



RESEARCH ARTICLE

# On the topicality and research impact of special issues

Maxime Sainte-Marie<sup>1,2,3,5</sup> , Philippe Mongeon<sup>3</sup> , and Vincent Larivière<sup>1,4,5</sup> 

<sup>1</sup>Canada Research Chair on the Transformations of Scholarly Communication, École de bibliothéconomie et des sciences de l'information, Université de Montréal, Canada

<sup>2</sup>Bureau des initiatives numériques, Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Canada

<sup>3</sup>Danish Centre for Studies in Research and Research Policy, Department of Political Science and Government, Aarhus University, Denmark

<sup>4</sup>Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Canada

<sup>5</sup>École de bibliothéconomie et des sciences de l'information, Université de Montréal

an open access  journal



Citation: Sainte-Marie, M., Mongeon, P., & Larivière, V. (2020). On the topicality and research impact of special issues. *Quantitative Science Studies*, 1(1), 303–319. [https://doi.org/10.1162/qss\\_a\\_00009](https://doi.org/10.1162/qss_a_00009)

DOI:  
[https://doi.org/10.1162/qss\\_a\\_00009](https://doi.org/10.1162/qss_a_00009)

Received: 24 April 2019  
Accepted: 26 August 2019

Corresponding Author:  
Maxime Sainte-Marie  
[maxime.sainte-marie@umontreal.ca](mailto:maxime.sainte-marie@umontreal.ca)

Handling Editor:  
Ludo Waltman

**Keywords:** scholarly publishing, journals, issues, special issues, citation impact, vector semantics

## ABSTRACT

The publication of special issues constitute an important yet underinvestigated phenomenon of scholarly communication. In an attempt to draw attention to the proliferation of special issues, Priem (2006) suggested that their commissioning has an underestimated opportunity cost, given the relative scarcity of publication space: by distorting the “marketplace for ideas” through the commanding of preselected topical distributions, special issues undermine the total research output by “squeezing out” high-quality but topically unrelated articles. The present paper attempts to test this hypothesis by providing a topicality and research impact analysis of conference-based, monographic, and regular issues published between 2010 and 2015 inclusive and indexed in Clarivate Analytics' Web of Science. The results show that the titles and abstracts of articles copublished are topically closer to each other than those copublished in regular issues, which suggests that their relative importance might influence the total topical distribution. However, disciplinary and overall comparison of relative citations for both special and regular issues shows that intrajournal averages and variances in the former case are respectively higher and lower than in the regular issue context, which undermines not only the abovementioned hypothesis, but also the belief that editors often “fill up” special issues by accepting substandard papers.

## 1. INTRODUCTION

As scholarly communication is increasingly shaped and driven by journal publications (Cope & Phillips, 2014; Wakeling et al., 2019), special issues (SIs) play a significant and lasting role in both knowledge production and dissemination. An SI can be defined as a journal issue “either completely or partly devoted to a single topic” (Olk & Griffith, 2004, p. 120), the latter referring to an area of study, a theoretical approach, or a methodology (Priem, 2006). Pervading the scholarly communication system as a whole, SIs can originate from various procedures.

Some special issues are generated from open calls for papers on a specific topic. Submissions are then subject to peer review, although the editors and reviewers may be

Copyright: © 2019 Maxime Sainte-Marie, Philippe Mongeon, and Vincent Larivière. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



unique to those special issues. Others grow out of miniconferences or sessions in larger conferences and typically contain papers by conference participants as their centerpieces (...). Finally, some special issues are comprised of articles from authors who were invited to write for them (Conlon, Morgeson, McNamara, Wiseman, & Skilton, 2006, p. 859)

Regardless of their type and despite their numbers, variety, and persistence, SIs are not unanimously welcomed within the scholarly community. Perhaps the most fierce criticism of SIs, stemming from the field of management, relates to the negative impact of their topicality. According to this view, the frequency and persistence of SI commissioning in the context of scarce publication space “distorts the marketplace for ideas by commanding particular frequency distributions of preferred topics in journals” (Priem, 2006, p. 383), which in turn squeezes out high-quality but topically unrelated articles that would otherwise have appeared in regular issues (RIs).

Even when special issues are peer reviewed and their topics are submitted from academy members, as most topic special issues are, the central commissioning decision (...) likely will only retard knowledge generation by placing boundaries on creativity. Importantly, that retardation will pertain even when those who do the choosing of special issues are thoughtful, well meaning, and eminently qualified (...). The great irony is that profligate commissioning of the very special issues that are intended to spur knowledge generation actually thwarts it (Priem, 2006, p. 387).

This hypothesis has led to various theoretical contributions surrounding the positive or negative impact of SIs in management journals (McKinley, 2007; Mowday, 2006, 2007; Priem, 2007). At first glance, both perspectives seem plausible: On the one hand, as “virtual organized symposia” (Eden, 2010, p. 904), 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 forcing journal editors to “squeeze in” thematically related but potentially substandard papers at the expense of regular ones, SIs might bear a significant opportunity cost, by reducing the total number of citations received and thus “damaging the image of the journal” (Sigué, 2011, p. 306).

Still in the field of management, interviews with journal editors have shown the absence of consensus on these matters (Olk & Griffith, 2004; Rynes, 2003). On the one hand, many of the editors interviewed pointed to the assumption that by “spurring research on new, innovative topics” (Priem, 2006, p. 384), SIs increase submissions to a journal as well as the latter’s prestige. Other mentioned benefits include showcasing work from smaller academic divisions and providing a “training ground for prospective editors” (Rynes, 2003, p. 536). By contrast, many editors were skeptical about the importance and impact of SIs:

They observed that special issues absorb scarce resource like journal page budgets, decreasing the number of pages available to publish regularly submitted papers. Some editors suggested that editor book volumes might be a more appropriate publication outlet for such special topics. They also raised a concern about the quality of special issue papers. Although in general most special issue articles were perceived as meeting journal standards through the blind review process, some editors remarked that rumours existed within the Academy to the effect that peer-reviewed journal standards are sometimes waived to complete a special issue. As the rumors went, some special issues were thought to have included substandard articles to fill the issue (Olk & Griffith, 2004).

Some empirical investigations were conducted in order to shed light on these matters. In the first important contribution to the study of SIs, Olk and Griffith (2004) analyzed journal issues

published between 1988 and 1999 in five mainstream management journals. Their analysis shows that SI articles have a higher rate of citations than RI articles. The authors also report no difference in variance in SI or RI citation counts for the studied sample, which allegedly invalidates the argument according to which SI editors may accept substandard articles in order to “fill up” the issue. Based on these results, the authors conclude that SIs have a clearly favorable impact on knowledge development: By improving citation impact “while simultaneously maintaining the journal’s normal standards for regular issue articles” (Olk & Griffith, 2004), SIs act as “vanguards” of knowledge development Fleck (2012), forging and widening new, explorative research paths in order to allow normal, mainstream science to exploit and develop them at a later stage.

