

Are Special Issues that Special? Distinctiveness and Impact of Special Issues in LIS Journals

Maxime Sainte-Marie¹, Philippe Mongeon² et Vincent Larivière³

¹ *sainte-marie.maxime@uqam.ca*

Centre Interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal
Canada research chair on the transformations of scholarly communication, Université de Montréal

² *philippe.mongeon@gmail.com*

Danish Center for Studies in Research and Resarch Policy, Aarhus University

³ *vincent.lariviere@umontreal.ca*

Canada research chair on the transformations of scholarly communication, Université de Montréal
Observatoire des sciences et des technologies, Université du Québec à Montréal

Abstract

Whether as periodical conference spinoffs or extraordinary journal digressions, Special Issues, are a common yet unexplored area of scholarly communication. In this research, content and citation analysis of Special Issues in Library & Information Science Journals indexed in the Web of Science shows that special issues are distinct from regular issues “contentwise”, yet indiscernible “citationwise”, thus leaving their existence and persistence unexplained from a publisher’s economic perspective.

Introduction

As scholarly communication is in most fields based on journal publications, Special Issues (SIs) play a significant and lasting role in both knowledge production and dissemination. A SI can be defined as a journal issue “either completely or partly devoted” to a single topic” (Olk & Griffith, 2004, p. 120), the latter either referring to an area of study, a theoretical approach or a methodology (Priem, 2006). Despite their ubiquity, SIs do not make unanimity within the scholarly community, especially as regards to whether or not their publication is detrimental to research impact (Conlon et al., 2006; Hendry & Peichel, 2016; McKinley, 2007; Mowday, 2006; Olk & Griffith, 2004; Schoonhoven, 2004; Siguffi, 2011). At first glance, both perspectives seem plausible. On the one hand, as SIs grant “increased legitimacy and attention” (Conlon et al., 2006, p 859) to relevant or unusual topics of interest, which helps extend the journal readership and potentially boost its citation rates. Inversely, in order “to either meet deadlines or to just fill budgeted pages”, journal editors may be forced to accept substandard papers, thus reducing the total number of citations received and “damaging the image of the journal” (Siguffi, 2011, p. 306).

Some empirical studies were conducted on the matter. Based on an analysis of journal issues published between 1988 and 1999 in 5 management journals, Olk and Griffith (2004) show that SI articles have a higher citation rate than regular issue (RI) articles. Expanding on this study, Conlon et al. (2006) show however that this citation boost is only apparent in lower-impact journals. Outside the field of management, Hendry and Peichel (2016) collected citation data of articles published in conference-based SIs published by the International conference on Stickleback Behaviour and Evolution. Their analysis shows that papers published in SIs have comparable citation impact and longevity to articles published in the same journal and year, as well as a lower but longer citation impact than topic-related papers published in RIs the same year. More recently, Sala, Lluch, Gil, and Ortega (2017) analyzed 1120 articles published in 10 Ibero-American psychology journals included in the Journal Citation Index and published between 2013 and 2015. By comparing RI articles to ones published in “open call” or invitation-

based SIs, the authors observe that SI papers receive a higher number of citations than RI articles, and that this higher citation impact is not the consequence of author or journal self-citations (Sala et al., 2017).

While these studies mostly agree on the research impact of SIs in their respective disciplines, they unfortunately suffer from two common shortcomings. First, their results are based on a rather small sample of issues. But most importantly, they take for granted what might be the most obvious and characteristic feature of SIs: their topicality. Regardless of discipline and whether based on open calls, conference presentations or invitations to publish, all SIs focus by definition on a more or less specific theme. And as with research impact, this topicality of SIs is not only far from trivial, but also and still in need of a proper bibliometric assessment. In light of these considerations, the aim of the present paper is to attempt a large-scale investigation of the topicality and impact of SIs. In the first step, vector semantic models are generated in order to assess the topicality of both SIs and RIs and thus verify whether or not SIs stand out in this respect. Following this, a citation analysis similar to but broader than the above-mentioned one is undertaken in order to determine whether publishing special issues contributes to a scholarly journal's influence and outreach.

