

How is CiteSpace used and cited in the literature? An analysis of the articles published in English and Chinese core journals

Xuelian Pan¹ Ming Cui² Xiaotong Yu³ Weina Hua⁴

¹ *xuelianpan@nju.edu.cn*
Nanjing University, Nanjing (China)

² *969010781@qq.com*
Nanjing University, Nanjing (China)

³ *981925271@qq.com*
Nanjing University, Nanjing (China)

⁴ *huawn@nju.edu.cn*
Nanjing University, Nanjing (China)

Abstract

This study investigates the use and citation of CiteSpace, a freely available tool for analysing, detecting and visualizing trends and patterns in scientific literature, by examining how it is used and cited among the articles published in English and Chinese core journals. Results show that CiteSpace is widely used in China along with a substantial uncitedness. The number of Chinese articles using CiteSpace is obviously increasing, while the citation rate of CiteSpace is not increasing over time. Many Chinese authors do not provide sufficient information for identifying CiteSpace. Findings also show that CiteSpace used in English core journal articles is more likely to receive citations than that used in Chinese core journal articles. Moreover, our results demonstrate that there are significant differences in citation counts between sections containing CiteSpace.

Conference Topic

Citation and co-citation analysis
Science policy and research assessment

Introduction

Software is important to scientific research: it assists scientists to identify research questions, analyse data, visualize the results and disseminate knowledge; indeed, “just about every step of scientific work is affected by software” (Howison et al., 2015, p. 454). However, the academic value of software has long been undervalued and, even worse, has been ignored in the current publication-driven scientific reward system. Recent years have witnessed a tremendous growth in software which is freely available for academic use (Hannay et al., 2009; Huang et al., 2013). As the value of data is increasingly recognized and a considerable amount of freely available software packages are used in the scientific community (Howison & Bullard, 2016; Thelwall & Kousha, 2016), some scholars argue that software should also be valued as an academic contribution (Hafer & Kirkpatrick, 2009; Piwovar, 2013). The US National Science Foundation (NSF) has recognized software as a valid research output since 2013 (NSF, 2013). Software has also been listed as a scholarly contribution in the UK Research Excellence Framework 2014 (Research Excellence Framework, 2013). Yet, many funding institutions, policy makers and administrators have not recognized software as a valid type of research products (Piwovar, 2013). Therefore, measuring the impact of software is imperative, because which will enable us to have a better understanding of the value of software and help to incorporate software as an integral component in research evaluations and scholarly communication.

Although citation counts is widely used to assess the impact of journal articles and monographs (Kousha, Thelwall, & Rezaie, 2011; Song & Kim, 2013; Cartes-Velásquez & Manterola Delgado, 2014), it still needs to investigate the current status of software citation practice before using citation counts to measure the impact of software. A few studies examining the citation of software in scientific articles have found that software citation practice is not yet customary. Howison's study on 90 biology articles has found that considerable software packages mentioned in the articles were not formally cited (Howison & Bullard, 2016). Our previous study on articles published in *PLOS ONE* has also found a similar result (Pan, Yan, & Hua, 2016). Therefore, it is worth investigating the usage of software in full-text scientific articles to demonstrate the academic impact of software. In this article, we examined how software was used and cited in different journal articles and factors affecting citation rate of software. CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace/>; Chen, 2004, 2006), a freely available software tool for analysing, detecting and visualizing trends and patterns in scientific literature, was chosen as the analysis target. As a representative bibliometric mapping analytical tool (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011), CiteSpace has been used around the world and obtained the most intensive usage in China (Ping, He, & Chen, 2017). We collected journal articles that mentioned CiteSpace by searching several databases including Web of Science and three Chinese databases, and then conducted a content analysis of these articles. We assessed the differences in citation rate of CiteSpace among journals and disciplines. Moreover, we examined whether there were differences between articles which mentioned CiteSpace in the topic fields (including title, keywords and abstract) and articles which did not mention CiteSpace in the topic fields. Overall, we aim to provide a solid foundation for understanding and improving software citation and description, which will enhance software's status as research outputs and benefit the developers of software.