Expanding on this study, Conlon et al. (2006) collected Web of Science (WoS) data for articles published between 1984 and 2001 inclusive in nine top management journals, including all five journals from the previous study. Of paramount importance here is the assumption that the citation impact of SIs may depend on the process leading to their publication:

Conference-based special issues are likely to have the highest impact. Such a special issue likely focuses on an emerging, topical subject and has a built-in audience familiar with the topic because of the conference connection; the prior visibility of the conference might lend these articles more impact. An open-call special issue may not have the same impact because it does not have the benefit of visibility lent by a conference focused on the special issue topic. An invited issue may have a lower impact because it benefits neither from association with a conference nor from vetting through a full peer review (Conlon et al., 2006).

The results from that study show that the citation boost attributed by the previous study to SIs only applies to open call and conference-based SIs. In addition, no positive impact was observed for more prominent journals. Based on these results, the authors argue that publishing open call SIs, in particular conference-based ones, may represent a useful strategy for lower impact journals to increase both article impact and readership.

Outside the field of management, few studies have been devoted to these empirical questions. In the field of biology, Hendry and Peichel (2016) analyzed citation data of articles published in the first seven SIs of the International Conference on Stickleback Behaviour and Evolution. By comparing the citation impact of these articles to that of other papers published in the same journals in the same years, the authors conclude that “journals do not suffer from publishing special issues based on conferences” (Hendry, 2016, p. 144), as papers published in SIs have comparable citation impact and citation longevity to articles published in the same journals in the same years. In a second series of analyses, the authors make use of the alleged topicality or content similarity of articles within SIs: By comparing mean annual citation rates for stickleback papers inside SIs to stickleback papers published in RIs the same year, the authors find that papers in SIs have a lower but longer citation impact than the latter. In light of these results, the authors conclude that the publishing of SIs is worthwhile and that scholars should not be afraid to publish in them, as their papers might fare better in such contexts than in RIs. More recently, Sala, Lluch, Gil, and Ortega (2017) analyzed 1,120 articles published in 10 Ibero-American psychology journals between 2013 and 2015. By comparing RI articles to “monographic” ones, that is, articles published in open call or invited SIs, the authors conclude that monographic SI papers receive a higher number of citations than nonmonographic ones, and that this higher citation impact is not the consequence of author or journal self-citations (Sala et al., 2017).

Overall, these studies agree on the fact that publishing SIs has an at worst negligible citation impact for journals, regardless of SI type. In this sense, one would be tempted to rule out the idea that SIs have a negative research impact. However, two important gaps in the literature preclude such a possibility. First of all, the data sets analyzed are all limited to one discipline, which prevents the discovery of disciplinary idiosyncrasies or cross-disciplinary patterns and thus seriously impedes the scope and generalizability of research findings. But most importantly, these various studies take for granted what might be the most obvious and characteristic feature of SIs: topicality. Whether based on open calls, conference presentations, or invitations to publish, all SIs focus by definition on a more or less specific theme. And according to the original argument against SIs, this topicality of SIs is precisely what disturbs the spontaneous and unconstrained flow of ideas that ensures the optimality of knowledge generation. However, this topical drift and its impact on research impact is not only far from trivial, but also in need of a proper and thorough bibliometric assessment.

In light of these considerations, the aim of this paper is to provide a large-scale comparative investigation of the relationship between topicality and impact investigation in RIs and SIs. Vector semantic models are first generated in order to assess the topicality of each issue type through disciplinary comparison of inraissue similarity scores for both titles and abstracts. It is here assumed that discovering an important difference in topical cohesion between RIs and SIs would be tantamount to conferring to the latter a voice of its own within the scholarly journal ecosystem, but also a decisive and potentially “disruptive” influence on the global topical landscape. Following this, a relative citation analysis is undertaken in order to determine the research impact of SIs. First, disciplinary relative citation by issue article averages for each issue type are compiled and compared in order to evaluate the impact differential related to the publication of SIs. An analysis of inraissue variance for each issue type is then conducted: Assuming that lower quality articles tend to get fewer citations, average relative citation variance at the issue level can be used as a proxy for article quality consistency and thus allow for the verification of the substandard article “squeeze in” hypothesis related to the publication of SIs.

## **2. METHODS**

Document and issue information for all articles, notes, and reviews considered in this study was extracted from WoS. For each distinct article extracted, the following attributes were considered: title, abstract, publication year, and journal and issue information. Of paramount importance here is the indexing of SI-related information. The latter is conveyed in the WoS database by the “SI” tag, whose value is of either string or numerical type, depending whether the corresponding SI is numbered or not. As this SI tag applies to whole issues rather than individual articles, we were unable to identify “mixed issues,” that is, issues including both regular and special articles.

The information regarding the special status of issues is directly supplied by the publishers themselves as part of the issue metadata. Unfortunately, issue metadata standards vary a lot depending on the publishers or journals (Marie McVeigh, Clarivate Analytics, personal communication). Despite the fact that issue numbering cataloging became an established practice long before the creation of citation indices, such variability does cast doubt on the reliability of SI information. In addition, the proportion of SIs has greatly increased in the wake of the regional expansion program undertaken in 2008 by Thomson Reuters (Cross & Jansz, 2009). Given these considerations, only articles published from 2010 to 2015 were considered for this study. The year 2010 was chosen as the starting point because the number and relative frequency of SIs as well as the frequency of WoS-indexed journals stabilizes from that year

onwards for all disciplines. Given that capture policies remain consistent within the different WoS citation indices (Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index, Emerging Sources Citation Index; Marie McVeigh, Clarivate Analytics, personal communication), such stabilization, in our opinion, reduces the probability that variations in cataloging and journal indexing practices might affect data reliability. A manual inspection of more than a hundred issues from the period under study was also conducted, and no SIs classified as RIs or vice versa were found. Finally, all issues published after 2015 were excluded from the study in order to allow for longer citation windows and thus optimize accuracy of research impact analyses (Wang, 2013).

Of all the distinct journals included in the WoS database, only those having published at least one SI during the observed period were considered. For reliability and sample representativity purposes, we chose to focus on established journals and practices; to achieve this, we kept only journals that published at least one issue in each year of the covered period, and all issues that had fewer than four articles written in English and with abstracts were removed. The resulting data set comprises 2,914,223 articles published in 202,767 issues (of which 23,055 are SIs) and 4,559 journals. Disciplinary distribution of issues and articles is shown in Table 1.

In order to identify conference-based SIs, amongst the set of extracted SIs, one article from each of the 23,055 SIs was retrieved from the online version of the WoS. All articles that had online conference-related information as well as the corresponding SIs were identified as conference-based, amounting to 2,498 issues. We manually verified a random selection of 100 monographic SIs and found that 10% were actually related to some type of scientific event

**Table 1.** Issue and article distribution by discipline

Discipline	Disciplinary group	#Issues	#Articles
Arts	AH	590	3,741
Biology	LS	15,216	208,520
Biomed.	LS	19,933	393,382
Chemistry	PSE	13,004	270,342
Clin.Med.	LS	34,103	503,917
Earth & Space	PSE	12,925	189,119
Eng.&Tech.	PSE	33,564	511,470
Health	SocSci	5,682	55,801
Humanities	AH	4,894	32,098
Mathematics	PSE	8,610	98,585
Physics	PSE	12,114	269,062
Prof.Fields	SocSci	15,777	134,067
Psychology	SocSci	8,701	90,591
Social Sciences	SocSci	17,654	153,528

*Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. = Engineering & Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences. *Disciplinary Groups:* AH = Arts&Humanities, LS = Life Sciences, PSE = Physical Sciences & Engineering, SocSci = Social Sciences.



