122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Paper ID #12745

A different view: An inquiry into visualization of bibliometric data

Mrs. Daniela Solomon, Case Western Reserve University

Daniela Solomon is Research Services Librarian for Biomedical Engineering, Civil Engineering, Electrical Engineering and Computer Science, Materials Science and Engineering, Mechanical and Aerospace Engineering at Kelvin Smith Library, Case Western Reserve University. She is interested in bibliometrics, altmetrics, data management, and library instruction.

A different view: An inquiry into visualization of bibliometric data

Abstract

Bibliometrics is a set of methods to quantitatively analyze scientific literature. The publication of the Science Citation Index has facilitated the development of bibliometric methods, and that development continued as computing power increased. Recently, bibliometric methods have been applied to measuring the impact of a work or of a scholar up for tenure or promotion; to facilitating the discovery of topical, spatial, and temporal scientific trends; to determining valuable resources for library collection development; and to evaluating students' information literacy skills. Unfortunately, processing, analyzing, and interpreting bibliometric data remains a challenge. Applying information visualization techniques to bibliometric maps creates graphical representations that enhance users' understanding of network structure. With the increasing availability of sophisticated tools that allow for easy data analysis, mapping bibliometric data has become more accessible to librarians.

This paper examines the value of visualizing bibliometric data. Specifically, it relates the experience of using the Sci2 Tool to analyze the data collected for a citation analysis project and to visualize the co-authors, co-citation, and bibliographic coupling networks revealed by that analysis. Furthermore, it discusses the challenges encountered while using the Sci2 Tool for data analysis, reflects on the benefits of visualizing bibliometric data, and future plans.

Introduction

As every librarian knows, department and subject knowledge are requisite for a successful liaison practice. This was not the case, however, when I became the Civil Engineering (CE) department's liaison: I was unfamiliar with the resources specific to the discipline and the department was not communicative. At the time, the CE department's webpage had not been updated in some time, so very little accurate information about the department was available. To further complicate matters, the nine faculty members had minimal (if any) interaction with the library: no purchasing or subscription suggestions, no cancelation complaints. The same was true for the CE department students. Because these conditions made collections management and the development of any liaison relationship difficult, it became clear that I must adopt a different approach in order to learn about the faculty members' research interests, productivity, scientific collaborations, and impact.

Performing citation analysis of the publications authored by the faculty members appeared an obvious first choice. With this background information in mind, I began the study of the Civil Engineering (CE) department in fall 2013 with the retrieval of publications by the faculty from the Web of Science Core Collections (WOS) using the "Author Finder" feature in combination with additional filters: "Technology" as the "Research Domain," "Case Western Reserve University" as "Institutional Affiliation," and "2007-2012" as "Time Period." The search yielded sixty-two articles with a combined total of approximately 3000 cited references. Upon exporting all of the cited references from each article to EndNote, I noticed that only 555 resources actually appeared in the citation management tool. Because such a significant discrepancy between the two numbers would skew the results of the analysis, it became clear that I must find a different tool capable of importing and analyzing complete bibliographic data for these publications in order to produce the desired results.

After researching the options available, I initially selected two free, open-source tools for consideration: Bibexcel and the Sci2 Tool. Although Bibexcel is designed specifically to analyze bibliographic data, it was difficult to use because the interface was not intuitive and available documentation was difficult to follow. The Sci2 Tool, on the other hand, worked well with bibliographic data extracted from ISI Web of Science and

Scopus, as well with other types of data formats. While it was easier to use in this regard, it did present different challenges along the way.

The primary challenges involved understanding how to prepare the data for uploading into the software and why specific types of networks could not be extracted directly. At that time, the software's documentation covered a previous version and was, as a result, often confusing. Additionally, some parts of the documentation were missing altogether which added to the overall difficulty of learning how to use it. Once the data was uploaded, however, it was relatively easy to extract some types of networks and to visualize them.

The goal of this study is to corroborate whether by conducting a comprehensive bibliometric study in conjunction with visualization techniques could provide liaison librarians invaluable information about a department.

Literature review

Bibliometrics is a set of methods to quantitatively analyze scientific literature using bibliographic data (De Bellis, 2009). Depending on its scope, a comprehensive bibliometric study may include a descriptive analysis, author's or journal's productivity analysis, scientific collaborations analysis, and citation analysis (Andres, 2009).

Citation analysis involves the systematic, quantitative study of cited references, based on the assumption that a citation reflects how an author is influenced by the work of another author and that each reference makes equal contribution to the citing article (Osareh, 1996). Garfield's revolutionary idea on the role played by bibliographic citations in the knowledge production process (1955) and Price's idea (1965) that scientific papers are altogether part of network initiated the study of citation relationships that culminated in the publication of the Science Citation Index. Despite recognized potential problems (MacRoberts & MacRoberts, 1989), citations became widely used in research evaluation (Moed, 2005). Specifically, they can be used for different purposes (Ashman, 2009), like a collection development tool (Hoffmann, 2012) and evaluation of students' information literacy skills (Denick, Bhatt, & Layton, 2010).

