CITATION NETWORKS: A NEW HUMANITIES TOOL?

Almila Akdag[a] , Zoe Borovsky[b]

[a] *Department of Art History, UCLA, LA, USA*
[b] *UDHIG, UCLA, LA, USA*

Abstract: This paper proposes to use citation networks as a tool for mapping out links among disciplines and research areas in & around humanities. By collecting papers from main journals in diverse areas like cognitive science, art history and psychoanalysis, this study aims to build a small citation network that will be visualized and published as a 3d web page. The end-map will be the backbone of a dissertation, which tries to analyze the influence of various movements and disciplines onto the art historical canon. Used as such, a citation network can become more than a frozen map that has more than a one-time use value.

Access to online electronic databases and tools such as Google scholar have led to a significant improvement in the discovery of secondary literature. However, organizing the vast amount of bibliographic information from various discipline-specific databases continues to be an impediment to truly interdisciplinary work. My project attempts to construct virtual maps of electronic databases or digital libraries capable of providing scholars with significant links between disciplinary relations, interdisciplinary research areas, or tendencies, approaches and methodologies inside a single discipline. Moreover, as I will show, if the third dimension, namely history, is added to these maps, the transformation of the disciplines, the merging of research areas, and the changes in the taxonomic structure of the academy will be revealed to the expert eye. As an example of how such a virtual time-map could become a valuable analytic tool during the research process, I will construct such a tool and, using my dissertation as a test case, demonstrate how it can be applied to map intersections between art historiography, psychoanalysis and cognitive science.

As a PhD candidate in UCLA's Art History program, my dissertation traces the changes of the critical discourse from the 1970s, when the so-called "New Art History" clashed with traditional art historians, and gave rise to a whole new approach--one that has now become known as Visual Cultural Studies. However, in such a broad context, one needs to handle a deluge of texts and interrelations. A simple timeline, a linear outline constituted of chapters and subchapters is not enough to depict the map of overlapping layers, concepts and relations between opposite but -in this case still- allied methodologies. In order to render all these visible, I would like to create a multidimensional space of such relations. To that end I propose to collect significant papers, extract citation information using various text-analysis programs and visualize the end results as a citation network that runs along three different trajectories, namely the history of psychoanalytical and cognitive scientific methodologies and their impact on the evolution of New Art History into Visual Cultural Studies.

HISTORY: A relatively new research venue, called 'scientometrics' or 'bibliometrics', specializes in creating such maps for delineating growth, relations and interactions in scientific fields.[1] Bibliometrics uses text analysis to extract citation data from papers and makes use of this data as a way of evaluating scientific publications. To obtain information from scientific papers and using

[1] To read more on the history of scientometrics see Katy Börner, Jeegar T. Marus, and Robert L. Goldstone. 2004. *The Simultaneous Evolution Of Author And Paper Networks*, PNAS 101 (suppl.1): 5266-5273.

the results in mapping out scientific relations has a long history. As early as 1964 Garfield and his colleagues suggested using citation data to evaluate the development of science.[2] From 80's on, the research in this area accelerated with the advancement of computers and various combinations of statistical methods used to extract and evaluate information such as citations, co-citations with reference of various bibliometric data. The end-results are usually rendered as so-called 'citation networks' which are a variation of social networks.[3] Now it is a common practice to evaluate a scholar or a journal according to how many times it/he/she is cited. Moreover, there is so much research done using citation networks as a methodology to analyze the disciplines, many scholars are now questioning the efficacy of this approach.[4]

Among the ample publications in this area, two general approaches distinguish themselves: the citation networks are either built to support an idea or to enhance the way in which such networks are composed. In the first instance, a search is done to filter out unnecessary papers. The maps generated in this way are limited with a scope, and generally give an overview of the research topic. These maps are not created from a relational database and are not published for further use. That means they are not applicable to other research questions and the enormous work put into collecting papers and preparing them for network analysis is done on a case-by-case basis because access to the databases is restricted and therefore the datasets themselves cannot be made available. In the second instance papers are extracted from the electronic database without any filtering. Instead a time limit is imposed. These types of publications tend to focus more on the technical details and explore the mathematical substructure of social networks. Usually open-source databases are preferred, since the main idea is to test the application's performance on huge datasets. Papers about such studies do not interpret the resulting map and instead detail the mathematical innovation of the applications.

Despite the popularity of this approach in the sciences, I have yet to find a paper that uses Humanities databases. Even the most comprehensive citation network, a data set that encompasses "7,121 journals covering over 1 million documents in the combined Science Citation and Social Science Citation Indexes" does not delve into the Arts and Humanities Citation Index. [5] Rather than simply applying the same techniques to Humanities materials, a fresh approach, one that is not only suits to humanities scholarship, but addresses some of the issues raised in the scientific communities seems desirable.[6] Although my long-term goal

[2] All these search engines providing citation index information are products of Thomson ISI. The original foundation was called simply Institute for Scientific Information. Garfield launched it, again in 1964, see ibid.