(e.g., workshop or symposium). Because we did not proceed to complete manual coding of every SI, one limit of our analysis is thus that some conference-based SIs may not be categorized properly. Another important limitation to our study relates to the inability to identify SIs that were the result of personal invitations. As a result, no distinction is here made between open call and invitation-based SIs. Following Sala et al. (2017), the present analysis will focus on two different types of SIs: monographic (MonoSIs) ones, which are based on either open calls or invitations, and conference (ConfSIs) ones. Following these various manipulation and categorization procedures, issue data for the present research is as follows: 179,712 (88.63%) RIs, 20,557 (10.13%) MonoSIs, and 2,498 (1.23%) ConfSIs, for respective totals of 2,614,053 (89.7%), 253,205 (8.69%) and 46,965 (1.61%) articles.

Discipline assignment of RIs and SIs was done using the National Sciences Foundation (NSF) field classification of journals used in the Science and Engineering Indicators (SEI) reports (National Science Foundation, 2006). NSF classification assigns only one of the 14 different disciplines to each journal; each document is thus assigned only one discipline, which allows complete disciplinary partitioning while avoiding any double counting of papers.

The research impact of MonoSIs, ConfSIs, and RIs was calculated following a three-step process. Field-and-year-normalized citation scores for all articles were collected and divided by the average annual number of citations received for all articles published in the same year and belonging to the same NSF discipline. Also, as SIs tend to contain fewer articles than RIs (SIs account for 11.37% of all issues but only 10.30% of all articles in the data set), relative citations by issue article (RCIA) ratios for all issue articles were obtained by averaging the score of all articles included in each issue. Finally, intraissue variance in relative citations was compiled for all issues.

Topicality analysis of RIs and SIs was done using word space models, a proven and lasting text modeling technique widely used in computational semantics (Baroni & Lenci, 2010; Gärdenfors, 2004, 2014; Sahlgren, 2006; Schütze, 1993; Turney & Pantel, 2010; Widdows, 2004). All article titles and abstract were first segmented in vectors of 3-grams (substrings of three characters) with TF-IDF-weighted dimensional values based on the 3-gram title or abstract lexicon of the corresponding discipline. 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 nonzero 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, et al., 1994; Damashek, 1995; McNamee & Mayfield, 2004). Then, for each issue and issue type, average cosine similarity scores were compiled for all article title and abstract combinations successively.

Following these various procedures, title and abstract topicality score distributions as well as RCIA ratio distributions were generated for each issue type and disciplinary context (all disciplines as well as the whole scholarly corpus), amounting to  $3 \times 15 = 45$  distributions. In the literature, pairwise mean comparisons of independent samples are usually done using Welch’s unequal variances t-test (Welch, 1947). Because this test requires that the distributions being compared follow a normal distribution,  $K^2$  normality tests (D’Agostino, Belanger, & D’Agostino, 1990) were conducted on all distributions. However, due to the strong positive skewness and high kurtosis of all distributions, the vast majority of distributions had inconsequential normality scores, which prevents the use of the abovementioned statistical tests. In such cases, the nonparametric Mann–Whitney U-test (Mann & Whitney, 1947) is often used in order to compare distribution medians based on rank sums comparisons (Nachar, 2008).

However, this can only be the case where the only difference between the compared distributions is a shift in location; when there is a difference in shape or spread, the Mann–Whitney test can indicate that two distributions are different even though their medians are similar. In this sense, using this test to compare medians “can lead to inadequate analysis of data” (Hart, 2001, p. 391).

Given these considerations, empirical likelihood-based confidence intervals were calculated for the distribution means of all relevant variables using the *statsmodels* package. Introduced by Owen (1988, 1990, 2001), empirical likelihood is a method of nonparametric inference and estimation which only requires data to be independent and identically distributed and performs well with asymmetric distributions of the sort that bibliometrics often has to deal with. Most importantly, no prior assumption regarding underlying distributions, scale, or skewness are required, as empirical likelihood automatically and uniquely determines confidence regions whose shape mirrors the shape of the data (Hall & LePage, 1996).

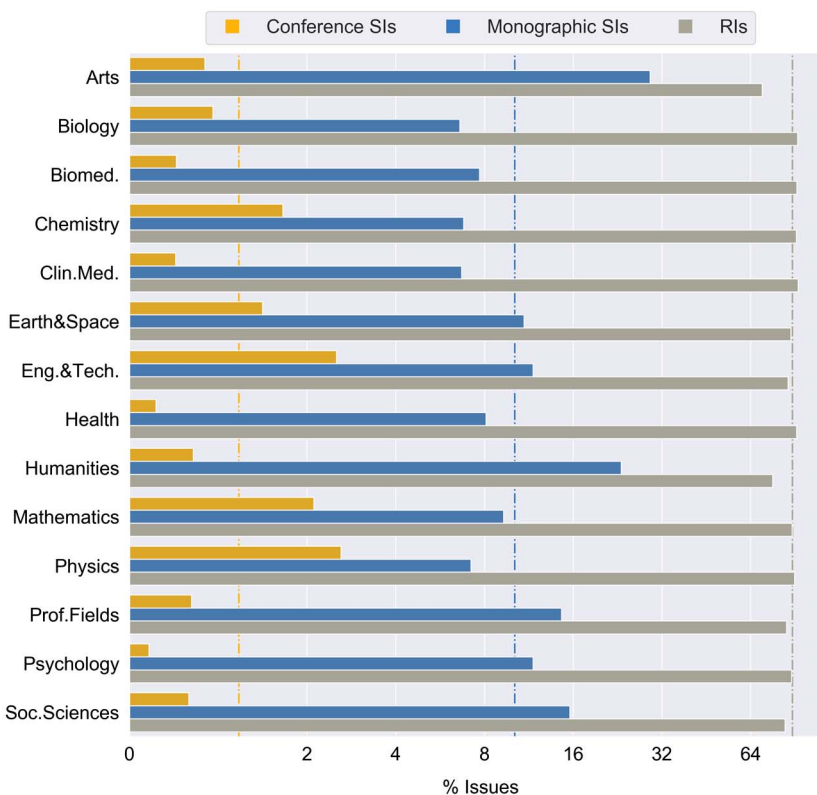
The main advantage of (...) empirical likelihood based method[s] is that they allow the data to determine the degree of asymmetry of the confidence interval. The endpoints of the confidence interval for the mean are the weighted averages of the sample observations. These weights are positive, therefore, the extreme observations influence the width of the confidence interval for the mean (Tursunalieva & Silvapulle, 2009, p. 15).