Data and Methods

The Web of Science (WoS) database and three Chinese full-text journal article databases, including CNKI, Wanfang and CQVIP, were selected as data sources for articles mentioning CiteSpace. "CiteSpace" and "Cite Space" were used as search terms. The time span was limited to 2004-2016 when CiteSpace was developed and largely spread in the world. First, we searched WoS (limited to Science Citation Index Expanded, Social Science Citation Index and Arts & Humanities Citation Index) for all journal articles containing search terms in the topic fields with title, keywords and abstract. Then we retrieved and downloaded the full text of these articles. In total, we obtained a set of 50 full text articles after discarding non-research articles (e.g., editorials, summary of conferences, comments and letters). Second, we searched for search terms in the topic fields (the same as WoS) of CNKI, Wanfang and CQVIP databases, and then limited articles to these published in Chinese core journals. The journals indexed in the Chinese Science Citation Database (CSCD) or Chinese Social Science Citation Index (CSSCI) or Chinese Core Journals List compiled by Peking University (CCJLPU) were regarded as core journals in this study, because CSCD, CSSCI and CCJLPU are widely used for research evaluation in China. We also downloaded the full text of these articles and manually discarded non-research articles to get the second article set. Among the three Chinese full-text journal article databases, only CNKI database provides full text search. To obtain more papers mentioning CiteSpace, we searched CNKI for journal articles containing search terms in the full-text field, and then limited the journals to Chinese core journals. Later, we further refined the results by selecting research articles mentioning search terms only in paper body as the third article set. A total of 1,418 articles (50 from WoS database, 1,368 from Chinese databases) were collected for this study. The search work was ended on 22nd February 2017. Because the three Chinese databases use different subject

classification systems, a journal indexed in more than one of the databases might be assigned to different domains. Since the journals indexed in CSCD has been included in WoS, we assigned each of these journals with articles using CiteSpace to a unique Journal Citation Reports (JCR) subject area based on the subject area that most of its articles belong to.

A content analysis was conducted to investigate the use and citation of CiteSpace in scientific articles. A content-analytic coding scheme for mentions and citations of CiteSpace was created based on the work of Howison & Bullard (2016) and shown in Table 1. Three coders were trained to code these articles. Before they began to code the articles separately, the inter-coder reliability was measured by calculating Fleiss' kappa (Fleiss, 1971) using ReCal3 (<http://dfreelon.org/utis/recalfront/recal3/>; Freelon, 2010). The Fleiss's kappa was scored at 0.951, which is considered as an almost perfect agreement by Landis and Koch (1977). Finally, Chi-square tests were used to explore the differences of citing behaviour for CiteSpace between English and Chinese journals. In addition, we also explored the differences between sections containing CiteSpace for citation counts.

Table 1. Coding scheme for mentions and citations of CiteSpace.

<i>Code</i>	<i>Description</i>
PaperID	ID of a particular paper mentioning CiteSpace.
Position	Location mentioning CiteSpace, including Title, Keywords, Abstract, Body, Acknowledgement, and Supplement.
Used	Indicates whether CiteSpace is used in this research.
Version number	Particular version of CiteSpace used.
Developer	Mention of the developer of CiteSpace
URL	Web address of CiteSpace.
Citation	Denotes whether this paper provides a formal citation to CiteSpace in the reference list.
Reference entry	Denotes an entry linked to CiteSpace in a reference list.
Cite to publication	Denotes citation of a particular publication.
Cite to manual	Denotes citation of a specific user's guide or manual.
Cite to website	Denotes citation of URL, project name, and other information.

Results

Overview

We identified 45 full text articles using CiteSpace in the 50 articles from WoS and denoted these articles simply as *WoS group*. It was noting that this study focused on papers using CiteSpace rather than papers mentioning CiteSpace. That is, only articles that were affected by CiteSpace were included for analysis. We allocated 45 articles to countries based on first author's country of origin. We found that China, USA and Brazil had 30 (67% of 45 articles), 11 (24%) and 4 (9%) articles respectively. The left four articles separately came from Romania, Turkey, England and South Africa. This finding is consistent with a recent study on logs of interactive events, which found that China and USA were the top two countries with the most intensity usage of CiteSpace (Ping, He, & Chen, 2017). We identified 1,020 articles using CiteSpace from the gathered 1,368 Chinese articles. Figure 1 displays the distribution of the 1,020 articles across the 274 journals. We observed that most journals (84% out of 274) published less than five articles using CiteSpace, while four journals published more than 50 articles.

