National and international scientific elites: an analysis of Chinese scholars[1]

Fei Shu[*], Vincent Larivière[**] and Charles-Antoine Julien[*]

[*]fei.shu@mail.mcgill.ca; charles.julien@mcgill.ca
School of Information Studies, McGill University, 3661 Peel, Montreal (Canada)

[**] vincent.lariviere@umontreal.ca
École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal (Canada)

ABSTRACT

The purpose of this study is to compare the WoS with a national Chinese bibliometric database at the level of individual authors and measure the extent of the overlap of the group of authors that are the most active in the two data sources. The results indicate that Chinese scholars do not have homogeneous publication patterns: some very productive scholars mostly publish in international (WoS) journals while others prefer to diffuse their research results in national Chinese journals. Disciplines that are most international in scope exhibit a much higher level of overlap than those of the social sciences and humanities. These results suggest that the WoS does not accurately represent Chinese research activities, especially in social science and humanities, but that it also has a relative overlap with the Chinese national scientific literature in the natural and medical sciences.

INTRODUCTION

Over the last 20 years, China's contribution to the world's scientific activity—as measured by its number of Web of Science (WoS) publications—has increased at an impressive rate (Zhou, 2013). While part of this trend might be due to an increase in the number research papers written in English by Chinese researchers (Montgomery, 2013), some Chinese scholars might still prefer to publish their manuscripts in Chinese academic journals (Jin, Zhang, Chen, & Zhu, 2002; Moed, 2002b) only indexed by national Chinese bibliometric databases. Hence, measuring China's research output remains a challenge, as no bibliometric database covers both Chinese and English scientific literature.

Many scholars have, thus, concluded that the WoS is not an appropriate tool to measure Chinese research performance (Guan & He, 2005; Jin & Rousseau, 2004; Zhou & Leydesdorff, 2007), as significant differences were found in the coverage of national and international bibliometric databases (Hennemann, Wang, & Liefner, 2011; Meho & Yang, 2007). While previous work has attempted to explain differences between WoS and Chinese bibliometric databases by looking at journal hierarchies and citation relations (Zhou & Leydesdorff, 2007), or regional publications (Liang, 2003), no research has yet analysed the discrepancies at the level of authors. For instance, little is not known on the extent to which scholars from Chinese institutions publish their articles in international journals, or whether "top" Chinese authors give up publishing papers in Chinese in order to be more visible

---

internationally. A better understanding of those trends might help to explain the differences between the international and Chinese bibliometric databases.

RESEARCH QUESTIONS

The purpose of this study is to 1) compare the WoS with a national Chinese bibliometric database at the level of individual authors, 2) measure the extent of the overlap of the group of authors that are the most active in the two data sources, and 3) compare the results across certain disciplines. Specifically, this study will answer following research questions:

1. In a given discipline, are the top Chinese scholars in terms of the number of publication in WoS also top scholars in the selected Chinese bibliometric database?
2. In a given discipline, are the rankings of Chinese scholars in WoS equivalent to their rankings in the selected Chinese bibliometric database?
3. How does the discipline affect the overlap between top Chinese scholars in international versus national Chinese databases?

In addition to providing a better understanding of Chinese researchers' publication patterns, these results will inform the extent to which the WoS can accurately represent Chinese scientific activity, and how this varies across disciplines.

METHODOLOGY

The Web of Science (WoS) and China Citation Indicators Database (CSI) are used as data sources because of their coverage and representation. WoS is the only bibliometric database covering a century of citation-based indicators for all disciplines, as well as, since 1973, all authors and their institutional affiliations (Moed, 2005). Along these lines, most previous bibliometric studies on China are based on WoS (Zhou & Leydesdorff, 2007). Although there are five major bibliometric databases in China, CSI has the largest coverage and offers author rankings in terms of publications and citations that are not provided by other databases (Zhao, Lei, Ma, & Qiu, 2008).