Empirical likelihood has been shown to be imprecise when the sample size is small or when dealing with distributions with infinite variance, as in some heavy-tailed contexts (Di Ciccio, Hall, & Romano, 1991; Hall & La Scala, 1990; Tsao, 2004). Sample size limitations are inoperant here, however, as the present data set exceeds in size those considered in the abovementioned studies by several orders of magnitude. As for variance constraints, Cheng, Liu, and Liu (2016) have recently demonstrated that empirical likelihood is still applicable under the infinite second moment condition, as infinite population variance slows down convergence but does not prevent it. In light of these considerations, empirical likelihood-based confidence intervals of distribution means represent a more than adequate means to assess the distinctiveness of SIs as regards to topicality and research impact.

### 3. RESULTS

Figure 1 shows the total proportion in percentage of MonoSIs and ConfSIs published during the 2010–2018 period for each NSF discipline. For visualization purposes, the x-axis has been scaled logarithmically; the total proportions of each issue type on the whole data set are also indicated by dotted horizontal lines. Although the share of MonoSIs obviously exceeds that of ConfSIs, both issue types represent the lesser part of journal issues. Indeed, the disciplinary share of MonoSIs never exceeds 30%, whereas the maximum share for ConfSIs is 10 times lower than that value. In both cases, however, the interdisciplinary variability in the proportion of SIs is noteworthy: Although one out of a little more than three arts issues is a MonoSI, that proportion drops to more than one out of 15 issues in the case of biology; in parallel, the publication frequency of ConfSIs for physics is one for 38.31 issues, whereas that number drops to one ConfSI for each 454.54 issues published in psychology.

Beyond this disciplinary variability, some interesting patterns also emerge from the data. First, the proportion of MonoSIs in arts and humanities journals (23.91%) is markedly higher than in other disciplines (9.76%). As for social sciences disciplines (represented here by health, professional fields, psychology, and social sciences), their proportion of MonoSIs (13.67%) is globally higher than that of both physical sciences & engineering (9.85%) and life



**Figure 1.** Issue and article distribution by discipline. *Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. Engineering&Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

sciences (6.94%) disciplines. With the sole exceptions of physics, chemistry, and health, there is no disciplinary overlap between each of the four disciplinary groups discussed here, which suggests the following pattern in the relative importance of MonoSIs in scholarly journals:

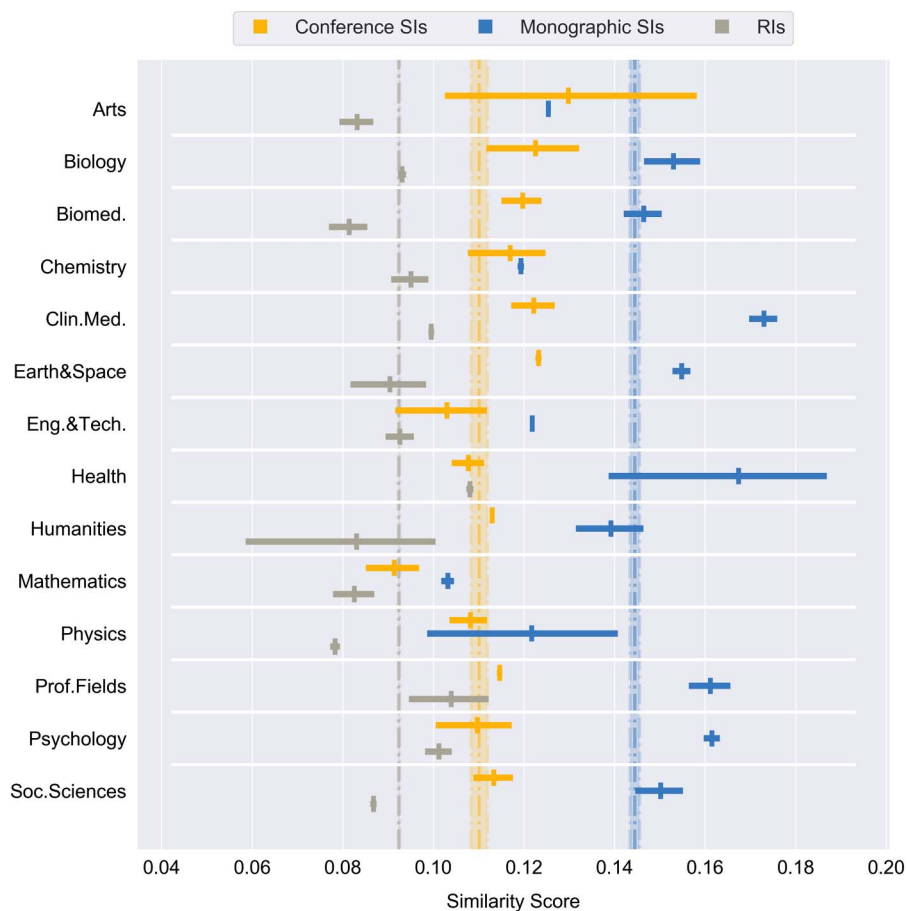
$$A \ \& \ H \ > \ SocSci \ > \ PSE \ \approx \ LifeSci$$

ConfSIs tell a different story, however. Physical sciences & engineering disciplines often referred to as the “hard” sciences (earth & space, engineering & technology, mathematics, and physics) have a distinctively higher publication frequency (one conference SI for each 45.45 issues) than other disciplines (one conference SI for each 166.67 issues). Given these strong contrasts between the different disciplinary groups with regard to SI publication practices, the same reading grid will be used when analyzing and comparing the topicality and impact results that follow.

### 3.1. Topicality Analysis

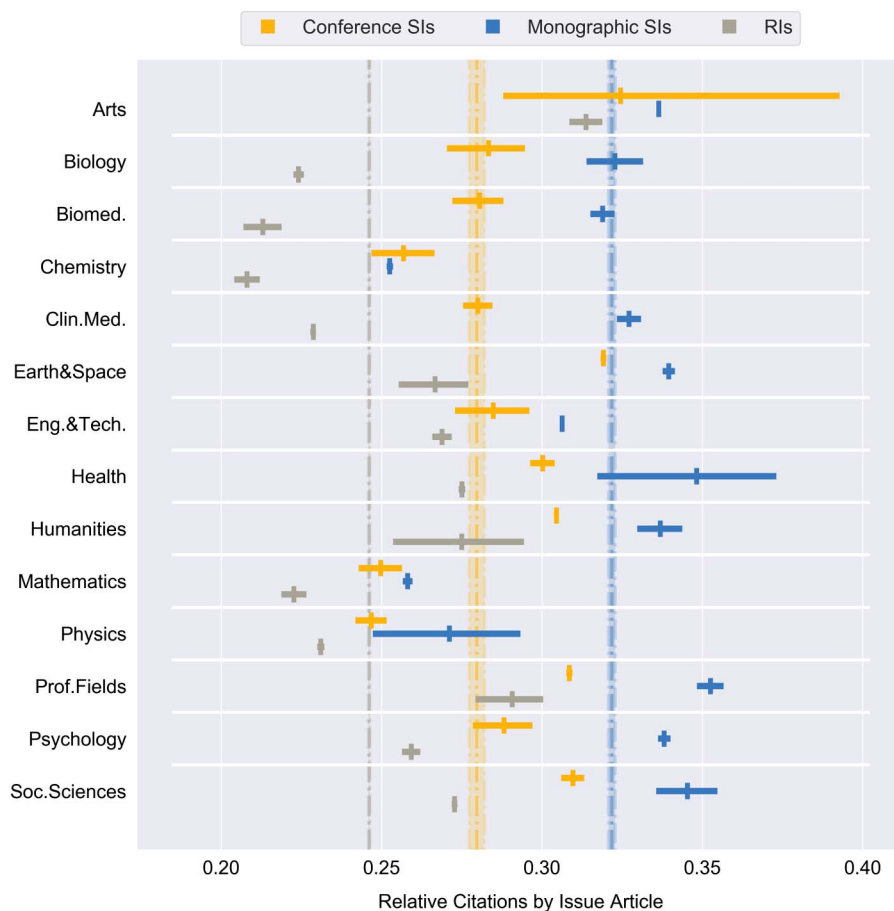
Topicality scores for titles and abstracts are shown logarithmically in Figure 2 and 3 respectively. Means for each discipline and issue type are indicated by small vertical lines, with horizontal confidence intervals on each side. Overall average and confidence intervals for each issue type are indicated by dotted lines and translucent bands in the background.





