Lexical diversity as an indicator of journal scope

Philippe Mongeon[1], Maxime Sainte-Marie[2], and Marc-André Simard[3]

[1] *philippe.mongeon @ps.au.dk*
Aarhus University, Danish Centre for Studies in Research and Research Policy, Aarhus (Denmark)

[2] *maxime.sainte-marie@umontreal.ca*
Université de Montréal, École de bibliothéconomie et des sciences de l'information, Montreal (Canada)

[2] *marc-andre.simard.1@umontreal.ca*
Université de Montréal, École de bibliothéconomie et des sciences de l'information, Montreal (Canada)

Abstract

This work-in-progress proposes that the semantic similarity of articles published in a journal can be used to quantify the scope of the journal. Preliminary results of an analysis of a set of Library and Information Science (LIS) journals show that journals with a known broader scope have flatter, left-skewed similarity distributions, with lower averages, that can be used to identify the papers that best fit the journals in which they are published. Finally, we find different patterns regarding the evolution over time of the journals' scope.

Format

Journals play a role in the legitimization and development of scientific communities (Gingras, 2014; Mullins 1972) and help delineate scholarly communities (Milard 2008) with specific epistemic cultures that shape scholarly discourse (Wakeling et al. 2019). The scope of a journal crystalizes when researchers consider journal fit as a criterion for submitting their work (Tenopir et al., 2019; Solomon & Björk, 2012; Cronin & Younce, 2010; Jamali et al., 2014).

Editors generally also evaluate the fit of submitted manuscripts with the journal before sending it to reviewers who will also assess the relevance of the paper for the journal's readership. These mechanisms consolidate the cohesiveness of the journal for the benefit of its readership and, from a semantic perspective, we can suppose that they increase the average content similarity of papers published in a given journal. Within a discipline, some journals will cover a broad range of topics while others will focus more on a specific area of research. The premise of this work-in-progress is that the semantic similarity of articles published in a journal can be used as a proxy for its scope.

Several factors could influence the evolution over time of the journals' scope. The cognitive extent of science is expanding (Milojevic, 2015), and the lexical concentration of titles of scientific articles is increasing (Bérubé et al., 2018). The transition of journals from print to digital format and the pressure to publish could encourage journals to accept more publications and thus possibly expand its scope. Finally, new models of scholarly publishing such as open access may also be conducive to less lexical cohesiveness, as journal fit may have less weight in editorial decisions. In sum, many factors in the broad scientific environment could have observable effects on the lexical diversity of journals.

Research objectives

Using nine LIS journals as a case study, this work-in-progress proposes a measure of the lexical diversity of a journal as an indicator of journal scope, and More specifically, this work provides preliminary answers to the following research questions:

RQ1. How concentrated or diverse is the content of distinct LIS journals?
RQ2. How is the diversity of scientific discourses in specific journals evolving over time?

Below, we present the data collection steps and the methods used to measure the discursive diversity of journals and the fit of individual papers within these discourse distributions. We then present the answers to the two research questions and lay out an agenda for further research.

Using the Web of Science, we collected all articles, notes and reviews published between 1991 and 2017 in journals in the Library and Information Science category of the NSF classification. We only kept journals that publish in English and that were active over the whole 1991-2017 period for a resulting dataset of 9 journals and 12,549 publications. The title and abstract of were merged and segmented in vectors of 3-grams with TF-IDF-weighted dimensional values. This approach allows for semantically-related words to have non-zero similarity scores and offers comparable results to traditional word-based approaches. We then calculated the average cosine similarity between the text vectors of all articles published in the same journal in the same year.

Results

Table 1 shows the list of journals, the number of publications and the average similarity score between each pair or articles. A visual inspection of the distributions of similarity scores allowed us to confirm that they were approximately normal. In a broader-scope journal such as JASIST, articles will tend to be more diverse in content, as indicated by a lower average similarity score; inversely, a narrower-scope journal like Library Trends will have a higher average lexical similarity score.

