



Scientific Mapping of Recent Research Trends on Novel Coronavirus (COVID-19): A Classical Bibliometric Analysis

¹Prabhat Kumar, ²S Suresh

Department of Computer Science, Institute of Science
¹prabhat.kumar13@bhu.ac.in, ²suresh.selvam@bhu.ac.in
Banaras Hindu University, Varanasi – 221 005, India

Abstract

INTRODUCTION: The COVID-19 pandemic has been declared as major threats to human beings worldwide. This outbreak is standing as a public health emergency problem in front of Research and Development (R&D). The bibliometric evaluation assists to quantitatively explore the scientific works and redirection of research trends on COVID-19. To support the R&D activities, various reputed publishers have already made the scientific literature related to coronavirus (CoV) open access to all researchers. The main objective in this article, carried out to extract the most productive authors, corresponding author's country, total citations per country, relevant sources, growth of publication citations, and frequent author keyword occurrence on CoV research.

METHODS: This paper is extremely dedicated to exploring the efforts and contributions of scientific researchers. We have wrapped published articles of the last four weeks (27th October – 27th November 2020), accessed from Web of Science (WoS). The searching strategies and dataset preprocessing tasks have also been performed to improve the quality of the dataset. We have used the bibliometrix R-package and VOSviewer as bibliometric analysis tools. At the initial phase, we have retrieved a total of 8,286 literature including duplicate records.

RESULTS: The International Journal of Environmental Research and Public Health Journal have ranked as the most productive journal. An article published in 2020 titled "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China" by C. Huang et al., recognized as the most cited article ($C_{2020} = 909$). Among the most productive authors, Wang Y published the most articles and the research was highly qualitative in terms of most citations and highest h-index and g-index value. The USA has been awarded as the most productive country in terms of the most productive corresponding author's country. Research trends on CoV rotate around the keywords COVID-19, SARS-CoV-2, Coronavirus, Pandemic, SARS, Infection, Pneumonia, Depression, Outbreak, and Virus, which were the most common keywords found.

CONCLUSION: This article has extremely dedicated to the researchers who currently working on the COVID-19 to shrink the effect of these outbreaks. In this study, bibliometric methods have been used to measure the research trends on CoV.

Keywords: Coronavirus, COVID-19, Bibliometric Analysis Tools, SARS-Cov-2, Scientific Research Trends.

1. Introduction

December 2019 has witnessed as the first case of pneumonia due to unknown causes existed in Wuhan province, China, and later this cause identified as a novel coronavirus (COVID-19) on January 07, 2020, by World Health Organization (WHO) [1]. The spreading and infection rate of

COVID-19 was unpredictable, affecting almost all countries of the world including China within four months and introduced as a public health emergency [2]. As of 28-Nov-20, the COVID-19 pandemic has claimed total 62,086,291 cases, 480,140 deaths, 42,881,032 recovered and 17,754,073 currently infected patients on November 28, 2020, 10:40

GMT [3]. Sustainable development in scientific research fields leads the advanced knowledge and availability of resources for both developing and developed countries against public health emergencies throughout the world-wide [4]. The researchers deserved the superior responsibilities by paying attention to controlling the pandemics, find out the trusted source of disease, symptoms, how to differ from other infections and ways of infection transmission then efforts to reduce the effects through precautions and finally discover the treatments, vaccines, medicines and prevention procedures [5]–[7]. Since 1931, the subject of CoV has acquired research attention for bibliometric analysis. The global research trends are varied closely according to the situation of the pandemic. This is difficult to follow the maturity of research trends and keep an eye on up-to-date. The scientific study is an opportunistic way to identify more valuable and responsible academic patterns towards controlling the outbreaks. The bibliometric methods can be used to evaluate the assessment of research trends. We carried out the bibliometric analysis of scientific literature on COVID-19, available on Web of Science (WoS). The scientific literature is evaluated and examined for extracting the publication patterns on the behalf of paper titles, authors, journals, organizations, funding agencies, Countries/Regions, etc [8]. This paper is extremely dedicated to explore the efforts and contributions of scientific researchers. We have wrapped published articles of the last four weeks (27th October – 27th November 2020).