**Figure 2.** Intra-issue similarity scores for titles. *Disciplines:* Biomed. = Biomedical Research, Clin. Med. = Clinical Medicine, Eng.&Tech. = Engineering&Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

A first glance at Figure 2 shows that for the vast majority of disciplines, titles of articles copublished in SIs are topically closer than in RI context. In the case of MonoSIs, the trend is clear: Except for arts, all score averages are higher than for other issue types of the same discipline, and their confidence intervals never overlap with those of other issues types of the same discipline, save for chemistry and physics ConfSIs. Although ConfSIs also show higher topicality than RIs, the results are, however, less unequivocal, as the confidence intervals for engineering & technology, health, maths, and psychology overlap with the average scores of RIs, which sheds doubt on the results obtained for these disciplines. Finally, one cannot but notice that interdisciplinary variation in inraissue similarity scores differs between issue types. Average similarity scores for RIs range from .08 to .10 and are thus more homogeneous than those of ConfSIs and MonoSIs, whose ranges are respectively [.9, .13] and [.10, .17]; confidence intervals also tend to be shorter in the RI context. As for the different groups and issue types, the average similarity score in MonoSI titles is at its lowest in physical sciences & engineering and highest in life sciences disciplines for both MonoSIs ([.12, .16]) and ConfSIs ([.10, .12]) and from arts & humanities to social sciences ([.8, .10]) in the RI context. Beyond these issue types and disciplinary differences, however, overall averages and confidence intervals for the different issue types are adamant: Topicality in SIs titles is higher than for RI titles, and even more so in the case of MonoSIs.



**Figure 3.** Intra-issue similarity scores for abstracts. *Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. = Engineering&Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

With regard to abstracts, Figure 3 show trends similar to those reported for titles. Here again, the values and variation in similarity scores are almost always highest for MonoSIs, with the possible exception of chemistry and physics. As for ConfSIs, the average scores are closer to those of RIs, but still higher most of the time. Regarding confidence intervals, overlaps can be found in physics for MonoSIs, and arts, chemistry, and physics for ConfSIs. Also, in both SI contexts, social sciences, arts & humanities, and life sciences issues all tend to have higher than average inraissue similarity scores (.35, .43, and .32 for MonoSIs, and .31, .31, and .28 for ConfSIs respectively), whereas physical sciences & engineering issues are last in both respects (.29 and .27); as for RIs, life sciences (.22) & engineering (.24) are in the below-average region. Finally, overall averages and their respective confidence intervals clearly show that topicality is higher for MonoSIs than in the ConfSI context, and higher for the latter compared to RIs. Beyond these similarities, however, one cannot but notice that similarity scores for all issue types are higher than in the title context. This difference can be explained by the fact that abstracts contain more text than titles, which increases both the number of common substrings between articles of the same issue and thus the inraissue similarity scores for ConfSIs and MonoSIs, as well as RIs. However important it may be, this generalized increase in similarity scores does not really change the various trends observed for the different disciplines and issue types.

In sum, save for a few possible exceptions, articles published in SIs are topically closer to each other than articles published in RIs. Also, with the possible exception of arts and physics titles, as well as arts, chemistry, and physics abstracts, ConfSIs have lower topicality than their monographic counterpart. SIs from life sciences disciplines receive the biggest intraissue similarity boost compared with their RI counterparts, regardless of SI type or text field type. At the other end of the spectrum, physical sciences & engineering disciplines have the lowest SI topicality boosts of all disciplines. In fact, except for earth & space SI titles and articles, and possibly engineering & technology ConfSI abstracts, all scores from physical sciences & engineering are below average.

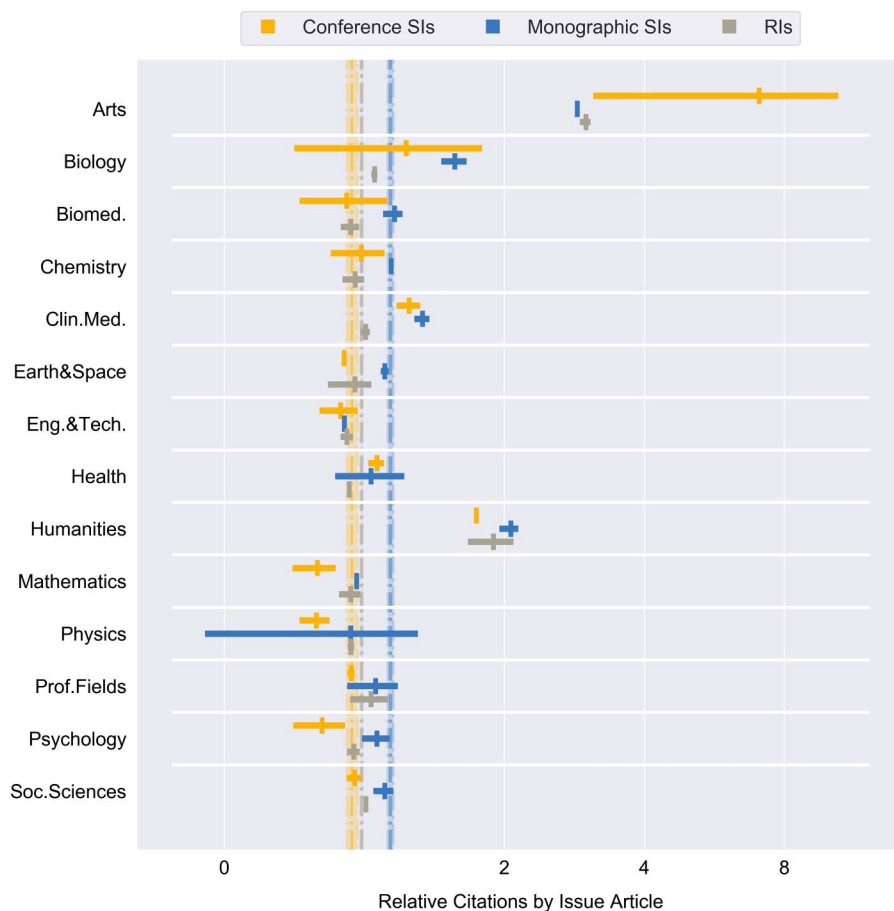
Beyond these disciplinary variations, however, the big picture is unequivocal and adamant: From a topical perspective and regardless of discipline, articles copublished in SIs are more similar to each other than those copublished in RIs. Given the higher topicality of SIs and their nonnegligible proportion in all disciplines, it is reasonable to assume that the top-down, planned initiatives leading to their publication do indeed have an impact on topical distributions, at both the disciplinary and general levels. However, whether or not this “topical drift” induced by SIs has a negative impact on the total research output remains to be seen. The following section will attempt a first step toward answering this question, by comparing the relative citation impact by issue article and intraissue variance of SIs and RIs in every discipline.