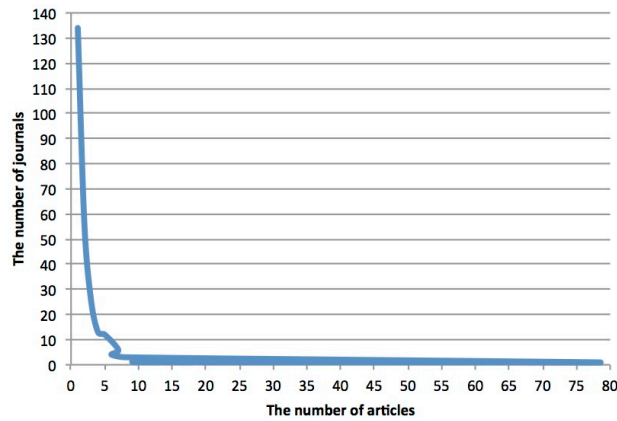


Figure 1. The distribution of 1,020 articles across Chinese core journals.

Characteristics of the mentions of CiteSpace

The 45 English journal articles are distributed across 18 JCR disciplines. Computer science, information and library sciences, and cell biology have more articles using CiteSpace than other disciplines. They have, respectively, 18, 4, and 4 articles. The 1,020 Chinese journal articles are distributed across 36 JCR disciplines. We found that most disciplines published less than 20 articles using CiteSpace. Table 2 presents the disciplines containing more than or equal to 20 articles using CiteSpace. It suggests that CiteSpace is more frequently used in information and library sciences, management, and education.

Table 2. The disciplines containing more than or equal to 20 articles using CiteSpace

Disciplines	The number of articles using CiteSpace
Information and library sciences	380
Management	158
Education	153
Sport sciences	51
Multidisciplinary social sciences	42
Geography	29
Economics	27
Communication	23
Environmental sciences	20

The descriptions of software provided in the articles are valuable for identifying and finding software. Among the 1,020 Chinese journal articles, 435 (43%) articles provided the version information of CiteSpace that they used in title, abstract, or article body. In addition, 433 (42%) and 21 (5%) articles, respectively, mentioned the author and the websites of CiteSpace in abstract or article body. It suggests that Chinese authors did not provide sufficient information for finding CiteSpace. In the WoS group, 17, 20, and 4 articles, respectively, provided version, author, and website information in the title, abstract, or article body. However, 282 (28% of 1020) articles in the Chinese article group and 20 (44% of 45) in the WoS group provided no further information than the name “CiteSpace” in the article body.

Characteristics of the citation of CiteSpace

Thirty-seven out of 45 (82%) English journal articles provided a formal citation of CiteSpace in the reference list, while 499 out of 1,020 (49%) Chinese journal articles made a formal citation to CiteSpace. The substantial higher citation rate of CiteSpace for the English journal articles might due to all the English journal articles containing CiteSpace in the topic field. In contrast to the citation rate of software (44%) found in the work of Howison & Bullard (2016), our citation rate of CiteSpace for Chinese journal articles is slightly higher. This might be explained by citable items are provided in the website of CiteSpace.

We then examined the cited references related to CiteSpace. Among the 37 English journal articles, 34 articles cited a publication related to CiteSpace, 2 provided a citation to a website, and 1 cited a user's manual. It shows that authors are more likely to cite publications than websites and user's manuals. Twenty-eight articles cite the article "CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature", which was published in *Journal of American Society for Information Science and Technology* in 2006. We also found that authors who published in Chinese journals prefer to cite related publications; 464 articles (93% out of 499 articles) cited a related publication. The top 10 most highly cited items for CiteSpace within Chinese journal articles are shown in Table 3. It is worth noting that the second most highly cited article is the Chinese version of the first most highly cited article. Among the top 10 most highly cited items, four publications and the blog are in Chinese. These Chinese references contribute to the wide use of CiteSpace in China.

Table 3. The top 10 most highly cited items for CiteSpace within Chinese article set.

Cited item	The number of articles that cited this item
Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature[J]. <i>Journal of the American Society for Information Science and Technology</i> , 2006, 57(3): 359-377	286
陈超美. CiteSpace II: 科学文献中新趋势与新动态的识别与可视化[J]. <i>情报学报</i> , 2009(6): 401-421.	45
Chaomei Chen's Homepage. http://cluster.cis.drexel.edu/~cchen/citespace/	25
Chen C. Searching for intellectual turning points: Progressive knowledge domain visualization[J]. <i>Proceedings of the National Academy of Sciences</i> , 2004, 101(suppl 1): 5303-5310.	17
刘则渊, 陈悦, 侯海燕. 科学知识图谱: 方法与应用[M]. 北京: 人民出版社, 2008.	18
Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis[J]. <i>Journal of the Association for Information Science and Technology</i> , 2010, 61(7): 1386-1409.	15
陈悦,陈超美,胡志刚等. 引文空间分析原理与应用: CiteSpace 实用指南[M]. 北京: 科学出版社, 2014.	9
侯剑华, 陈悦. 战略管理学前沿演进可视化研究[J]. <i>科学学研究</i> , 2007, 25(A01): 15-21.	5
Chaomei Chen's blog. http://blog.sciencenet.cn/u/ChaomeiChen .	4
Chen Chaomei. The CiteSpace Manual. http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf .	4