Garfield also conceptualized citation relationships as direct lines connecting later documents with earlier ones, initiating the idea of citation mapping (Garfield, 1983). Science mapping (or bibliometric mapping) is a "spatial representation of how disciplines, fields, specialties, and individual documents or authors are related to one another" (H. Small, 1999). Most of the elements of the bibliographic information could be used to create networks that can be analyzed to determine relations such as co-occurrence, coupling, and direct linkage.

One method employed to investigate citation networks involves using strategies common to social network analysis (SNA). SNA is based on the premise that, in a social network, relations between relevant nodes and the patterns formed by these relations are the most important (Marin, 2012). As such, SNA can be used to investigate the relations and relations patterns between the nodes consisting of bibliographic information (i.e. individual documents, authors, keywords, organizations, cited references, etc.).

The two most important SNA methods applied to bibliographic data include bibliographic coupling (Kessler, 1963) and co-citation (H. Small, 1973)--concepts that help determine the intellectual structure of a scientific research field. Since its introduction, others proposed variants or enhancements to Small's original co-citation concept, such as journal bibliographic coupling (H. G. Small & Koenig, 1977), author co-citation (White & Griffith, 1981), and journal co-citation (White & McCain, 1998). Another common analysis type using SNA is the co-author or co-authorship network which looks at authors and their affiliation to study social structures and collaboration networks (Peters & Van Raan, 1991).

Science maps can easily become very complicated and difficult to follow. Application of visualization techniques to the bibliometric maps facilitates gaining new depths of understanding into network's structure and uncovers relations between the nodes otherwise difficult to observe. With the increase in computing power, many software tools for science mapping analysis became available (Cobo, Lopez-Herrera, Herrera-Viedma, & Herrera, 2011).

One such tool is the Sci2 Tool, "a modular toolset specifically designed for the study of science [that] supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets" (Sci2 Team, 2009). Sci2 can create, analyze, and visualize three types of networks: collaboration networks, semantic networks, and publication citation networks. As applied to bibliometric analysis, collaboration networks show the collaboration between authors, institutions, or countries in the production of scientific research; semantic networks analyze the occurrence of certain words in a set of publications; and publication citation networks show the relationships among scientific publications based on their citations.

Methods

The study resumed in spring 2014. Once again, I retrieved the sixty-two articles from Web of Science and added them to "Marked List." This allowed for a quick analysis of the whole set of publications using the bibliometric tools available in Web of Science. The results included citations per article, h-index, top publications venues, WOS top categories, collaborators, and funding organizations. However, this analysis did not reveal any information on who are the collaborators, what research topic is each faculty publishing on, what resources are more important for their research, or any details on faculty members' information seeking behavior. Sci2 Tool looked promising in this regard.

Next, the records on the "Marked List" were exported to "Other File Formats" with "All Fields Available" selected. The result was a .txt file that was converted to an .isi file and uploaded into the Sci2 Tool. Once the data was available in Sci2 Tool, it was possible to extract several networks: co-authors, bibliographic coupling, and co-citation networks. Using the exploratory data analysis and visualization tool for graphs and networks GUESS that is integrated within the Sci2 Tool, it was also possible to visualize these networks.

To increase general understanding, the networks were weighted, meaning that the nodes size and the edges thickness, as well as their color, changed based on specific criteria.

Results

The WOS features "Citation Analysis" and "Create Report" used for the group of sixty-two articles produced the first set of results. The number of published items each year (Fig. 1) increased from five in 2007 to fifteen in 2012, demonstrating an increase in research activity for the period of this analysis. The sixty-two identified publications by CE faculty members included a total of 198 citations, showing a significant increase from 2008 to 2013 and a slight decline in 2014. Analysis by document type lists 50 articles, 10 proceedings papers, and 2 review papers that could demonstrate a strong interest in publishing journal articles and less interest in participating to conferences. The top ten publication venues (Fig. 2) includes titles specific to CE field as well as some unexpected titles like SPIE Proceedings, IEEE conferences, and IBM journal. All of the top ten WOS categories for the sixty-two publications (Fig. 3) align with the research interests listed on the departmental website, with the exception of "Optics" which relates to the publications found in SPIE Proceedings. The analysis based on authors' organization affiliation demonstrates limited external collaboration within the US (Fig. 4a) and internationally (Fig. 4b). The analysis of funding agencies (Fig. 5) indicates that only 41% of the publications have included information on the funding agency: the majority mentions the NSF, while the remaining mention other governmental and local agencies.

