TechMining LexisNexis: Challenges in Using Legal Journal Articles in Cocitation Analysis and Text Mining

Lexi	White	acwhite5@asu.edu	USA	Arizona State University, School of Human Evolution and Social Change

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 1,060 articles on SCOPUS and 125 articles on LexisNexis. The two primary legal databases (LexisNexis and Westlaw) were both considered, however the exporting documents on Westlaw is only available in PDF or WordDocument instead of text file. LexisNexis allows export to text file, but requires that all articles be exported together as one continuous text file. Researchers have written Python scripts that parse the LexisNexis exports to break down these giant text files to assist in data cleaning. Another major challenge is that LexisNexis legal articles do not use a traditional bibliography or reference list. Instead they use footnotes with a proprietary legal citation system (BlueBook) that utilizes references repeatedly across footnotes and combines aside information with substantive citation information. These footnotes need to be parsed to remove repeat references and additional information. Researchers used a combination of hand cleaning and Python scripts to pull citation data from the 125 articles to combine the data with SCOPUS data to run a cocitation analysis and LDA topic model of the abstracts.

Initial Findings

LexisNexis articles alter the profile of the corpus when included with SCOPUS articles. Disciplinarily, legal articles make up 7% of the entire corpus (after medicine, nursing, and biological sciences). 16% of legal articles are also written by students.

Initial LDA analyses show a high proportion of not only "legal" topics across LexisNexis articles (comprising words like "state, city, regulation, local, government, ban") but also topics that include

alternate substances of regulatory concern like "tobacco" and "alcohol". LexisNexis articles show little proportion of topics that focus on study design and details ("cross, sectional, sample, variable"; "meta, cohort, prospective, evaluate, evidence").

Conclusions

There remain many questions about the viability of including legal academic articles in large scale scientometric analyses. While this research has expanded the technical viability of including LexisNexis articles in analyses on scientific literature, the epistemological and cultural differences regarding the article review process and article motivation leave questions about whether or not legal academic articles are completely comparable to scientific academic articles. Regardless of these epistemological and cultural differences, exploring the inclusion of this large body of literature is important, because it offers a view of a related academic domain.

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.

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.