tab-delimited, etc.).



Document the process.

With the concurrent use of many tools, even an hour's work without documentation can sometimes make it impossible to retrace your steps.



Experiment with multiple tools.

In addition to Python and R libraries, tools for bibliometric analysis and network visualizations include CiteSpace, the Science of Science (Sci2) tool, VosViewer and Gephi. Applications are frequently upgraded or discontinued (Moral-Munos 2019).



Consult with subject matter experts.

Patron communication for bibliometric projects typically extends well beyond a single reference interview. This is an iterative process, and a subject matter expert can help to distinguish between the obvious and the illustrative.

REFERENCES

Crowdflower (2016). 2016 Data Science Report. Online at https://visit.figure-eight.com/rs/416-ZBE- 142/images/CrowdFlower DataScienceReport 2016.pdf

Franssen, T., & Wouters, P. (2019). Science and its significant other: Representing the humanities in bibliometric scholarship. Journal of the Association for Information Science and Technology, 70(10), 1124-1137. doi: 10.1002/asi.24206

MacDonald, K. I., & Dressler, V. (2018). Using citation analysis to identify research fronts: A case study with the internet of things. Science & Technology Libraries, 37(2), 171-186. doi: 10.1080/0194262X.2017.1415183

Presented at the SLA Annual Conference. Online at https://www.sla.org/wp-

Moral-Munoz, Jose A., et al. "Science Mapping Analysis Software Tools: A Review." Springer Handbook of Science and Technology Indicators. Springer, Cham, 2019. 159-185.

Mooney, P. (2018). 2018 Kaggle Machine Learning and Data Science Survey. Online at https://www.kaggle.com/paultimothymooney/2018-kaggle-machine-learning-data-science-survey.

Tregoning, J. (2018). How will you judge me if not by impact factor? Nature, 558(7710), 345. doi: 10.1038/d41586-018-05467-5.

Network Legend **Node Size:** Total number of word occurrences for Edge thickness: Weighted degree of connectivity. a single key word or phrase. 490 **Node colors:** Differentiate topical clusters

Makar, S., & Trost, A. (2018). Operationalizing Bibliometrics as a Service in a Research Library. content/uploads/2018/06/Operationalizing-Bibliometrics.pdf.

Presented at the UMD Libraries Research and Innovative Practice Forum, June 2020