### **3.2. Citation Analysis**

Figure 5 shows RCIA scores for ConfSIs, MonoSIs, and RIs. RCIA averages are indicated by small vertical lines, with their confidence intervals on each side. The overall averages for each issue type and their confidence interval are respectively shown by vertical dotted lines and translucent bands. For visualization purposes, the x-axis has been scaled logarithmically.

Globally, comparison of citation rates of RIs and SIs provides a different portrait than that of topicality scores. For once, the tangle of confidence intervals is more the norm than the exception here, as every discipline has at least two issue type scores whose confidence intervals overlap with each other. All issue types from arts & humanities fare exceptionally well; given that the scores presented here are based on field-normalized citation counts, one possible explanation for this is that journals that publish SIs in that disciplinary group are in fact higher impact journals. Here again, issues from physical sciences & engineering disciplines score relatively poorly in most cases. Scores between issue types and disciplines also tend to be more homogeneous, whereas the issue type averages are closer to each other. Also, in contrast to the preceding analyses, SIs often fare worse than RIs, all the more so in the case of ConfSIs. This, however, may be due to the higher disciplinary variability and wider confidence intervals of both SI types, as aggregated averages for MonoSIs and ConfSIs are respectively higher and barely lower than those of RIs.

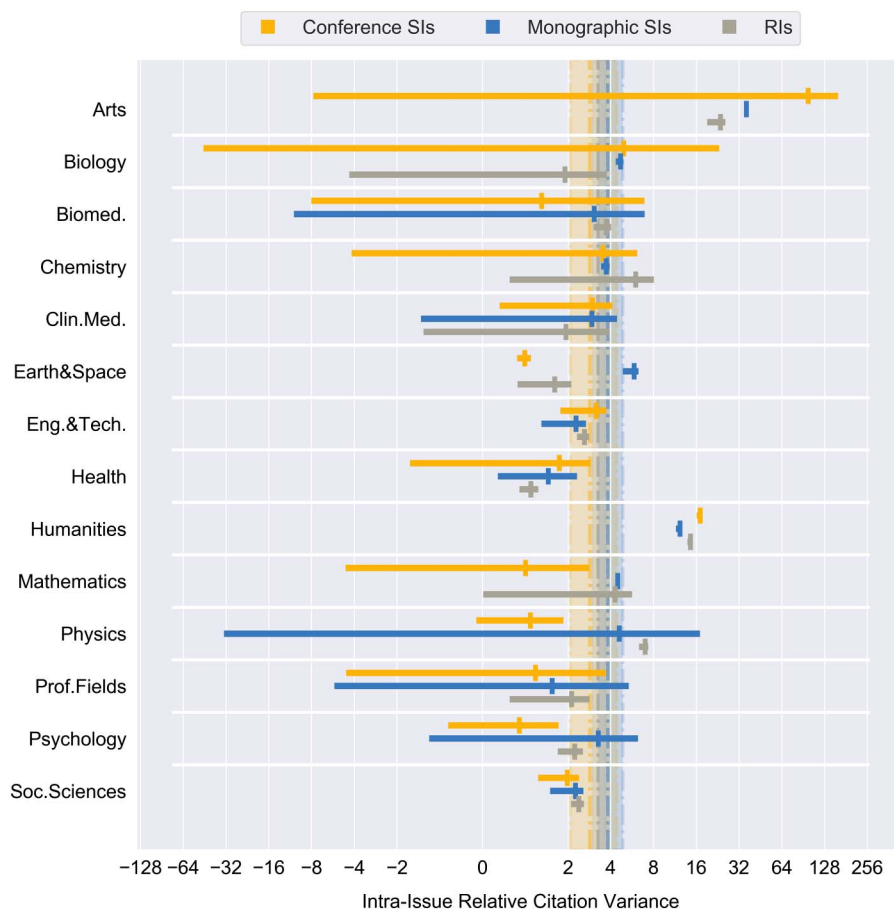
Looking closer at each issue type, MonoSIs published in arts journals tend to have lower relative citations by articles than articles published in RIs of the same year and discipline; engineering & technology, health, humanities, mathematics, and physics are also possible candidates. This lower research impact of MonoSIs in Arts journals certainly appears to be counterintuitive: The latter publish a higher proportion of monographic SIs than any other discipline, yet their articles receive fewer citations than regular ones. Another differential trend that stands out is the above- and below-average impact of life sciences (1.39) and social sciences (1.11) disciplines respectively. By contrast, the results for ConfSIs stand out in terms of confidence interval wideness, score variability, and score average, as most disciplinary



**Figure 4.** Relative citation by issue article scores for all disciplines. *Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. = Engineering&Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

scores are below those of RIs. Scores for life sciences disciplines are here again higher (1.20) than average for other disciplines (0.85). Arts stand out as being the only discipline or disciplinary group where the scores for ConfSIs are undoubtedly the highest; due to confidence interval overlaps, biology, health, and psychology are also possible cases. But what is really puzzling about arts ConfSIs is that despite having a higher impact than other issue types, these issues represent only 0.73% of all issues published in the field. Admittedly, the number of ConfSIs published by arts journals is certainly not as important as in more productive disciplines such as engineering & technology, but the trends are nonetheless striking. Finally, ConfSIs articles published in physical sciences & engineering journals tend to get lower citation scores (.8) than the issue type average as well as scores from RI articles of the same year and discipline. However interesting they may be, these lower level trends do not change the overall picture: MonoSIs globally receive more citations than RIs, while the latter attracts slightly more citations than ConfSIs.

Regarding research quality, Figure 5 shows the mean intraissue variance for all issue types and disciplines. Once again, mean values and their confidence intervals are indicated by vertical lines and horizontal bars respectively, while issue type averages are shown by dotted lines, flanked by their respective and translucent confidence intervals. As with the preceding plot, the x-axis has been scaled logarithmically.