Inconsistencies are found between WoS and CSI in terms of the discipline classification. WoS assigns journals to 232 subject categories while the CSI classifies journals into 35 fields and 457 subfields. Based on the descriptions of each subject category, 114 one-to-one matches between WoS and CSI were identified. This study analyses data for these 114 disciplines in both WoS and CSI. For each of these disciplines, the number of papers per author was compiled in order to produce ranked lists of top Chinese authors in WoS and CSI. Since CSI only offers bibliometric data from 2000 onwards, the top 100 authors between 2000 and 2014 in the 114 identified disciplines will be retrieved, and form 114 pairs of author groups for comparison. A total of 22,800 records (114 disciplines × 100 authors × 2 databases) were retrieved. The amount of overlap between these 114 pairs is of particular interest since it indicates if, for any of the 114 disciplines, the elite of scholars (i.e. top 100 authors) found in the WoS is the same as the one found in the CSI.

Author name ambiguity is a known threat to the reliability and validity of bibliometric analysis (Moed, 2002a; Strotmann & Zhao, 2012). This is especially the case for Chinese author, as traditional author identification using surname and given name initial(s) cannot distinguish Chinese authors having the same translated English name (Qiu, 2008; Strotmann & Zhao, 2012). In this study, in order to identify the Chinese authors, a Chinese author name database (CAND) has been developed based on the 11,400 records retrieved from CSI. The author's Chinese name, as well as all possible translated English names, and the author's

affiliation were recorded. Chinese authors' English names in WoS were then distinguished by matching the data in CAND. These records were validated manually to produce a list of author name pairs groups in each of the disciplines. Groups of top author groups were then compared in terms of their overlap (i.e., percent of shared authors) and the rankings of individual authors. While the overlap between top authors answers the first research question, the difference in rankings between the groups provides insights on the second research questions. The results of the comparison in different disciplines will answer the third research question.

PRELIMINARY RESULTS

Table 1 lists the 10 specialties selected for the pilot test[2]; they represent the disciplines of engineering (2), natural sciences, (3) medical sciences (2), social sciences (2) and humanities (1). For each discipline, the lists of top 100 authors in terms of number of publications were retrieved from both WoS and CSI.

Table 1. Number of Chinese papers per discipline, WoS and CSI

Category	Discipline	Papers indexed in WoS (2014)				Papers in CSI (2014)
		From China	Total	China's Ratio	China's Rank	
Engineering	Automation & Control System	2,573	8,261	31.1%	1	20,777
	Engineering, Environmental	2,820	11,240	25.1%	1	21,922
Natural Sciences	Chemistry, Physics	14,730	56,991	25.8%	1	4,133
	Mathematics	4,890	25,296	19.3%	2	12,797
	Optics	8,371	27,125	30.9%	1	2,623
Medical Sciences	Gastroenterology & Hepatology	1,648	11,304	14.6%	2	10,612
	Oncology	8,907	34,960	25.5%	2	42,054
Social Sciences	Management	825	8,528	9.7%	3	3,243
	Education & Education Research	354	9,820	3.6%	9	71,531
Humanities	Linguistics	248	4,795	5.2%	7	3,092

The preliminary results show overlaps between the top 100 authors in the WoS and CSI groups and the Kendall's Tau correlation coefficient between the author rankings. The overlaps between top 10, 20, 50 authors between WoS and CSI groups were also measured as reference.

As Table 2 shown, the overlaps vary from 1% (in *Education & Education Research*) to 21% (in *Engineering, Environmental*). Discipline has a strong effect on the size of the overlap between the groups: while overlaps in engineering and medical science are around 20%, those in the social sciences and humanities are less than 10%. As one might expect, the composition and rank of Chinese scientific elites in WoS and CSI are more closely related in the medical sciences as compared with other categories. The negative correlations were found in three disciplines including two natural science disciplines: *Chemistry, Physics* and *Optics*.

---

[2] The CSI data was retrieved for the time period from 2000 to 2015 due to the limited access to the database.