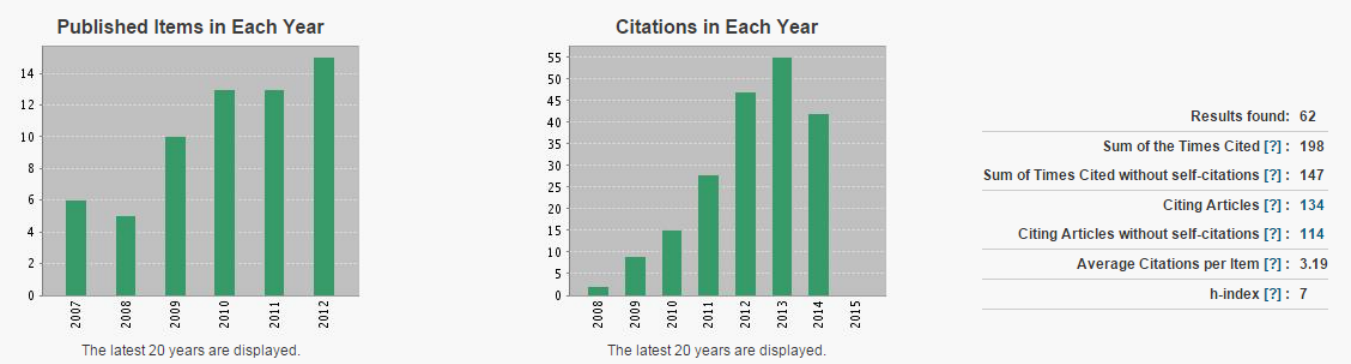

Fig. 1. Citation Analysis

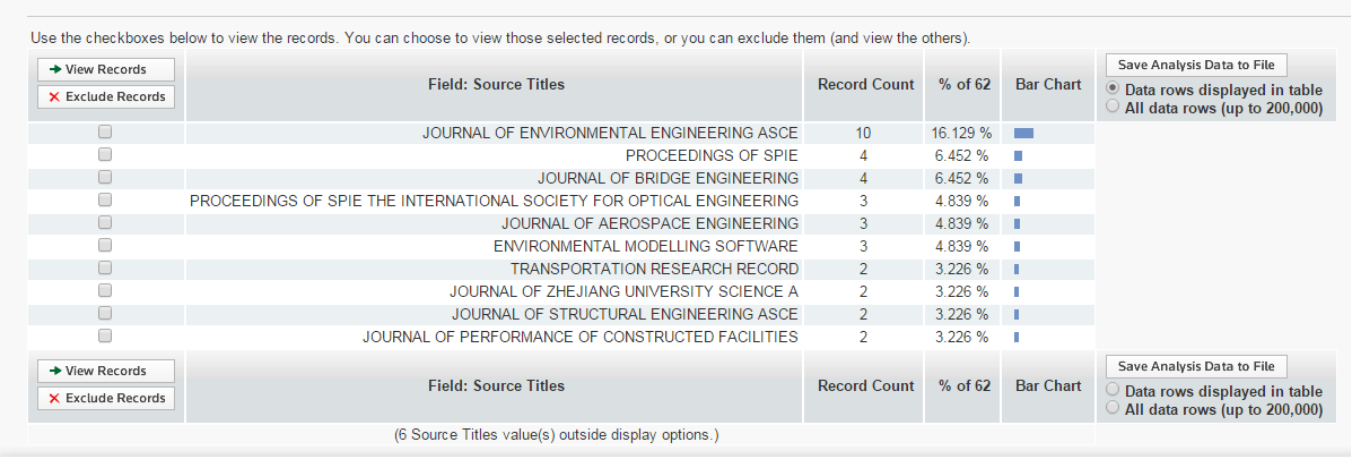

Fig. 2 Top ten publication venues

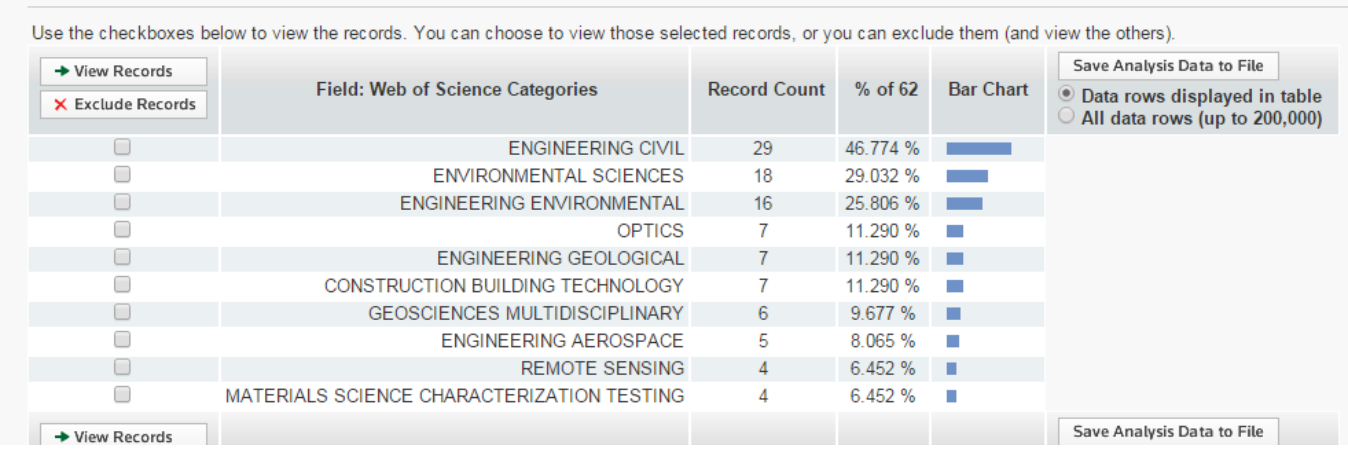

Fig. 3 Top ten WOS Categories

Use the checkboxes below to view the records. You can choose to view those selected records, or you can exclude them (and view the others).

View Records / Exclude Records	Field: Organizations	Record Count	% of 62	Bar Chart	Save Analysis Data to File
☐	CASE WESTERN RESERVE UNIV	60	96.774 %		Data rows displayed in table / All data rows (up to 200,000)
☐	NASA	4	6.452 %		
☐	SUNY BUFFALO	3	4.839 %		
☐	ZHEJIANG UNIV	2	3.226 %		
☐	URS CORP	2	3.226 %		
☐	UNIV ALASKA FAIRBANKS	2	3.226 %		
☐	SIMPSON GUMPERTZ HEGER INC	2	3.226 %		
☐	RE WARNER ASSOCIATES	2	3.226 %		
☐	MALCOLM PIRNIE INC	2	3.226 %		
☐	CONESTOGA ROVERS ASSOCIATES INC	2	3.226 %		
View Records / Exclude Records	Field: Organizations	Record Count	% of 62	Bar Chart	Save Analysis Data to File / Data rows displayed in table / All data rows (up to 200,000)

(2 Organizations value(s) outside display options.)

Fig. 4.a. Authors Organization affiliation in US

Use the checkboxes below to view the records. You can choose to view those selected records, or you can exclude them (and view the others).

View Records / Exclude Records	Field: Countries/Territories	Record Count	% of 62	Bar Chart	Save Analysis Data to File
☐	USA	62	100.000 %		Data rows displayed in table / All data rows (up to 200,000)
☐	PEOPLES R CHINA	5	8.065 %		
☐	HUNGARY	2	3.226 %		
View Records / Exclude Records	Field: Countries/Territories	Record Count	% of 62	Bar Chart	Save Analysis Data to File / Data rows displayed in table / All data rows (up to 200,000)

Fig. 4 b. Authors Organization affiliation worldwide