Methodology

We retrieved from the Web of Science (WoS) all articles, notes and reviews published in the last 10 years (between 2009 and 2018 inclusively) in Library & Information Science journals, as classified by the National Science Foundation. We chose 2009 as starting year for our study because SIs in LIS journals are very scarce in WoS before that date. We further limited our dataset to journals that have published at least one special issue and that published at least eight issues with at least four articles each over the whole period studied. All articles that are not written in English were also removed. These dataset restrictions were applied in order to allow for reliable and robust similarity computations and comparisons. The resulting dataset contains a total of 14 132 documents published in 1 335 issues distributed amongst 34 distinct journals; of the lot, 122 (9.14%) issues and 1 213 (6.96%) articles are of the special kind. For each relevant article entry, the following attributes were extracted: article ID, title, journal, publication year, issue ID, special issue status as well as total citations, normalized by year and discipline.

Issue-level data was then obtained through the following processing tasks. Field-and-year-normalized relative citations were first aggregated at the issue level, then divided by the number of articles contained in each issue. Also, in order to obtain content similarity scores for both RI and SI articles, article title and abstract data for all collected LIS articles were merged into one single text string attribute, then segmented in vectors of 3-grams (substrings of 3 characters) with TF-IDF-weighted dimensional values. The main reason for using word substrings instead of whole words is that it allows semantically-related words such as 'science', 'scientific', 'scientifically' and 'scientist' to have non-zero similarity scores. This character sequence segmentation procedure has also been shown to offer comparable results to traditional word-based approaches over various Natural Language Processing-based tasks (Cavnar & Trenkle, 1994; Damashek, 1995; McNamee & Mayfield, 2004). Following these data processing steps, the topicality or thematic cohesiveness of each issue was obtained by calculating the average cosine similarity between the text vectors of all articles included in the issue.

Analysis

Table 1 shows the distribution by year of all LIS RIs and SIs. In the case of RIs, if we exclude journal issues published in 2018 (year for which data has yet to be collected), the number of different LIS journals having published special issues over the year remains relatively stable, while the number of issues and articles slightly increases over the period. More surprising however is the case of SIs: while the number of journals collecting special issues has remained relatively constant over time, there has been an impressive surge in number and proportion of SIs. Indeed, while counting for around 3% of all publications in 2009, the share of SIs and special articles increased to around 14% and 13% of their respective publication types in less than 10 years.

Table 1. Descriptive Stats for LIS Journals, 1991-2018.

Year	Journals	Special Issues		Articles		Regular	
		#	%Total	#	%Total	Issues	Articles
2009	31	4	3,2	25	2,4	120	1014
2010	31	8	5,8	65	5,5	130	1113
2011	32	12	8,3	112	8,9	133	1153
2012	31	13	8,9	99	6,4	133	1443
2013	32	15	9,9	93	7,1	137	1214
2014	33	15	8,6	132	7,7	159	1589
2015	32	15	9,0	110	6,6	152	1555
2016	31	14	7,9	112	5,8	163	1831
2017	30	17	9,9	154	8,3	154	1699
2018	25	9	14,3	81	13,1	54	538

A number of reasons can be put forward to explain this recent surge in the publication of SIs. The simplest one is that SIs were not indexed as such in the WoS until then. If that were the case, however, the relative importance of SIs would probably have increased suddenly rather than over several years. Another possible explanation may be that journal editors tend more and more to find and appoint special editors to specific issues in order to reduce their own workload. One could also suggest that more and more organization committees try to attract submissions to conferences by announcing that some or all papers chosen by the program committee will be published in a SI of a given journal. Beyond these speculations, one thing is certain though: this surge in publication and relative importance of SIs certainly warrants further investigation, be it exploratory and cross-disciplinary or explanatory. More on that matter will be said in the Discussion section.