We next compared the citation rates of CiteSpace between different article sets. Thirty-seven out of 45 (82%) English journal articles provided a formal citation of CiteSpace in the reference list, while four hundred and ninety-nine out of 1,020 (49%) Chinese journal articles made a formal citation to CiteSpace. The substantial higher citation rate of CiteSpace for the English journal articles might due to all the English journal articles containing CiteSpace in

the topic field. In contrast to the citation rate of software (44%) found in the work of Howison & Bullard (2016), our citation rate of CiteSpace for Chinese journal articles is slightly higher. This might be explained by citeable items are provided in the website of CiteSpace. We then classified the 1,020 Chinese journal articles into two groups based on whether they contained CiteSpace in the topic fields or not. The first group, including 528 articles that contained CiteSpace in the topic fields, was denoted as Chinese topic group. The second one, including 492 articles which did not contain CiteSpace in the topic fields, was denoted as Chinese main group. Table 4 shows the usage and citation of CiteSpace in the three article groups during the period of 2007-2016. We can find that there is an over upward trend in the number of articles using CiteSpace in Chinese journal article sets. The average citation rate of the three paper sets is 0.82, 0.59 and 0.39 respectively. The citation rate of CiteSpace out of WoS group is higher than that of two Chinese groups. Moreover, the citation rate of CiteSpace out of Chinese topic group is higher than that of Chinese main group.

Table 4. The usage and citation of CiteSpace in the scientific lecture over time.

Year	<i>WoS group</i>			<i>Chinese topic group</i>			<i>Chinese main group</i>		
	A	B	C	A	B	C	A	B	C
2007	0	0	/	2	2	1.00	1	1	1.00
2008	2	1	0.50	5	4	0.80	0	0	/
2009	0	0	/	12	9	0.75	8	4	0.50
2010	0	0	/	22	15	0.68	15	4	0.27
2011	4	4	1.00	42	30	0.71	24	16	0.67
2012	1	1	1.00	55	32	0.58	44	17	0.39
2013	4	4	1.00	52	38	0.73	75	35	0.47
2014	11	9	0.82	90	53	0.59	90	44	0.49
2015	13	10	0.73	104	51	0.49	105	31	0.30
2016	10	8	0.80	144	75	0.52	130	38	0.29
Total	45	37	0.82	528	309	0.59	492	190	0.39

Note: A indicates the number of papers using CiteSpace; B indicates the number of papers that provided a formal citation of CiteSpace in the reference list; C indicates the citation rate of CiteSpace which was calculated by B/A.

To further determine whether there is a statistical difference of citing behaviour for CiteSpace between English and Chinese journals, we employed Chi-square test to compare the citation rate of CiteSpace between *WoS group* and *Chinese topic group* using SPSS (SPSS, version 20; IBM Corp., Armonk, NY). A statistically significant difference between the two groups was found ($p < 0.05$, Table 5). CiteSpace used in English core journal articles is more likely to receive citations than that used in Chinese core journal articles. Moreover, we also assessed the differences between sections containing CiteSpace for citation counts by comparing the citation rate of CiteSpace between *Chinese topic group* and *Chinese main group*. We found that *Chinese topic group* has a lower uncitedness than *Chinese main group* ($p < 0.05$, Table 5). We hold that authors put CiteSpace in different sections based on the importance of CiteSpace to their research. That is, authors tend to mention CiteSpace in their articles' title, keywords and abstract when CiteSpace is very important to their research. While CiteSpace is not very important to their research, they are more likely to mention CiteSpace in article body. Therefore, to some extent, CiteSpace that is important to research is more likely to receive citations.

Table 5. Chi-square tests for comparison of differences in uncitedness.

Test	<i>WoS VS. Chinese topic group</i>	<i>Chinese topic VS. Chinese main group</i>
------	------------------------------------	---------------------------------------------

Chi-square	9.736	40.381
P	0.020	0.000

We classified the 1,020 Chinese articles, according to the disciplines that their journals belong to. We then calculate the citation rates of CiteSpace for the disciplines which contain more than or equal to 20 articles. As shown in Table 6, the citation rate of CiteSpace varies from one discipline to another ranging from 0.29 (Sport sciences) to 0.57 (Communication). It demonstrates that software citation practices are far from common within the scientific community. A lot of efforts are needed to examine software attribution and improve software citation practices.