**Figure 5.** Relative citation intraissue variance by discipline. *Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. = Engineering&Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

Here, as in Figure 4, wide and entangled confidence intervals abound. Overall, MonoSIs have a higher intraissue variance than ConfSIs, and their confidence intervals do not overlap. However, no clear conclusion can be reached regarding RIs and either SI type on this matter, as the confidence intervals of the latter overlap on each side. At the disciplinary level, conclusive results are scarce: The variance of MonoSIs is highest and lowest of all issue types in earth & space issues and humanities respectively, higher than that of ConfSIs in mathematics, and higher than that of RIs in arts and biology; also, variance is higher for RIs than ConfSIs in humanities and physics. However, the most striking feature of these results is certainly the high number of confidence intervals that cross over the vertical intercept of the graph. The means are of course all positive, but most empirical likelihood estimates for the lower bounds of all confidence intervals are surprisingly long and unbalanced, most likely due to the high skewness of the data distribution.

Given the inconclusiveness of these results, we conducted a Levene test with trimmed means to determine if issue type distributions for the different disciplines have equal variances. As the main homoscedasticity and heteroscedasticity tests are all null-hypothesis-based, we chose this test due to its proven robustness in dealing with heavy-tailed distributions (Brown & Forsythe, 1974; Derrick, Ruck, Toher, & White, 2018). Also, given that the question addressed here is whether or not the different SI types have greater intraissue variance than

**Table 2.** Brown-Forsythe test results for disciplinary distributions

Discipline	MonoSIs/RIs		ConfSIs/RIs	
	<i>W</i>	<i>p</i> -value	<i>W</i>	<i>p</i> -value
Arts	1.54	0.22	41.15	1.44e-09
Biology	0.59	0.44	10.67	1.12e-03
Biomed.	0.33	0.56	6.91	8.68e-03
Chemistry	2.81	0.09	20.76	5.74e-06
Clin.Med.	0.04	0.84	6.03	1.41e-02
Earth&Space	0.94	0.33	8.76	3.13e-03
Eng.&Tech.	3.63	0.06	22.88	1.78e-06
Health	0.21	0.64	16.58	5.52e-05
Humanities	0.01	0.92	10.52	1.22e-03
Mathematics	5.87	0.02	36.97	1.68e-09
Physics	5.47	0.02	16.11	6.31e-05
Prof.Fields	0.17	0.68	2.05	1.52e-01
Psychology	0.04	0.85	0.17	6.83e-01
Soc.Sciences	0.00	0.96	1.25	2.64e-01

*Disciplines:* Biomed. = Biomedical Research, Clin.Med. = Clinical Medicine, Eng.&Tech. = Engineering & Technology, Prof.Fields = Professional Fields, Soc.Sciences = Social Sciences.

RIs, two different tests were conducted: one for MonoSI and RI distributions and another for ConfSI and RI distributions. The results are shown in Table 2.

Using  $p = 0.001$  as an upper bound for heteroscedasticity, results indicate that MonoSI and RI distributions have equal variances in every discipline, while ConfSI distributions are distinct from the other two in a few cases, namely in arts, chemistry, engineering & technology, health, mathematics, and physics. In sum and from a bibliometric point of view, if we assume that higher quality papers tend to receive more citations than papers published in the same year and discipline, one cannot but acknowledge that the hypothesis that journal editors may be forced to accept substandard papers to “fill up” SIs does not hold: ConfSIs of many disciplines have lower intrainstance variance than RIs, while the results for MonoSIs are undistinguishable from those of RIs.

#### 4. DISCUSSION

As this research shows, Conference SIs, MonoSIs, and RIs represent distinct issue types, both in terms of topicality and impact. As regards to topicality, what is often assumed in the literature but never properly assessed is here demonstrated: The titles and abstract of articles copublished in issues that are integrally or partially special are more similar to each other than those copublished in RIs, even more so in the case of MonoSIs. At the disciplinary level, SIs from all disciplines have higher cosine similarity scores than their RI counterparts, with the possible exception of engineering & technology and psychology ConfSI titles, as well as arts ConfSI abstracts. However, SI “similarity score boosts” vary a lot depending on issue type and



discipline. Differences in editorial practices and choices may partly explain this variance: For example, SI topics may be broader in some cases or disciplines and narrower in others, and some fields may tend to include a greater or lower proportion of regular articles in what was previously called “mixed issues.” Such decisions would necessarily affect intraissue similarity scores; however, given that citation indexes and algorithms in their current state cannot allow for a proper assessment of topic scope or distinguish special and regular copublished articles, one can only speculate on these matters. Despite these shortcomings, the topical analyses conducted indicate that the commissioning of SIs clearly “distorts the marketplace for ideas” (Priem, 2006), a conclusion that further shows the usefulness and potential of word space modeling and computational semantics in general for bibliometric research purposes.

With regard to citation analysis, the results presented here suggest that the research impact of SIs is strongly dependent on issue type: MonoSIs attract more citations and show similar intraissue variance than RIs, while ConfSIs tend to show lower citation and citation variance scores than the latter. Disciplinary differences are also of paramount importance here. SIs in physical sciences & engineering tend to attract less impactful articles, as their RCIA scores are systematically lower than those of their RI counterparts. These results are reminiscent of those reported for conference proceedings, which have stressed the systematically lower scientific impact of the latter in various disciplines (Lisée, Larivière, & Archambault, 2008). On the other end, life sciences journals fare better at editing high-impact SIs, whether in MonoSI or ConfSI settings. These results, in line with the findings and conclusions previously reported in the literature for biology (Hendry & Peichel, 2016), are all the more surprising given the relatively low proportion of both MonoSIs and ConfSIs in all life sciences disciplines. Indeed, because SIs in these fields seem to attract higher impact articles, one would expect not only researchers, but also editors and publishers, to encourage their publication; yet our research shows that this is far from the case. Also intriguing is the case of SIs in arts journals: On the one hand, this discipline has the worst citation turnout for MonoSIs, yet has the highest proportion of such issues; on the other hand, ConfSIs in arts journals have the highest citation impact reported here, regardless of issue type and discipline, but represent less than 1% of all issues published over the observed period. The present research also sheds doubt on the positive citation impact of MonoSIs previously reported both for management (Conlon et al., 2006; Olk & Griffith, 2004), which is part of the professional fields discipline, and psychology (Sala et al., 2017) journals, as the current data fails to conclusively reproduce these results: In the first case, all confidence intervals overlap with each other; as for psychology, while confidence interval width makes the difference in impact between the different issue types marginal at best, the alleged higher impact of SIs can only apply to MonoSIs. But most importantly, this research shows that the assumption underpinning the main criticism on SIs does not hold: SIs do disturb the flow of ideas within the scholarly communication system, but with regard to the alleged opportunity cost related to this shift, the results show that monographic SIs tend to have more citations than RIs, while the loss for conference SIs is only marginal. In addition to this, the present research has invalidated the hypothesis according to which SI editors “fill up” issues by accepting substandard papers, as intraissue variance in citation scores for SIs is either lower than or undistinguishable from those of RIs published in the same year and field. In our view, these different refutations demonstrate the importance of large-scale quantitative studies in tackling issues pertaining to scholarly communication, culture, and practices, as personal experiences or even findings reported by small-scale or regional studies often cannot be generalized at the disciplinary or international level due to their limited confirmatory and explanatory scope.