[3] Doreian Patrick, *A Measure Of Standing Of Journals In Stratified Networks*. 1985. Journal of the American Society for Information Science, 8(5/6), 341-363.

[4] See R. E. Rice et al., Journal-To-Journal Citation Data. 1989, Scientometrics, Vol. 15~ No 3-4: 257-282; Lindsey D, Using Citation Counts As A Measure Of Quality In Science Measuring What's Measurable Rather Than What's Valid. 1989, Scientometrics, Vol. 15, No 3-4: 189-203; Nederhof, A. J., and Zwaan, R. A. *Quality Judgments of Journals as Indicators of Research Performance in the Humanities and the Social and Behavioral Sciences*.1991. Journal of the American Society for Information Science 42, no. 5: 332–40; Nederhof Anton J., *Bibliometric Monitoring Of Research Performance In The Social Sciences And The Humanities: A Review*. 2006, Scientometrics, Vol. 66, No 1: 81-100; Leydesdorff Loet, *Can Scientific Journals Be Classified in Terms of Aggregated Journal-Journal Citation Relations Using the Journal Citation Reports?* 2006, Journal Of The American Society For Information Science And Technology—March: 601-614

[5] Boyack Kevin W., Klavans Richard, Börner Katy, *Mapping The Backbone Of Science*. 2005. Scientometrics, Vol. 64, No. 3: 351.374

[6] Even though the end results of the "citation networks" give a scholar a good overview, they are still not integrated into the academic research facilities like Web of Science, Science Citation Index, Social Citation Index or Arts & Humanities Citation Index etc. The main reason is that once a citation network is derived, it becomes a static entity; it covers a limited time and scope. Thus citation networks, by their very definitions and aims, fail to keep up with new publications.

is to create a dynamic citation network that can become a part of an ongoing research project in digital humanities, in this paper I will focus on creating a static citation network using open-source tools. My research plan to achieve this aim is as follows:

1) Collecting papers: I have already collected around 2000 papers, mostly from prominent journals in Art History (such as Art Bulletin, October, Art Journal and Leonardo), and in Cognitive Science (Trends in Cognitive Sciences). Beside these resources I would like to include the classical texts in psychoanalysis; which can be found in electronic format in the Psychoanalytic Electronic Publishing database. The main criterion used for selecting pertinent papers is to use keywords that are relevant for the research topic. For example I used keywords such as "Freud", "psychoanalysis", "aesthetic", "artist" while searching in a cognitive science journal whereas I chose keywords like 'cognition", "cognitive science", "vision", "artist" etc. for the search in the domain of psychoanalysis.

2) Preprocessing the database for text analysis & extraction the needed information: The acquired corpus is preprocessed for text mining and analysis. Preparing a list of keywords and bibliometric data (author name, date, journal name, title, etc.) will be enough for the preprocessing stage. We will experiment with different text-mining and text analysis programs and report on those that work well at this stage of data preparation.

3) Construction of social networks: A social network is a graph representation of social relations. Graphs are the most popular and widely researched data structure for representing and processing relational data. In a graph, each node represents one entity (a person in a social network; a researcher or a work in a citation network) and the edges (or arcs, if they are directed) of the graph represent some relation. One can also indicate the strength of the relation by associating weights with the edges of the graph. Then, by using tools like Pajek[7], the graph nodes can be placed in a 3D space in such a way to minimize an energy term. Thus, the nodes that are close to each other semantically (through the interpretation of graph edges) are placed in proximity, even when they don't have actual links. On a social network, clusters and cliques can be identified, indirect relations can be uncovered, and relevance judgments can be made based on quantitative or qualitative measures. Even the location of the nodes (center vs. periphery) can be informative for a person thoroughly acquainted with the represented structure.

The use of such a graph tool is simple; nodes and arcs are read from a file, and the graph visualization is accomplished with a few commands. For a citation or semantic network, text mining tools can be employed to derive the entities in relation automatically. Once such a network is built, Pajek can import the graph in 3D file formats like 3xd and VRML; both are now becoming standards for internet publication in 3D.[8]

CONCLUSION

The proposed tool serves not only as a bibliographic aid, but will become the main framework for my dissertation. By incorporating the 3D virtual map into my dissertation, I hope to demonstrate a new form of digital scholarship--one that springs from the new possibilities that digital technologies affords scholars in the humanities whose work is inherently, and exuberantly, interdisciplinary.

[7] You can find more information about Pajek, its history and application areas at http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm, last accessed 11.15.2006.

[8] Please check http://www.w3.org/ to see the latest standards of World Wide Web in relation to 3d publishing, last accessed 11.15.2006.