

Rete-Netzwerk-Red: Network Workbench – Analyzing and Visualizing Scholarly Networks Extracted from Online Bibliographic Resources

Micah Linnemeier, Timothy Kelley, Weixia Huang, Bruce Herr, Russell Duhon, Katy Börner

mwlinnem | kelleyt | huangb | bherr | rduhon | katy@indiana.edu

Cyberinfrastructure for Network Science Center, School of Library and Information Science
Indiana University, Bloomington, IN 47405, USA

Theme: Network Science Cyberinfrastructure

Introduction

Today, terabytes of scholarly data are available from different data providers.^{1,2,3} The datasets differ by type, cost, size, quality, coverage, contents, and ease of access. All data providers support search: some support local citation and coauthor network traversal. None supports the large-scale analysis, modeling and visualization of scholarly networks. This is where the Network Workbench (NWB) Tool^{1,5,6} comes in, online available at <http://www.cis.indiana.edu/~nwb/>. It allows users to extract co-author/co-inventor as well as direct citation, co-citation^{7,8} and bibliographic coupling networks⁹, from different data sources. In our talk for NetSci'08, we will detail how the NWB Tool can be used to study scholarly networks derived from publication and funding data.

Results

The NWB Tool provides easy access to more than 22 preprocessing, 23 analysis, 8 modeling, and 17 visualization algorithms. Recently, the NWB Tool has been extended to read publication data from Thomson Scientific/ISI, Scopus, Google Scholar via Publish or Perish¹⁰, personal references collected via reference management software such as Endnote (<http://www.endnote.com>), Reference Manager (<http://www.refman.com>), or the Bibtex format (<http://www.bibtex.org>), funding data downloaded from the National Science Foundation, and other scholarly data available in plain comma-separated value files. Extraction of networks from a dataset is easy: simply load the dataset, and run the extraction algorithm of your choice. The resulting network can be analyzed using diverse combinations of the many data analysis and visualization algorithms available in the NWB Tool. Common workflows include general network analysis, text analysis via burst detection¹¹, network visualization, and exporting to various standard network formats. Figure 1 shows a small section of a visualization generated using the network extraction workflow.

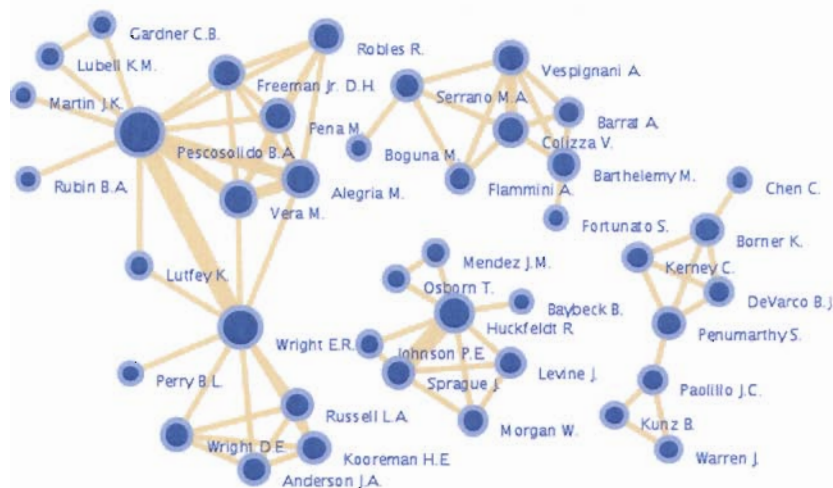


Figure 1: Four clusters from the co-authorship extraction of a search for "Indiana University" and "Social Networks" on Scopus, visualized using GUESS¹². Edge width corresponds to number of co-authored works, and node size corresponds to total degree.

Discussion and Outlook

Today, the NWB Tool provides basic functionality, and we welcome all feedback for extensions. In the future we plan to support more input formats, additional data cleaning, co-word analysis and network extraction, extraction for multi-type networks, integration of new visualization algorithms, exportation of bibliographic formats to a single common format, and additional support for scientometric workflows. Existing workflows will also be enhanced and expanded upon through a follow-up NSF project that aims to build a macroscope for science policy makers.

References

1. Thomson Reuters. 2008. ISI Web of Knowledge. <http://www.isiwebofknowledge.com/>. (accessed April 22, 2008)
2. Elsevier B.V. 2008. Scopus. <http://www.scopus.com/>
3. National Science Foundation. 2007. About Awards. <http://www.nsf.gov/awards/about.jsp> (accessed April 22, 2008)
4. Herr H, Bruce W., Weixia (Bonnie) Huang, Shashikant Penumarthy and Katy Börner. 2007. Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179). Annals of the New York Academy of Sciences, Boston, MA.
5. Weixia Huang, Bruce Herr, Russell Duhon, Katy Börner. 2007. Network Workbench – Using Service-Oriented Architecture and Component-Based Development to Build a Tool for Network Scientists, *Proceedings of the International Workshop and Conference on Network Science*.
6. NWB Team. 2006. Network Workbench Tool. Indiana University and Northeastern University. <http://nwb.slis.indiana.edu> (accessed April 22, 2008).
7. Marshakova, I. V. 1973. "A System of Document Connections Based on References." *Scientific and Technical Information Serial of VINITI* 6: 3-8.
8. Small, Henry. 1973. "Co-Citation in Scientific Literature: A New Measure of the Relationship Between Publications." *Journal of the American Society for Information Science*, 24: 265-269
9. Kessler, Michael M. 1963. "Bibliographic coupling between scientific papers." *American Documentation*, 14 no. 1: 10-25
10. Harzing, Anne-Wil. 1997-2008. Publish or Perish: A citation analysis software program. <http://www.harzing.com/resources.htm>. (accessed April 22, 2008).

- 11 Kleinberg, J. M. 2002 "Bursty and Hierarchical Structure in Streams" In *8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 91-101. ACM Press.
- 12 Adar, Eytan. 2007. Guess: The Graph Exploration System. University of Washington.
<http://graphexploration.cond.org/> (accessed April 22, 2008)