Use the checkboxes below to view the records. You can choose to view those selected records, or you can exclude them (and view the others).

View Records / Exclude Records	Field: Funding Agencies	Record Count	% of 62	Bar Chart	Save Analysis Data to File
☐	NATIONAL SCIENCE FOUNDATION	15	24.194 %		Data rows displayed in table / All data rows (up to 200,000)
☐	U S DEPARTMENT OF TRANSPORTATION	5	8.065 %		
☐	NASA	3	4.839 %		
☐	OHIO DEPARTMENT OF TRANSPORTATION	2	3.226 %		
☐	GREAT LAKES PROTECTION FUND	2	3.226 %		
☐	FEDERAL HIGHWAY ADMINISTRATION	2	3.226 %		
View Records / Exclude Records	Field: Funding Agencies	Record Count	% of 62	Bar Chart	Save Analysis Data to File / Data rows displayed in table / All data rows (up to 200,000)

(37 records(59.677%) do not contain data in the field being analyzed.)

Fig. 5 Analysis by Funding Agencies

More interesting results were obtained by applying visualization techniques to the networks extracted using Sci2 Tool--co-authorship, bibliographic coupling, and co-citation networks.

Co-authorship network represents the size and structure of the collaborators network for each author in the study. For this network, size of the nodes depends on the number of paper authored by a specific author, and the thickness of the edges depends on the number of common papers between two authors. The co-authorship network (Fig. 6. a.) suggests that the CE faculty rarely collaborate with one another; instead, they tend to collaborate within their own research group. This visualization also made it apparent that consistency in authors' names is important. After identifying and correcting the name variants for the each author, the corrected co-author network was created (Fig. 6. b.). As it can be observed in this network, consolidating name variants into one modifies nodes size and gives a better image of the author's collaborators network.