Table 1. Average lexical similarity scores of nine LIS journals.

Journal	N	Avg. sim.
Info. Processing & Management	1,586	0.28
Info. Systems Journal	481	0.26
Info. Technology and Libraries	368	0.26
J of Documentation	877	0.32
J of Info. Science	893	0.25
J of the Society for Info. Sci. & Tech.	3,193	0.24
Lib. & Information Science Research	586	0.25
Lib. Trends	990	0.30
Scientometrics	3,575	0.27

Figure 1 shows the evolution of the average lexical similarity score of the nine journals over the 1991-2017 period. No general trend is followed by all journals. The average similarity score decreased for the Information Systems Journal, increased for the Journal of Documentation and Library Trends, and remained relatively stable for the other journals. This heterogeneity of trends between journals suggests the absence of a global trend in the lexical diversity of journals. Instead, the changes could result from editorial decisions, or changes in the research interests of the community of researchers that publish in a given journal.
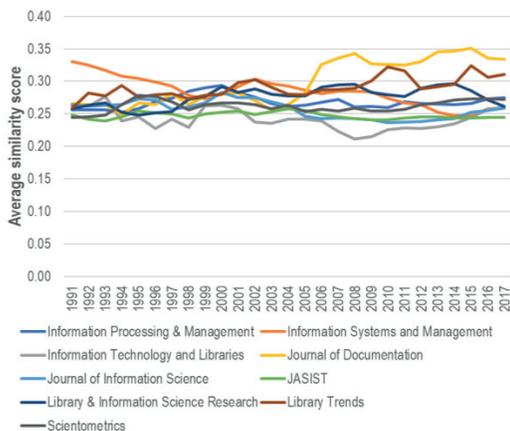

Figure 1. Average lexical similarity (1991-2017)

Discussion and conclusion

The methods and analyses presented in this work in progress lay the foundations for further in-depth analyses of the scope of journal and, more broadly, of the role that journals play in structuring knowledge dissemination. Furthermore, by enabling the identification of core, average, and peripheral papers in terms of similarity with the other papers in the journals, this method might be useful to investigate the publication practices of researchers and the relationship between journal fit and impact.

References

Bérubé, N., Sainte-Marie, M., Mongeon, P, & Larivière, V. (2017). Words by the tail: Assessing lexical diversity in scholarly titles using frequency-rank distribution tail fits. *PLoS ONE 13*(7): e0197775.

Cronin, B., & Younce, L. M. (2010). Publishing in open access education journals: The authors' perspectives. *Behavioral & Social Sciences Librarian, 29*(2), 118-132.

Gingras, Y. (2017). Sociologie des sciences. Paris cedex 14, France: Presses Universitaires de France.

Hagstrom, W. (1965). *The Scientific Community*. New York, basic books.

Jamali, H. R., Nicholas, D., Watkinson, A., Herman, E., Tenopir, C., Levine, K., ... & Nichols, F. (2014). How scholars implement trust in their reading, citing and publishing activities: Geographical differences. *Library & Information Science Research, 36*(3-4), 192-202.

Milard, B. (2008). La soumission d'un manuscrit à une revue: quelle place dans l'activité scientifique des chercheurs? Schedae (Presses Universitaires de Caen), 1, 1-12.

Milojevic, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics*, *9*(4): 962-973.

Mullins, N. C. (1972). The development of a scientific specialty: The phage group and the origins of molecular biology. *Minerva, 10*(1), 51-82.

Solomon, D. J., & Björk, B. C. (2012). Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology, 63*(1), 98-107.

Tenopir, C., Dalton, E., Fish, A., Christian, L., Jones, M., & Smith, M. (2016). What motivates authors of scholarly articles? The importance of journal attributes and potential audience on publication choice. *Publications, 4*(3), 22.

Wakeling, S., Spezi, V., Fry, J., Creaser, C., Pinfield, S., & Willett, P. (2019). Academic communities: The role of journals and open-access mega-journals in scholarly communication. *Journal of Documentation, 75*(1), 120-139.