As regards to content similarity, the percentile rank distribution of content similarity scores for both RIs and SIs is presented in Figure 1. Both distributions are positively skewed, with a little more than 18% of the highest-scoring SIs and RIs accounting for slightly more than half of the total cumulative score of their respective issue types. Additionally, while both distributions have short heads and tails, similarity scores for SIs tend to be generally higher than those of RIs. Comparison of mean content similarity scores for both types of issues further confirms this intuition, as articles within a SI are on average 17% more similar in content (0.33 vs. 0.28) than articles from any given RI. Comparison of median values leads to the same conclusion, as the median for the SI distribution (0.32) is 18% higher than that of its counterpart (0.27). Given the similar shapes of both distributions and in order to confirm their distinctiveness as regards to content similarity, we conducted a Mann-Whitney U test (Mann and Whitney, 1947) on the

similarity distributions of SIs and RIs and obtained a U-statistic of 37,970 and a p-value of 3.51×10^{-19} . Since the null hypothesis of that test states that the random variables corresponding to the two independent groups being compared are stochastically equal (“each datum of the first group will have an equal chance of being larger or smaller than each datum of the second group” (Nachar, 2008)), rejecting this hypothesis due to the very low p-value obtained entails that elements of both groups belong to distinct populations (Nachar, 2008). In the present one-tailed context, this discrepancy in central tendencies means that the elements as well as the median of one group, namely the SIs similarity score distribution, are significantly higher than those of the other group. Point biserial correlation (Lev, 1949) further confirms the higher content similarity of SIs over RIs, with r and p values of 0.26 and 5.74×10^{-22} respectively. In sum, SIs in Library and Information Sciences, whether conference-related or not, are not only a recent emerging phenomenon, but also one that also has to be distinguished from RIs “contentwise”, as the articles they contain tend to form a semantically more consistent whole.