In the bibliographic coupling network (Fig. 7), two documents are coupled if they both cite a third document: the nodes represent the documents and the edges represent the common references. This network is also weighted: the more common cited references are between the documents, the thicker and darker the edges are. The bibliographic coupled network and the co-citation networks reveal the similarities among the articles' topics, as is this case with the articles on worldwide regulations. This network analysis helps identify the topics that specific authors are working on and, more importantly for collection development, the research interests of a given department are. A co-citation network (Fig. 8) represents the occurrence of two papers being jointly cited by a third paper: the nodes represent the documents and the edges represent the relationship with a different document. In this network, nodes size and edges thickness depend on the relationship strength. The network helps identify the most cited documents; as it is revealed by this network, the theses of graduate students from the department are often cited.

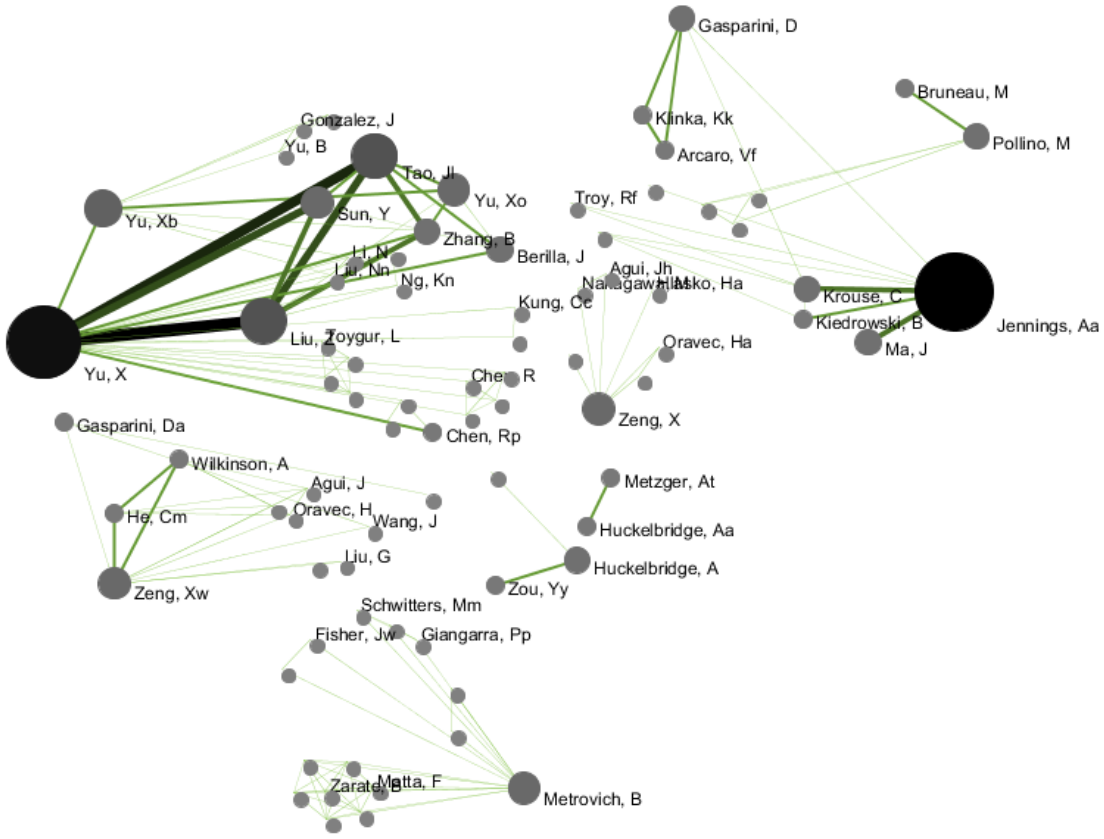

Fig. 6.a. Co-authors network

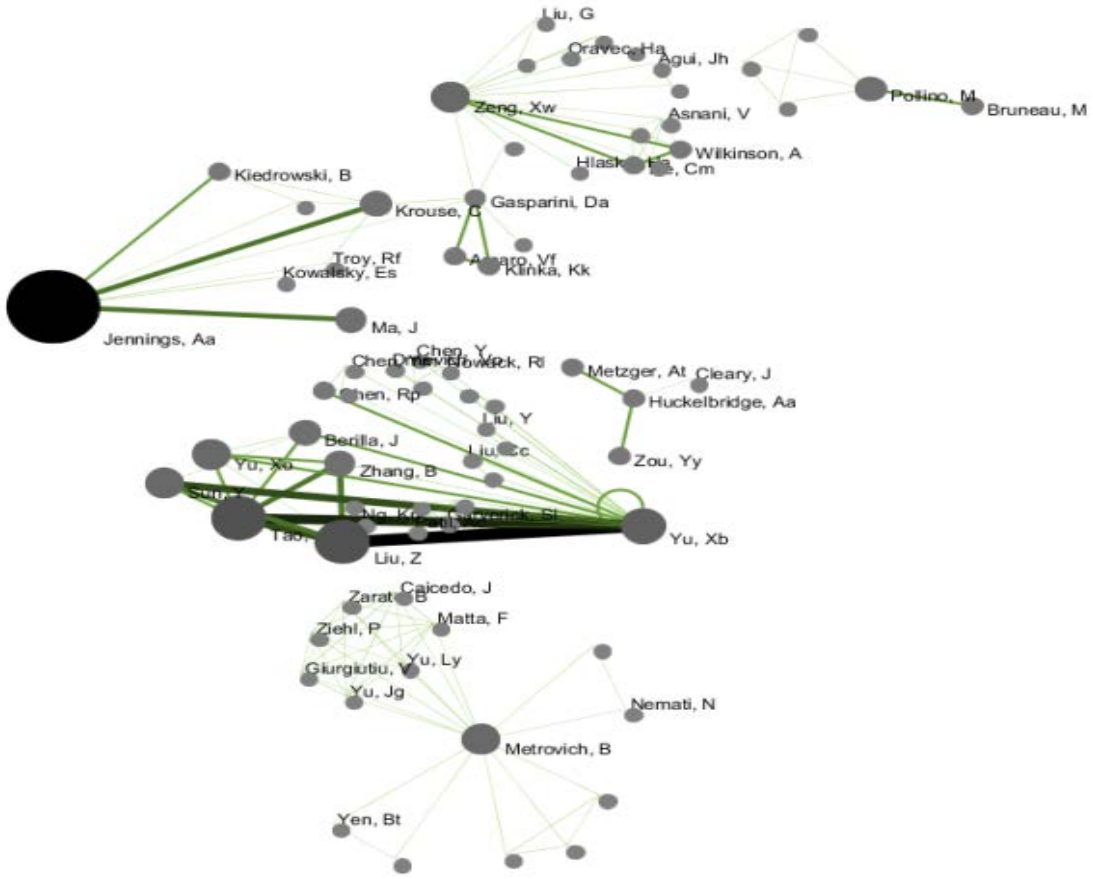

Fig. 6. b. Corrected co-authors network

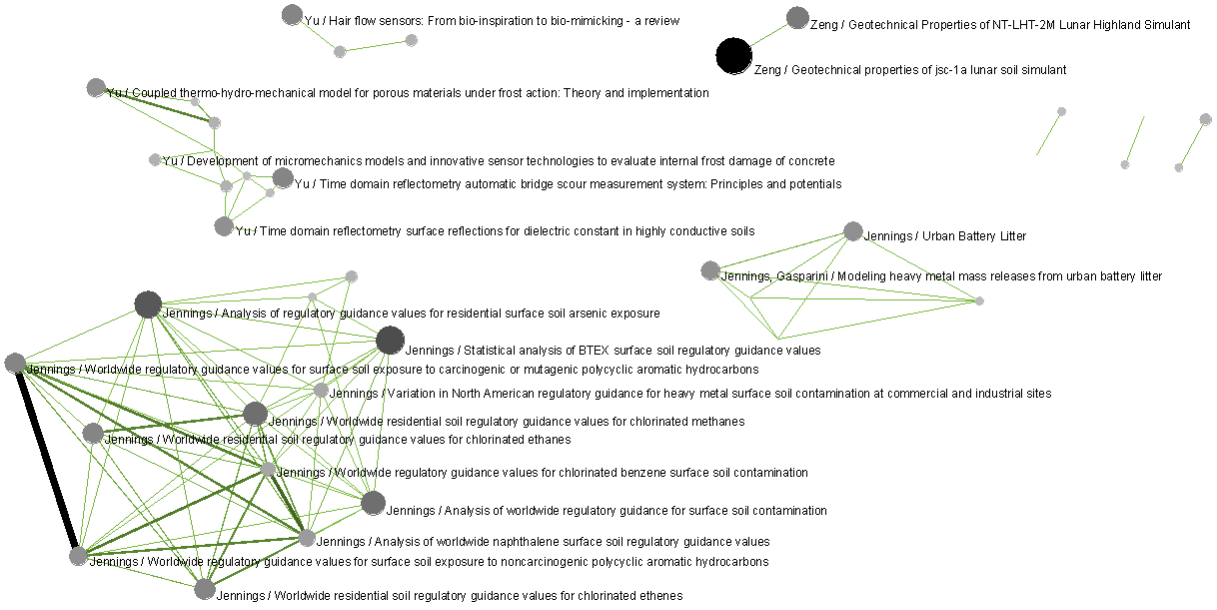

Fig. 7. Bibliographic coupling network with edges above 2.0 by weight & isolates removed