Table 6. The citation rate of CiteSpace of each discipline

Disciplines	The citation rate of CiteSpace
Communication	0.57
Information and library sciences	0.52
Education	0.52
Environmental sciences	0.50
Management	0.49
Multidisciplinary social sciences	0.38
Geography	0.34
Economics	0.26
Sport sciences	0.29

Conclusion and Future Work

This paper is part of a large effort to examine the attribution, citation and impact of scientific software. In this article, we collected a total of 1,418 full-text articles from English and Chinese databases, and then undertook a context analysis of these articles to how CiteSpace is used and cited among the articles published in English and Chinese core journals. Moreover, we explored the differences of citing behaviour for CiteSpace between English and Chinese Journals, as well as the differences between sections containing CiteSpace for citation counts. Results show the number of Chinese articles using CiteSpace has increased year by year. CiteSpace was more frequently used in information and library sciences, management, and education. Our results also demonstrate that many Chinese authors did not provide sufficient information for identifying CiteSpace in their articles or provide a formal citation of CiteSpace in the reference list. In addition, findings show that there is a significant difference in citation rate of CiteSpace between English and Chinese Journal articles. Findings also demonstrate that articles containing CiteSpace in title, keywords and abstract are more likely to make a formal citation to CiteSpace than those containing CiteSpace in article body.

Our future work includes collecting more English full-text journal articles using CiteSpace to corroborate the findings of this study. We tend to gather still more articles citing the publications related to CiteSpace. We are also interested to explore the difference of citing behaviour for CiteSpace between Chinese and foreign authors. In addition, we would like to investigate how CiteSpace is diffused over time and across disciplines. Research helps to incorporate software as an integral component in research evaluations and scholarly communication.

References

- Cartes-Velásquez, R., & Manterola Delgado, C. (2014). Bibliometric analysis of articles published in ISI dental journals, 2007-2011. *Scientometrics*, 98(3), 2223–2233.

- Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, *101 Suppl*, 5303–5310.
- Chen, C. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Proceedings of the National Academy of Sciences*, *57*(3), 359–377.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, *62*(7), 1382-1402.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, *76*(5), 378-382.
- Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, *5*(1), 20–33.
- Hafer, L., & Kirkpatrick, A. E. (2009). Assessing open source software as a scholarly contribution. *Communications of the ACM*, *52*, 126. <http://doi.org/10.1145/1610252.1610285>
- Hannay, J. E., MacLeod, C., Singer, J., Langtangen, H. P., Pfahl, D., & Wilson, G. (2009). How do scientists develop and use scientific software? *Proceedings of the 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, SECSE 2009*, 1–8.
- Howison, J., & Bullard, J. (2016). Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. *Journal of the Association for Information Science and Technology*, *67*(9), 2137–2155.
- Howison, J., Deelman, E., McLennan, M. J., Da Silva, R. F., & Herbsleb, J. D. (2015). Understanding the scientific software ecosystem and its impact: Current and future measures. *Research Evaluation*, *24*(4), 454–470.
- Huang, X., Ding, X., Lee, C. P., Lu, T., Gu, N., & Hall, S. (2013). Meanings and Boundaries of Scientific Software Sharing. *Proc. Conf. Computer Supported Cooperative Work (CSCW)*, 423–434.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174.
- NSF. (2013). GPG summary of changes. Retrieved from https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp
- Pan, X., Yan, E., & Hua, W. (2016). Disciplinary differences of software use and impact in scientific literature. *Scientometrics*, *109*(3), 1593–1610.
- Ping, Q., He, J., & Chen, C. (2017). How many ways to use CiteSpace? A study of user interactive events over 14 months. *Journal of the Association for Information Science and Technology*.
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*(7431), 159-159.
- Research Excellence Framework. (2013). Output information requirements. Retrieved from: <http://www.ref.ac.uk/about/guidance/submittinresearchoutputs/>
- Song, M., & Kim, S. Y. (2013). Detecting the knowledge structure of bioinformatics by mining full-text collections. *Scientometrics*, *96*(1), 183–201. <http://doi.org/10.1007/s11192-012-0900-9>
- Thelwall, M., & Kousha, K. (2016). Academic software downloads from Google Code. *Information Research*, *21*(1), n1.