Beyond these dissonances, however, the present results agree with the existing literature on SIs in conferring on that publication type a special status within the scholarly communication

system, be it from an empirical or theoretical standpoint. Indeed, and as Mowday's long-overdue proposal for an SI on SIs (Mowday, 2007) implies, SIs are far more interesting and worthwhile than the paucity of literature thereon suggests. In our opinion, this scarcity is not so much related to the subject matter itself as to enduring data-related deficiencies in the identification and indexation of SI-related information. In support of this claim and in order to both contribute to the study of SIs and attest to the importance of their proper indexing, we aim to pursue the current line of research further. In this regard, research opportunities abound, from large-scale and cross-disciplinary extensions of previous SI-related analyses to the investigation of unexplored themes of high relevance to science studies, such as authorship practices and gender biases. This line of research would be of obvious relevance not only to the field of bibliometrics, but also to the scholarly community in general, as a better use of SIs and a better knowledge of their impact can benefit both authors and journals in ways that may go beyond what can currently be quantified or measured.

#### ACKNOWLEDGMENTS

This research has benefited from relevant and insightful comments from both Professor Doctor Ludo Waltman and Professor Andrew Hendry, to whom we are much indebted.

#### AUTHOR CONTRIBUTIONS

Maxime Sainte-Marie: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Supervision; Validation; Visualization; Writing—original draft; Writing—review & editing. Philippe Mongeon: Conceptualization; Data curation; Investigation; Methodology; Validation; Writing—review & editing. Vincent Larivière: Conceptualization; Funding acquisition; Resources; Validation; Writing—review & editing.

#### COMPETING INTERESTS

The authors have declared that no competing interests exist.

#### FUNDING INFORMATION

This research was funded by the Canada Research Chair program and the Social Sciences and Humanities Research Council of Canada.

#### DATA AVAILABILITY

Data and scripts used in the research are available at [osf.io/6rezg/](https://osf.io/6rezg/).

#### REFERENCES

- Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4), 673–721.
- Brown, M., & Forsythe, A. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 364–367.
- Cavnar, W. B., Trenkle, J. M., et al. (1994). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval* (Las Vegas, NV) (pp. 161–175).
- Cheng, C., Liu, Y., & Liu, Z. (2016). Empirical likelihood ratio under infinite second moment. *Communications in Statistics – Theory and Methods*, 46(14), 6909–6915.
- 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.
- Cope, B., & Phillips, A. (Eds.). (2014). *The future of the academic journal* (2nd ed.). Oxford: Chandos.
- Cross, J., & Jansz, K. (2009). Thomson Reuter's regional expansion programme. *Editors' Bulletin*, 5(2), 64–68.
- D'Agostino, R., Belanger, A., & D'Agostino, R. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4), 316–321.
- Damashek, M. (1995). Gauging similarity with n-grams: Language-independent categorization of text. *Science*(267), 843–846.
- Derrick, B., Ruck, A., Toher, D., & White, P. (2018). Tests for equality of variances between two samples which contain both paired

- observations and independent observations. *Journal of Applied Quantitative Methods*, 13(2), 36–47.
- Di Ciccio, T., Hall, P., & Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *The Annals of Statistics*, 19(2), 1053–1061.
- Eden, L. (2010). Adding spice to our scholarly journals: The JIBS experience. *International Studies Quarterly*, 54(3), 901–907.
- Fleck, L. (2012). *Genesis and development of a scientific fact*. Chicago, IL: University of Chicago Press.
- Gärdenfors, P. (2004). *Conceptual Spaces: The geometry of thought*. Cambridge, MA: MIT Press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.
- Hall, P., & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58, 109–127.
- Hall, P., & LePage, R. (1996). On bootstrap estimation of the distribution of the studentized mean. *Annals of the Institute of the Statistical Mathematics*, 48, 403–421.
- Hart, A. (2001). Mann–Whitney test is not just a test of medians: Differences in spread can be important. *British Medical Journal*, 323(7309), 391–393.
- Hendry, A. (2016, February 15). *Special issues: What are they good for? Eco-Evo Evo-Eco: Academic musing by Hendry, Bolnick, and awesome guests*. [Blog Entry]. Retrieved from <http://ecoevoevoeco.blogspot.com/2016/07/i-have-helped-to-edit-nine-journal.html>
- Hendry, A., & Peichel, C. (2016). Are conference special issues worthwhile? *Evolutionary Ecology Research*, 17, 141–155.
- Lisée, C., Larivière, V., & Archambault, É. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology*, 59(11), 1776–1784.
- Mann, H., & Whitney, D. (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(1–2), 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.
- Mowday, R. T. (2007). Vertical integration as an explanation for the creation of special journal issues. *Journal of Management Inquiry*, 16(3), 250–253.
- 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.
- National Science Foundation. (2006). *Science and Engineering Indicators. Chapter 5: Academic Research and Development. Data and Terminology*. Retrieved from <http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1>
- Olk, P., & Griffith, T. L. (2004). Creating and disseminating knowledge among organizational scholars: The role of special issues. *Organization Science*, 15(1), 120–129.
- Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237–249.
- Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18, 90–120.
- Owen, A. (2001). *Empirical likelihood*. Boca Raton, FL: Chapman and Hall.
- Priem, R. L. (2006). What happens when special issues just aren't "special" anymore? *Journal of Management Inquiry*, 15(4), 383–388.
- Priem, R. L. (2007). Let's make special issues "special" once again. *Journal of Management Inquiry*, 16(3), 246–249.
- Rynes, S. (2003). Special research forums: Past, present and future. *Academy of Management Journal*, 46, 535–537.
- Sahlgren, M. (2006). *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces* (Unpublished doctoral dissertation).
- Sala, F. G., Lluch, J. O., Gil, F. T., & Ortega, M. P. (2017). Characteristics of monographic special issues in Ibero-American psychology journals: Visibility and relevance for authors and publishers. *Scientometrics*, 112(2), 1069–1077.
- Schütze, H. (1993). Word space. In S. Hanson, J. Cowan, & C. Giles (Eds.), *Advances in neural information processing systems* (Vol. 5, pp. 895–902). San Mateo, CA: Morgan Kaufmann.
- Sigué, S. (2011). On special issues. *Journal of African Business*, 12(3), 305–307. Retrieved from <https://doi.org/10.1080/15228916.2011.620919>
- Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *The Annals of Statistics*, 32, 1215–1221.
- Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.
- Tursunaliyeva, A., & Silvapulle, P. (2009). *Estimation of confidence intervals for the mean of heavy tailed loss distributions: A comparative study using a simulation method*. Retrieved from <https://tinyurl.com/yjt4tept>
- 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.
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851–872.
- Welch, B. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1–2), 28–35.
- Widdows, D. (2004). *Geometry and meaning* (Vol. 773). Stanford: CSLI publications.