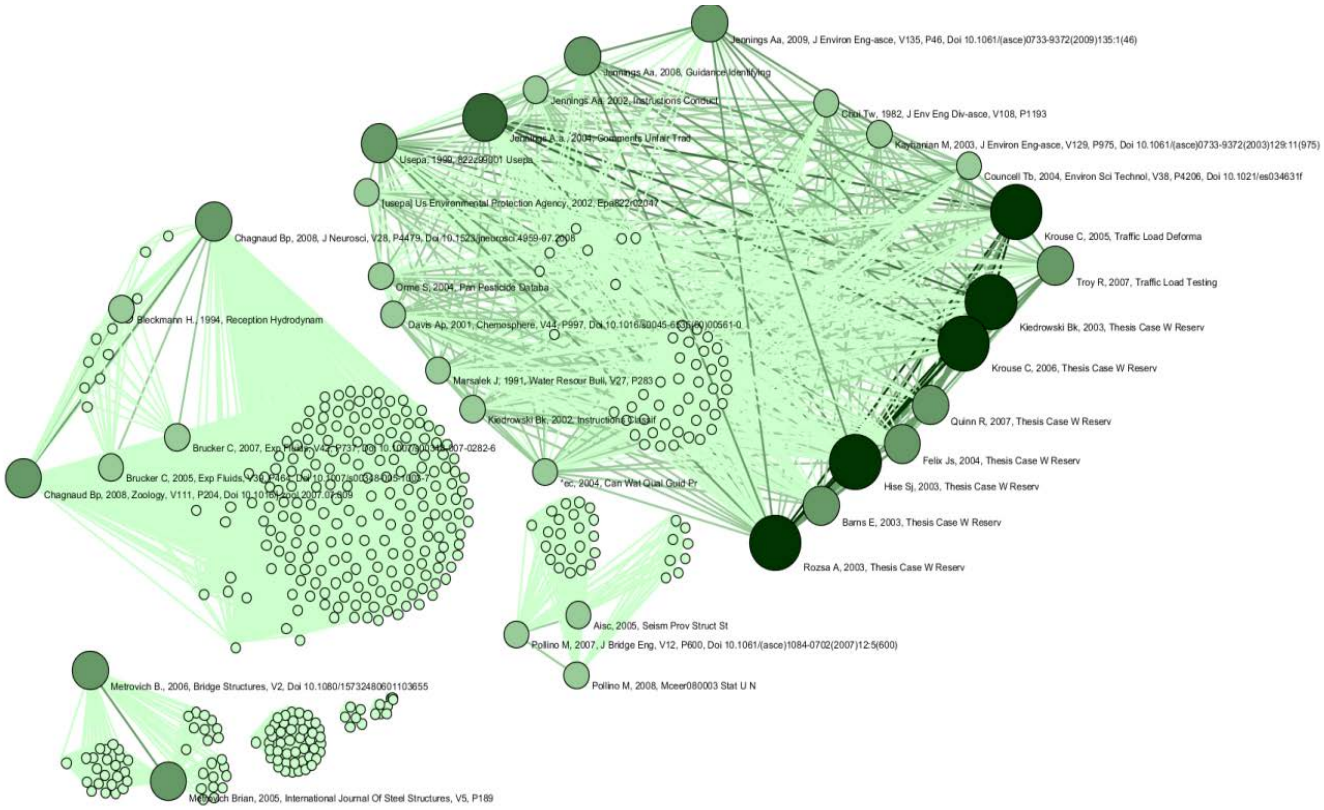

Fig. 8. Co-citation network

Conclusion

A comprehensive bibliometric study of faculty publications creates a multifaceted overview of a department's activity that offers the liaison librarian more insight into the faculty members' information seeking behavior than a basic citation analysis. The information generated could be used to make collection management decisions, to develop liaison relationships, to identify information needs, or to reveal outreach topics specifically related to the group of faculty included in the study. The value of having this information outweighs any challenges encountered and the time required to complete the project.

Collection management relies on a liaison librarian's familiarity with and understanding of faculty research interests, publications, and resources used for research and teaching. In this case, the bibliometric study confirmed the currency of the research interests listed on the department's webpage by matching them with the WOS categories of analyzed publications. Such information could prove useful when selecting resources to add or exclude from the library collections. This study demonstrated how important the ASCE journals are to CE faculty, providing the reason to protect them from subscription cuts. Furthermore, it revealed an upward trend in publication numbers over the selected time period. It also indicated that the number of published scholarly articles was overwhelmingly higher than any other types of publications, including conference proceedings. This preference seems to indicate that journals are the favorite type of resources; however, this conclusion is dismissed by the large number of conference papers cited, demonstrating that conferences are also very important. It could be inferred that the difference between the number of published conference papers and the number of cited conferences may be due to limited traveling funds, and by extension, limited research funds, or that for tenure and promotion more weight is being given to articles than conference papers

Moreover, bibliometric studies confirm the preferred types of resources for a given department. Publications included in this study referenced most often websites for domestic and foreign governmental agencies related to pollution. This correlated with the number of articles that analyzed worldwide pollution policies. The other types of documents referenced in these articles include journal articles, proceedings, technical reports, theses, books, and standards which confirmed my expectations for the types of resources used in an engineering program. Analysis of included citations revealed discrepancies in ASCE citation style knowledge: while the entries for journals and books were complete most of the time, citations for proceeding papers, technical reports, and websites were not as clear or complete. As a result, these entries were more difficult to locate. By making such an observation, a librarian can better advise faculty on the correct citation format for these types of documents and introduce them to WebCite, the on-demand archiving system for websites and other Internet-accessible digital objects.

