Bibliometric Analysis based on Highly Cited Papers in Computer Science

Fangyu Jiao¹, Ji Fang¹, Yutong Ci and Wenyan Tu⁺
¹Huazhong Univ Sci & Technol, Lib, Wuhan, Hubei, Peoples R China
⁺Corresponding author’s e-mail: wenyan@hust.edu.cn

Abstract. From the perspective of bibliometrics, the important indicators of 4282 ESI highly cited papers in the field of computer science are analysed in this article, including publication time, countries/regions, institutions, journals, and keywords. China Mainland and the USA occupy a dominant position. The institutions with high number of papers and citations are mainly from China Mainland and the USA. The number of highly cited papers in China Mainland is increasing year by year. Most of the highly cited papers are dominated by international cooperation. Clusters of keywords through VOSviewer is to show the research hotspots that focus on algorithm optimization, isogeometric analysis, 5G communication, cloud computing, group decision making, etc.

1. Introduction

Essential Science Indicators (ESI) is an analytical database built on the basis of academic literature and their references included in the core database (SCIE/SSCI) of Web of Science [1]. ESI Highly Cited Papers are high-level papers that reflect the level of ESI discipline construction. It greatly reduces the influence of factors such as subject type and publication time on the citation of papers. It represents the high-quality and influential scientific research results in the subject area [2-3]. The output of highly cited papers is one of the important evaluation indicators of scientific research strength. As an important indicator for evaluating the influence of researchers, academic institutions, national and regional scientific research, it has always received extensive international attention [4-8].

This article conducts an in-depth analysis of ESI highly cited papers in the field of computer science, with the purpose of providing reference for researchers to expand their research directions and participate in international cooperation.

2. Data sources and Research methods

2.1 Data source

Essential Science Indicators (ESI) is an analytical tool that helps to identify top-performing research in Web of Science Core Collection. ESI surveys more than 11,000 journals around the world, ranking authors, institutions, countries, and journals in 22 broad areas based on publication and citation performance [9]. Highly cited papers are the top 1% in each of the 22 ESI subject areas per year. They are based on the most recent 10 years of publications.

The data source is the current ESI data updated on January 21, 2021. Current ESI Data covers over a 10-year and 10-month period (January 1, 2010 - October 31, 2020). There are 4282 highly cited papers in the field of computer science in this study.
2.2 Research methods
Bibliometrics refers to an interdisciplinary science that uses mathematical and statistical methods to quantitatively analyze literature. It can use quantitative statistics to depict the overall development trend of a specific subject area and the cooperation network of research subjects [10-11]. This research mainly uses bibliometric indicators and analysis tools such as InCites and VOSviewer to analyze highly cited papers in the field of computer science, and obtains the main countries/regions, institutions, distribution of journals, and research hotspots of highly cited papers in the recent 10 years.

3. Bibliometric Distribution

3.1 Academic influence and distribution of publication time
In January 2021, there were 4,282 highly cited papers in the field of computer science. The total citations generated by these papers in the Web of Science core collection reached 1,156,172 (retrieved on March 10, 2021), and the average citations is 269.82. Among them, the proportion of papers published in the past five years (2015-2020) accounted for 63.33%. It shows that the technology in the field of computer science has developed rapidly. New research results are more likely to be noticed.

![Figure 1. Year distribution of ESI Highly Cited Papers in Computer Science](image)

3.2 Main Countries/Regions
These 4,282 highly cited papers belong to 96 countries/regions. China Mainland ranked first with 1,944 highly cited papers, accounting for 45.4% of the world. The top 10 countries/regions ranked by the number of highly cited papers are shown in Table 1. The number of highly cited papers from China Mainland (1,944) and USA (1,230) is far greater than other countries.

| Ranking | Country/Region   | Papers | %Papers | Total citations | Average citations | International Collaboration | %International Collaboration |
|---------|------------------|--------|---------|----------------|------------------|-----------------------------|-------------------------------|
| 1       | CHINA MAINLAND   | 1,944  | 45.40%  | 260,213        | 133.85           | 1,148                       | 59.05%                        |
| 2       | USA              | 1,230  | 28.72%  | 429,058        | 348.83           | 795                         | 64.63%                        |
| 3       | ENGLAND          | 450    | 10.51%  | 117,397        | 260.88           | 365                         | 81.11%                        |
| 4       | AUSTRALIA        | 400    | 9.34%   | 83,488         | 208.72           | 336                         | 84.00%                        |
| 5       | CANADA           | 313    | 7.31%   | 106,239        | 339.42           | 257                         | 84.11%                        |
| 6       | GERMANY          | 246    | 5.74%   | 84,909         | 345.16           | 194                         | 78.86%                        |
| 7       | SINGAPORE        | 228    | 5.32%   | 57,736         | 253.23           | 176                         | 77.19%                        |
| 8       | INDIA            | 192    | 4.48%   | 31,478         | 163.95           | 124                         | 64.58%                        |
| 9       | FRANCE           | 174    | 4.06%   | 72,977         | 419.41           | 143                         | 82.18%                        |
| 10      | SOUTH KOREA      | 171    | 3.99%   | 42,509         | 248.59           | 147                         | 85.96%                        |

When citations are used as indicators, a population or an aggregate of articles representing perhaps
a university department or a scientific area is usually being analyzed [9]. Although China Mainland ranks first in the number of papers, in terms of total citations, USA (429,058) is 1.65 times higher than China Mainland (260,213).