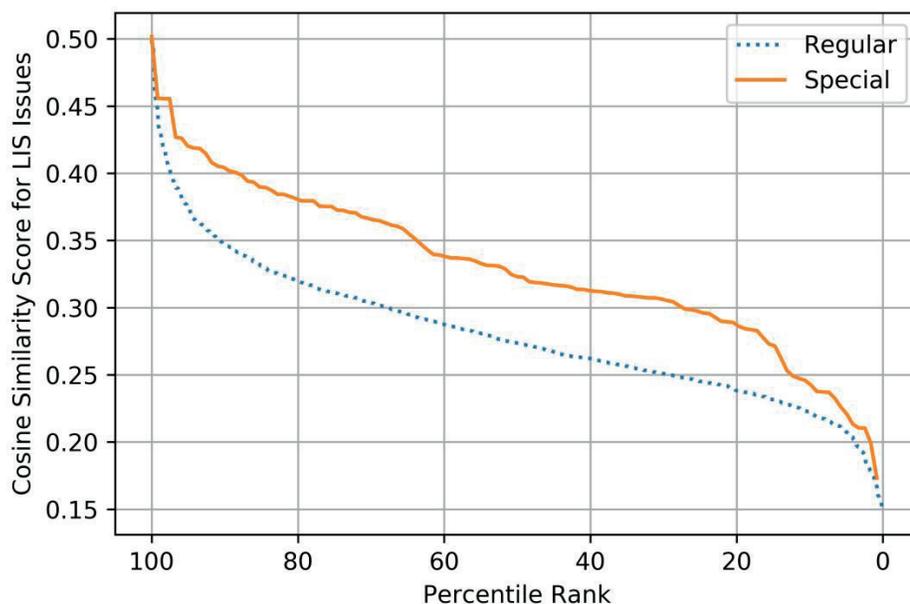


Figure 1. Frequency-Rank Distribution of Cosine Similarity Scores for LIS Issues

As regards to research impact, the frequency-rank semi-logarithmic distributions for both special and regular LIS issues are shown in Figure 3. At first glance, both distributions have strikingly similar shapes, slopes, heads, and tails. As can be expected from previous bibliometric literature, the two distributions are highly skewed, with 2.85% and 4.3% of regular and special issues accounting for more than half of the relative citations of their respective issue types. Both distributions have strikingly similar shapes, slopes and heads. As regards to the central tendencies observed for both distributions, the mean relative citation score by article value for SIs (1.18) is only 4% higher than that of the RIs distribution (1.13), whereas the median value for RIs (0.88) is 5% higher than the value for SIs (0.83). Resorting once more to point biserial correlation, we obtained a negligible r value of 0.01 for the relationship between special issue status and relative citation score per article; however, a high p -value of 0.67 prevents us from drawing any statistically significant conclusion on that matter. The Mann-

Whitney test however provides more adamant results, with U- and p- values for both RI and SI citation score distributions of 70422 and 0.19 respectively. Based on the high p-value, the null hypothesis cannot be rejected, which means that the random variables corresponding to each distribution are stochastically equal, and thus that the elements and the medians of both groups are statistically undistinguishable. These results thus suggest that the citation cost of publishing articles in SIs rather than in RIs tends to be negligible.

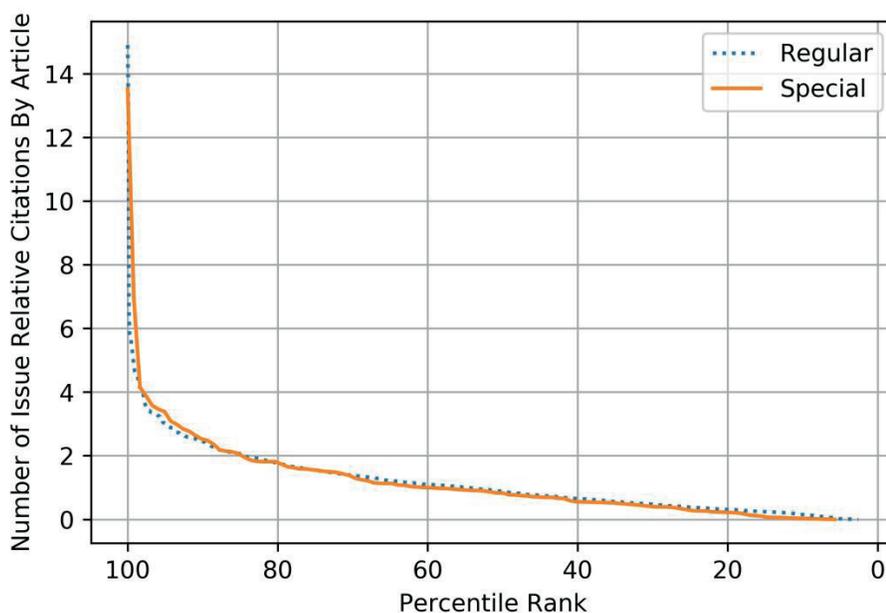


Figure 2. Frequency-Rank Distribution of Relative Citations by Article for LIS Journal Issues

Discussion

As the previous section has shown, RIs and SIs in LIS journals are distinct “contentwise”, yet indiscernible “citationwise”. Taken together, conference-related and unrelated SIs thus distinguish themselves from RIs from a thematic standpoint, but this specificity does not impede nor boost their research impact. Even though the explicit topicality of SIs may attract new readers or discourage others, these readership dynamics result in a zero-sum citation game. And while it is not unreasonable to suppose that some special editors may have, at times and for thematic cohesiveness purposes, accepted lesser-quality articles in order to complete their issue, nothing in the data suggests that this practice may have had an effect on other articles from the same issue or special issues in general.