In addition, bibliometric studies can support librarians' consultations with the faculty about collaboration networks, name consistency, and funding. As demonstrated, CE faculty members do not often collaborate with one another or with other faculty or researchers at other institutions in the US and abroad. Instead, their co-authorship is limited to students in their research groups. Likewise, librarians could use the co-author network visualization to show the importance of name consistency and the tools to use for name disambiguation. In this case, the co-citation map revealed that local "theses"--used here as a generic term for both theses and dissertations--appear very important, which may demonstrate that the department is more teaching-oriented than research-oriented. The perceived value of theses published by the department identifies graduate students as target group for outreach and library-related educational programs. Similarly, the insight into funding organizations gleaned from the publications analyzed points librarians to those which are relevant to the discipline and should be kept on their radar for resources about data management plans and public access requirements.

The visualizations of bibliometric networks extracted using Sci2 Tool provided an easier way to understand bibliographic data and to identify outreach opportunities. Analyses of a variety of networks have the potential to offer additional insights. Analysis of journals co-citation network seems to be most promising as it helps differentiate frequently used sources from those with little usage and guide the collection management process. Furthermore, collected data can be used to create additional visual representations to utilize in outreach efforts.

References

Andres, A. (2009). *Measuring Academic Research: How to Undertake a Bibliometric Study*. Oxford, Cambridge, New Dehli: Chandos Publishing.

Ashman, A. B. (2009). An Examination of the Research Objectives of Recent Citation Analysis Studies. *Collection Management, 34*(2), 112-128. doi: 10.1080/01462670902725885

Cobo, M. J., Lopez-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. *Journal of the American Society for Information Science and Technology, 62*, 1382-1402. doi: 10.1002/asi.21525

De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, Maryland ; Toronto; Plymouth, UK: The Scarecrow Press, Inc.

Denick, D., Bhatt, J., & Layton, B. E. (2010). *Citation Analysis of Engineering Design Reports for Information Literacy Assessment | iDEA: DREXEL LIBRARIES E-REPOSITORY AND ARCHIVES*. Paper presented at the American Society for Engineering Education Conference, Louisville, Kentucky. https://idea.library.drexel.edu/islandora/object/idea%3A3475

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science, 122*(3159), 108-111. doi: 10.1126/science.122.3159.108

Garfield, E. (1983). *Citation Indexing, Its Theory and Application in Science, Technology, and Humanities*. Philadelphia, US: ISI Press.

Hoffmann, L. D., Kristin; . (2012). A Review of Citation Analysis Methodologies for Collection Management. *College & Research Libraries, 73*, 321-335. doi: 10.5860/crl-254

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10-25. doi: 10.1002/asi.5090140103

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science, 40*(5), 342-349. doi: 10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U

Marin, A. W., Barry. (2012). Socal network analysis: an introduction In J. C. Scott, PJ. (Ed.), *The SAGE handbook of social network analysis* (pp. 11-25). Los Angeles, London, New Dehli: SAGE.

Moed, H. (2005). *Citation Analysis in Research Evaluation*. Dordrecht, [Great Britain]: Springer.

Osareh, F. (1996). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature I : Libri. *Libri, 46*, 149-158. doi: 10.1515/libr.1996.46.3.149

Peters, H. P. F., & Van Raan, A. F. J. (1991). Structuring scientific activities by co-author analysis. *Scientometrics, 20*(1), 235-255. doi: 10.1007/BF02018157

Price, D. J. d. S. (1965). Networks of Scientific Papers. *Science, 149*(3683), 510-515. doi: 10.1126/science.149.3683.510

Sci2 Team. (2009). Science of Science (Sci2) Tool Indiana University and SciTech Strategies. Retrieved from https://sci2.cns.iu.edu

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269. doi: 10.1002/asi.4630240406

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science, 50*(9), 799-813. doi: 10.1002/(sici)1097-4571(1999)50:9<799::aid-asi9>3.3.co;2-7

Small, H. G., & Koenig, M. E. D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management, 13*(5), 277-288. doi: http://dx.doi.org/10.1016/0306-4573(77)90017-6

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*(3), 163-171. doi: 10.1002/asi.4630320302

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327-355. doi: 10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-4