Among the top 10 countries, the top three average citations are France (419.41), USA (348.83), and Germany (345.16). China Mainland, which has the largest number of papers, has only 133.85 average citations. It shows that China has an advantage in the number of papers published, but the quality and attention of papers need to be improved.

Figure 2 shows the number of highly cited papers in different countries and different years. The number of highly cited papers in CHINA MAINLAND shows a linear upward trend, the United States keeps a stable level, while other countries shows no obvious change.

![Figure 2. Number of papers in each year of TOP10 countries/regions](image)

3.3 International Cooperation
As shown in the data (Table 1), the proportion of international collaborative papers in top 10 countries/regions all exceeds 50%, with the highest in SOUTH KOREA (85.96%) and the lowest in CHINA MAINLAND (59.05%).

Figure 3 is an international cooperation network of highly cited papers in the field of computer science. The nodes represent countries/regions, and the size of the nodes represents the number of papers in each country/region. The connections represent cooperation between countries/regions, and the thickness of the connection represents frequency of cooperation between countries/regions.

3.4 Distribution of Institutions
A total of 1,968 institutions around the world have published highly cited papers. The top 10 institutions in terms of total citations and the number of highly cited papers are shown in Table 2&Table 3. Among the top 10 institutions of total citations (Table 2): The top 6 are all from the USA; Among the Top 10 institutions of papers (Table 3): 5 institution are from CHINA MAINLAND.
Figure 3. International cooperation network of highly cited papers.

Table 2. Top 10 institutions in terms of total citation.

| Ranking | Institutions                     | Country/Region | Total citations |
|---------|----------------------------------|----------------|-----------------|
| 1       | University of California System  | USA            | 43,028          |
| 2       | Google Incorporated              | USA            | 38,775          |
| 3       | University of Wisconsin System   | USA            | 36,850          |
| 4       | University of Texas System       | USA            | 36,668          |
| 5       | University of Wisconsin Madison  | USA            | 36,444          |
| 6       | University of Texas Austin       | USA            | 31,288          |
| 7       | Universite Paris Saclay          | FRANCE         | 27,015          |
| 8       | Nanyang Technological University | SINGAPORE      | 26,497          |
| 9       | Chinese Academy of Sciences     | CHINA MAINLAND | 26,380          |
| 10      | ETH Zurich                      | SWITZERLAND    | 25,389          |

Table 3. Top 10 institutions in terms of the number of highly cited papers.

| Ranking | Institutions                                           | Country/Region         | Papers |
|---------|--------------------------------------------------------|------------------------|--------|
| 1       | Chinese Academy of Sciences                           | CHINA MAINLAND         | 186    |
| 2       | University of California System                        | USA                    | 115    |
| 2       | University of Texas System                             | USA                    | 115    |
| 4       | Southeast University                                   | CHINA MAINLAND         | 107    |
| 5       | University of Electronic Science & Technology of China | CHINA MAINLAND         | 102    |
| 6       | King Abdulaziz University                             | SAUDI ARABIA           | 93     |
| 7       | Nanyang Technological University                       | SINGAPORE              | 91     |
| 7       | Tsinghua University                                   | CHINA MAINLAND         | 91     |
| 9       | National University of Singapore                       | SINGAPORE              | 90     |
| 10      | Central South University                              | CHINA MAINLAND         | 83     |
3.5 Distribution of Journals

4,282 papers were published in 267 journals. Among the Top 10 journals, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS is the most authoritative review journal in the field of computer and communications, which listed with the highest JIF (Table 4). This journal’s latest impact factor is 23.7. It represents the academic frontier and some future development directions.

| Ranking | Journal                                                                 | Papers | Total citations | JIF   | Country/Region |
|---------|-------------------------------------------------------------------------|--------|-----------------|-------|----------------|
| 1       | IEEE TRANSACTIONS ON CYBERNETICS                                        | 251    | 27,982          | 11.079| USA            |
| 2       | IEEE COMMUNICATIONS SURVEYS AND TUTORIALS                               | 196    | 54,953          | 23.7  | USA            |
| 3       | INFORMATION SCIENCES                                                   | 195    | 30,662          | 5.91  | USA            |
| 4       | IEEE COMMUNICATIONS MAGAZINE                                            | 179    | 59,748          | 11.052| USA            |
| 5       | IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS              | 179    | 25,285          | 8.793 | USA            |
| 6       | IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS                           | 135    | 40,541          | 6.779 | USA            |
| 7       | IEEE INTERNET OF THINGS JOURNAL                                        | 113    | 16,884          | 9.936 | USA            |
| 8       | APPLIED SOFT COMPUTING                                                  | 101    | 16,961          | 5.472 | NETHERLANDS    |
| 9       | BMC BIOINFORMATICS                                                     | 97     | 51,400          | 3.242 | ENGLAND        |
| 10      | KNOWLEDGE-BASED SYSTEMS                                                | 93     | 15,709          | 5.921 | NETHERLANDS    |

4. Keyword clustering analysis

Research hotspots refer to the focus issues that researchers have paid attention to in a certain period of time. Visual metrological analysis can quantitatively analyze published research literature, and display the results in the form of a scientific knowledge map, revealing research hotspots and development trends in a certain area. This method has now been widely used in the research summary of the subject area. VOSviewer is a software that can avoid the mutual coverage of important nodes and labels in visual display, especially in cluster analysis [12].