Table 2 The Comparison of 10 Disciplines between WoS and CSI

Category	Discipline	Kendall's Tau Correlation	Overlap in			
			Top 10	Top 20	Top 50	Top 100
Engineering	Automation & Control System	0.111	1	4	8	19
	Engineering, Environmental	-0.134	1	2	12	21
Natural Sciences	Chemistry, Physics	-0.340	0	0	5	18
	Mathematics	0.200	0	1	4	6
	Optics	-0.448	0	0	3	12
Medical Sciences	Gastroenterology & Hepatology	0.162	1	3	10	20
	Oncology	0.517	2	3	8	18
Social Sciences	Management	0.287	0	2	4	8
	Education & Education Research	N/A	0	0	1	1
Humanities	Linguistics	0.300	0	0	0	5

CONCLUSION

This research in progress paper has shown that Chinese scholars do not have homogeneous publication patterns. While some very productive—or *elite scholars*—mostly publish in international (WoS) journals; others prefer to diffuse their research results in national Chinese journals. Unsurprisingly, disciplines that are most international in scope—such as those of the natural and medical sciences—exhibit a much higher level of overlap than those of the social sciences and humanities. Along these lines, we find a higher overlap of top authors between WoS and CSI in disciplines where the proportion of China's contribution to WoS is higher. On the whole, these results suggest that the WoS does not accurately represent Chinese research activities, especially in social science and humanities, but that it also has a relative overlap with the Chinese national scientific literature in the natural and medical sciences.

These preliminary results are solely based on the ranking of top authors according to their number of publications, and do not consider the impact of research (i.e. citation). Future work will investigate the Chinese scientific elite according to its average number of citations per paper and total number of citations.

REFERENCES

Guan, J., & He, Y. (2005). Comparison and evaluation of domestic and international outputs in Information Science & Technology research of China. *Scientometrics, 65*(2), 215-244.

Hennemann, S., Wang, T., & Liefner, I. (2011). Measuring regional science networks in China: a comparison of international and domestic bibliographic data sources. *Scientometrics, 88*(2), 535-554.

Jin, B., & Rousseau, R. (2004). Evaluation of research performance and scientometric indicators in China. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research : the use of publication and patent statistics in studies of S&T systems* (pp. 497-514). Dordrecht; London: Kluwer Academic Publishers.

Jin, B., Zhang, J., Chen, D., & Zhu, X. (2002). Development of the Chinese Scientometric Indicators (CSI). *Scientometrics, 54*(1), 145-154.

Liang, L. (2003). Evaluating China's research performance: how do SCI and Chinese indexes compare? *Interdisciplinary Science Reviews, 28*(1), 38-43.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *ASI Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Moed, H. (2002a). The impact-factors debate: the ISI's uses and limits. *Nature, 415*(6873), 731-732.

Moed, H. (2002b). Measuring China's research performance using the Science Citation Index. *Scientometrics, 53*(3), 281-296.

Moed, H. (2005). *Citation analysis in research evaluation*. Dordrecht, UK: Springer.

Montgomery, S. L. (2013). *Does science need a global language? English and the future of research*. Chicago, IL: The University of Chicago Press.

Qiu, J. (2008). Scientific publishing: Identity crisis. *Nature, 451*(7180), 766-767.

Strotmann, A., & Zhao, D. (2012). Author Name Disambiguation: What Difference Does it Make in Author-Based Citation Analysis? *Journal of the American Society for Information Science and Technology, 63*(9), 1820-1833.

Zhao, R.-Y., Lei, j., Ma, R., & Qiu, J. (2008). A comparison among five bibliometric database in terms of citation index. *Information Studies: Theory and Application, 2008*(4), 589-605.

Zhou, P. (2013). The growth momentum of China in producing international scientific publications seems to have slowed down. *Information Processing & Management, 49*(5), 1049-1051.

Zhou, P., & Leydesdorff, L. (2007). A comparison between the China Scientific and Technical Papers and Citations Database and the Science Citation Index in terms of journal hierarchies and interjournal citation relations. *Journal of the American Society for Information Science and Technology, 58*(2), 223.