

Mapping Sugar Sweetened Beverage Policy: TechMining Across SCOPUS and LexisNexis

Introduction

The study of citation patterns in scientific research has been a fruitful area of study in recent years. Many scientometric researchers have investigated networks of research publications and indicators in a variety of databases. Little research, however, has focused on citation patterns among legal publications (*See*, Shapiro & Pearse, 2006). Almost no research has looked at citation patterns between legal and scientific publications (*See*, Pasadeos et al. 2006). Since legal publications are housed in different databases than scientific publications, they are excluded from large citation studies like those done by Leydesdorff and colleagues (2015) that probe databases such as SCOPUS and Web of Science. While the two primary legal publication databases LexisNexis and Westlaw are owned by science-giants Elsevier and Thompson Reuters respectively, they operate very differently and are not optimized to allow exploration of network patterns among the articles. This research seeks to explore citation patterns on a specific, bounded topic—sugar sweetened beverages—across not only scientific research, but also legal research.

Using Burt's (1992) structural holes, which examines the position of actors across network gaps, and the newer area of cultural holes, which adds a cultural dimension through linguistic networks (Pachucki & Breiger, 2010), this research will use network analysis of citation patterns, as well as latent Dirichlet allocation (LDA) of article abstracts, to look at whether, in addition to structural divides in patterns of citations between legal academic and scientific publishing, there are also language and contextual differences. By investigating these key issues this research will not only expand the applications of these well validated scientometric techniques to new databases, but also will explore the intersection of two academic publication areas, the way they communicate, and how information across both is attempting to influence policy.

Methods

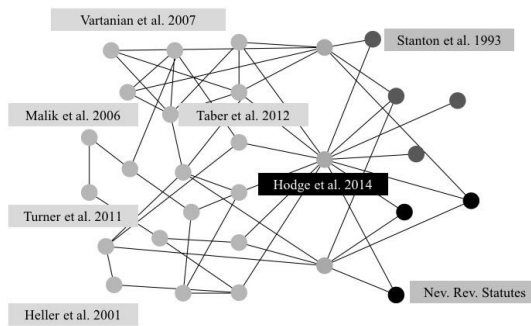
This research uses a census of all articles available on SCOPUS and LexisNexis that are retrieved using the search terms “sugar sweetened beverages.” This yields 1,281 articles on SCOPUS and 113 articles on LexisNexis. While SCOPUS article metadata has already been imported into Vantage Point, cleaned, and subject to initial cocitation analyses and multidimensional scaling, articles from LexisNexis are not easily translated into a format that is analysable. LexisNexis requires that all articles be exported together as one continuous text file. This research has and will continue to involve extensive cleaning of this text file to yield individual text files and spreadsheet data that is compatible with VantagePoint. Cleaning and curating of this text file and its data have been done both manually and through Python scripts. In order to assess the structural network of citation patterns, the data from the LexisNexis documents will be converted into a matrix format, and then will be combined with the article-author-journal-citation data from SCOPUS. All of the data will then be combined in Vantage Point. It will then be subjected to a cocitation analysis and multidimensional scaling, and structural holes will be calculated in UCINET.

In order to assess the cultural network of the text among the papers, the text of each article abstract will be cleaned using AutoMap to remove numbers, capitalizations, punctuation, and stop-words. Additional stop-words more related to the topic may be removed as the data are processed to increase clarity. The text files for each article will be run through MALLETT, or Machine Learning for Language Toolkit, a software program that implements LDA in order to generate a series of topic models. Using UCINET, I will create a matrix of the distribution of topics across documents and then calculate the gaps between clusters using cultural hole analysis.

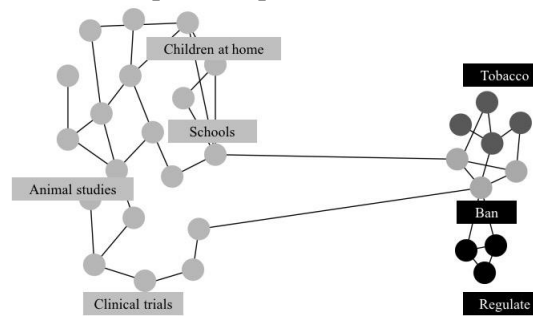
Predictions

Since this research is currently in progress, I have not completed all of the analyses. My prediction is that this analysis will likely show that while structurally many legal papers occupy a broker position between legal and scientific frameworks, their textual context is distinct and clusters exclusively with other legal academic papers (Figure 1).

Figure 1: Paper to Paper Citation Network



Topic to Topic Network



In the context of an author cocitation network, those occupying broker positions between clusters of similar authors are those who cite across disciplines and engage in what may be interdisciplinary research. These authors are in structurally interdisciplinary positions (Porter et al., 2007). By calculating the structural holes between clusters, and thus the positions of the brokers in the author cocitation network, we can identify these structurally interdisciplinary authors.

The LDA topic model of the text of the article abstracts will provide a cultural lens that will allow for an investigation of the content of the articles. While the linguistic cultural element of legal and academic papers could be explored in other ways (such as simply looking at the frequencies of descriptive versus prescriptive words), topic modeling allows for the text itself to generate a series of topics that may indicate not only different types of prescriptive or descriptive language as more associated with legal or scientific papers, but also other themes that may emerge and be strongly associated with one area. For example, legal scholars writing on sugar sweetened beverages may frequently compare SSBs to tobacco in a way that scientists do not.

Conclusion

Through examining the structural position of articles involved in policy advocacy as well as the textual cultural content of these articles, information can be gleaned about potential strategies for better communicating with policy intermediaries like legal academic scholars. Given that legal authors frequently cite to scientific papers, but the reverse is not true, it is likely that a structural network will show that legal papers frequently occupy broker positions across structural holes between scientific and legal journals.

References

- Burt, R. S. (1992). *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
- Leydesdorff, L., de Moya-Anegón, F., & de Nooy, W. (2015). Aggregated journal-journal citation relations in scopus and web of science matched and compared in terms of networks, maps, and interactive overlays. *Journal of the Association for Information Science and Technology*.
- Pasadeos, Y., Bunker, M. D., & Kim, K. S. (2006). Influences on the Media Law Literature: A Divergence of Mass Communication Scholars and Legal Scholars? *Communication Law and Policy*, 11, 179–204.
- Pachucki, M. a., & Breiger, R. L. (2010). Cultural Holes: Beyond Relationality in Social Networks and Culture. *Annual Review of Sociology*, 36(1), 205–224.
- Porter, A. L., Cohen, A. S., Roessner, J. D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, 72(1), 117-147.
- Shapiro, F. R., & Pearse, M. (2015). The Most-Cited Law Review Articles of All Time. *Michigan Law Review*, 110, 1483 – 1520.