In this study, 4,282 literature data were preprocessed and imported into VOSviewer software for calculation, and a total of 16,329 keywords were obtained. The keywords with the greatest total link strength will be selected. Set the minimum number of occurrences of a keyword to 30, there are 139 keywords that meet the threshold.

Combined with the frequency of words, 139 keywords form 5 clusters in the space. After synonym merging and non-meaningful words removal, the following effective keywords are obtained (Table 5).

| Number | Keywords                                                                 |
|--------|--------------------------------------------------------------------------|
| Cluster1 (27 items) | algorithm, classification, clustering, convergence, convolutional neural networks, database, decomposition, deep learning, diagnosis, differential evolution, diversity, evolutionary algorithms, feature extraction, feature selection, genetic algorithms, global optimization, identification, machine learning, multiobjective optimization, neural-network, particle swarm optimization, prediction, recognition, regression, segmentation, strategy, support vector machines |
| Cluster2 (20 items) | adaptive control, approximation, behavior, complex networks, criteria, delay, design, discrete, dynamic surface control, isogeometric analysis linear-systems, neural networks, nonlinear-systems, output-feedback control, robust, simulation, stability analysis, state estimation, synchronization, time-varying delays, tracking control |
| Cluster3 (17items) | 5G, cellular networks, channel estimation, cognitive radio, communication, energy efficiency, energy harvesting, interference, massive mimo, nonorthogonal multiple-access, performance analysis, power allocation, protocols, resource allocation, sensor networks, stochastic geometry, wireless networks |
5. Conclusion

This article explores the high-level output of scientific research in the field of computer science from the perspective of bibliometric: the leading countries/regions in scientific research are China Mainland, the USA, and England. Among them, China Mainland has the highest number of highly cited papers; The USA has the highest total citations. Most of the highly cited papers are dominated by international cooperation, and the proportion of international cooperation exceeds 50%.

The institutions with high number of papers and citations are mainly from the USA and China Mainland. The University of California System has the most citations. The Chinese Academy of Sciences has the most papers.

The journal IEEE COMMUNICATIONS SURVEYS AND TUTORIALS ranks among the top in terms of the number of highly cited papers and journal impact factors. It has a great academic influence in the field of computer science.

The research hotspots of highly cited papers in the field of computer science mainly focus on algorithm optimization, isogeometric analysis, 5G communication, cloud computing, group decision making, etc.

In the future, researchers can strengthen cooperation, share research results and pay more attention to research hotspots and frontiers in this field around the world, so as to produce more high-quality academic results and promote the further development of computer science.

References

[1] Huo,Y. (2003) Essential Science Indicators Dataset. Library and Information Service, (1):56-59.
[2] Liao,H., Tang,M., Li,Z., et al. (2018) Bibliometric Analysis for Highly Cited Papers in Operations Research and Management Science from 2008 to 2017 Based on Essential Science Indicators. Omega.
[3] Fu,H.Z., Chuang,K.Y., Wang,M.H., et al. (2011) Characteristics of research in China assessed with Essential Science Indicators. Scientometrics, 88(3):841-862.
[4] Liu,X. (2012) Method of defining highly cited papers based on web of science and esi database. Chinese Journal of Scientific and Technical Periodicals, (06): 975-978.
[5] Yi,H., Jie,W. (2011) A bibliometric study of the trend in articles related to eutrophication published in science citation index. Scientometrics, 89(3): 919-927.
[6] Wei,R., Chen,D. (2011) An Empirical Research on the Highly Cited Documents Based on the Citation Network. Journal of Modern Information, 31(3):117-121.
[7] Su,L.W., Qiao,L.L.(2019) Output and Impact Study of Highly Cited Papers Under International Academic Collaboration Based on ESI. Journal of Modern Information, 039(004):143-152.
[8] Wang,X.F., Chen,Y., Huang,Y., et al. (2016) Research of Highly Cited Papers of China Based on ESI Category Overlay Maps. Journal of Intelligence, 35(10):106-113.
[9] Aksnes,D.W. (2003) Characteristics of highly cited papers. Research Evaluation, (3):159-170.
[10] Liu,J.Y. (2013) Bibliometric Analysis of Engineering Science Based on ESI Database. Higher Education Development and Evaluation, (04):10-18.
[11] Qiu,J.P., Yang,R.X. (2010) Bibliometrics Research Based on ESI:A Case Study on Material Science Field. Information Science, (08):3-8.
[12] Eck,N., Waltman,L. (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2):523-538.