Given this impact neutrality of SIs, why publish them at all then? And why are they increasing in number? In our opinion, there are necessarily incentives to publish SIs, but these benefits have to lie beyond the quantitative realm of bibliometrics. It was earlier hypothesized that journal editors may appoint special editors in order to reduce their workload. SIs may even be proposed in order to attract submissions to conferences, as mentioned before. Editing special issues or publishing in them may help gain recognition and strengthen bonds within a more local community of peers and collaborators. Editing a special issue can also be more enjoyable than editing a regular one: for the editor, “the collection of papers in a special issue can be more interesting; the review process is more collegial, constructive, and efficient; editorial decisions

are more enjoyable; and the opportunity to advance the field is greater” (Hendry & Peichel, 2016: 141). Finally, SIs are arguably more enjoyable to read than regular ones, as all articles contained within a given SI are potentially of higher relevance to the interested reader. However, given that most articles can now be directly accessed electronically, independently of any consideration at the issue level, one may arguably wonder whether speaking of issue readability or attractiveness is still relevant at all. And while all these conjectures are plausible, no scenario that would directly benefit journals and compensate for the opportunity cost mentioned in the introduction can be reasonably thought of, thus leaving the sudden and surprising increase of SIs in LIS journals without a proper and rational economic explanation.

Of course, various unaccounted factors might have affected the results presented here. For once, the quality of the database used in this research might also be questioned on various grounds. First, the accuracy of the classification of LIS issues as RIs or SIs in the Web of Science cannot be exhaustively assessed. As regards to the text-based methods used here, similarity computations between articles are entirely dependent on the wording behavior of authors, which is inherently subjective and as such often elusive to scientific inquiry. However, these limitations do not invalidate the results presented: SIs in LIS represent a scholarly communication form that is both distinct “contentwise” and indiscernible “citationwise”; whether this situation also applies to other disciplines is a matter for future research.

References

- Cavnar, W.B. & Trenkle, J.M. (1994). N-gram-based text categorization. In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval (Las Vegas, NV, 1994), 161-175.
- Conlon, D. E., Morgeson, F. P., McNamara, G., Wiseman, R. M., & Skilton, P. F. (2006). From the editors: Examining the impact and role of special issue and regular journal articles in the field of management. *Academy of Management Journal*, 49(5), 857-872.
- Damashuk, M. (1995). Gauging Similarity with *n*-Grams: Language-Independent Categorization of Text. *Science*, 267: 843-846.
- Hendry, A.P. & Peichel, C.L. (2016). Are conference special issues worthwhile? *Evolutionary Ecology Research* 17: 141-155.
- Lev, J. (1949). The point biserial coefficient of correlation, *Annals of Mathematical Statistics* 20(1): 125-126.
- Mann, H.B. & Whitney, D.R. (1947). On a test of whether one of 2 random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18: 50-60.
- McKinley, W. (2007). Special issues as vertical integration: A rejoinder to priem and mowday. *Journal of Management Inquiry*, 16(3), 240-245
- McNamee, P. & Mayfield, J. (2004). Character *N*-Gram Tokenization for European Language Text Retrieval. *Information Retrieval* 7, 73-97.
- Mowday, R. T. (2006). If special issues of journals are not so special, why has their use proliferated? *Journal of Management Inquiry*, 15(4), 389-393
- Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology*, 4, 1: 13-20.
- Olk, P. & Griffith, T. L. (2004). Creating and disseminating knowledge among organizational scholars: The role of special issues. *Organization Science*, 15(1), 120-129.
- Priem, R. L. (2006). What happens when special issues just aren't “special” anymore? *Journal of Management Inquiry*, 16(3), 246-249.
- Schoonhoven, C. B. (2004). Putting rumors to rest: the impact, quality, and speed of journal special issues. *Organization Science*, 15(1), 119-119
- Schütze, H. (1993). Word space. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems* (Vol. 5, pp. 895-902). San Mateo, CA: Morgan Kaufmann.
- Siguffi, S. P. (2011). On special issues. *Journal of African Business*, 12(3), 305-307.