Our contributions in this literature are summarized as follows. The first section describes the strategy used for data acquisition, methodology, and bibliometric analysis tools used to produce the results of the scientomatic study. The next section illustrates the discussion and analyzed results based on most productivity authors, corresponding author's country, and total citations per country, relevant sources, cited publications, and frequent author keyword occurrence analysis. The last section concludes the summary of our contributions in this paper and future research directions.

2. Data acquisition strategies and methodology

In this section, we have detailed the data acquisition strategies from the academic repository (WoS) and methodology used for bibliometric analysis, which has also been discussed.

2.1. Data source

The bibliometric details of published research works were accessed from Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI), Conference Proceedings Citation Index-Science (CPCI-S), and Index Chemicus (IC) of the Clarivate Analytics's Web of Science (WoS) Core Collection. The dataset used in this literature were last accessed on November 27th, 2020. The bibliometric details such as literature title, authors, publication years, document types, funding agencies, source titles, countries/regions, research areas, abstracts, keywords, list of references,

volume, issues, page number, and some other details were exported and examined through scientometrical methods. The dataset contained literature i.e. existed in different document types such as journal articles, reviews, letters, case studies, preprints, news items, etc. The graphical representation of document types has shown in table 1. The total number of publications retrieved from WoS was 8,286. This was most probably that same articles may be existed more than one time. We need to preprocess the new dataset to improve consistency and integrity by removing duplicate records.

Table 1. Document types of scientific literature on COVID-19 research

Description	Results
Timespan	27 th October – 27 th November 2020
Articles	4484
Editorial Material	1120
Letters	1240
Meeting Abstracts	325
News Items	144
Reviews	888
Others	85
Total Documents	8286

2.2. Searching Strategy and Dataset preprocessing

The searching strategies have been used to retrieve the research works using keywords related to COVID-19 research and its variations related to the paper title and abstract of literature, as mentioned in table 2.

Table 2. Searching strategies used to retrieve the published works.

Database	Keywords	Number of publications
Web of Science (WoS)	TS = ("COVID-19" OR "CORONAVIRUS" OR "2019-NCOV" OR "SARS" OR "SARS-COV-2" OR "covid" OR "covid19")	8,286

A total of 8,286 research works were retrieved from the WoS database. The searched keywords "COVID-19" OR "CORONAVIRUS" OR "2019-NCOV" OR "SARS" OR "SARS-COV-2" OR "covid" OR "covid19" were used to identify and matched with scientific works of literature, obtained on November 27th, 2020. The real-world dataset is often incomplete and inconsistent and is likely to contain many errors and may produce irrelevant results. The preprocessing step is essential for improving the quality of the dataset and transforms it into the knowledge discovery form. At the initial phase, we have retrieved a total of 8,286 documents including the duplicate records. After verification of titles and abstracts, we have found a total of 55 duplicate

records. Finally, the duplicate records have been removed from the dataset and we retained 8,231 documents that met our inclusion criteria.

2.3. Bibliometric Analysis Methodology

The bibliometric analysis tools are used to measure the impact of research trends in a specific domain. The data preparation for analysis contained three basic steps, first retrieved the data from different bibliographic databases such as WoS (<http://www.webofknowledge.com>), Scopus (<https://www.scopus.com>), PubMed (<https://pubmed.ncbi.nlm.nih.gov>), Dimensions (<https://app.dimensions.ai>), etc. The second step is to combine and transform the data into a suitable format as applicable for bibliometric tools. The final step is data cleaning, which consists of several preprocessing methods such as removing the duplicate data, missing and misspelled data [9]. The data are analyzed with respect to most productive authors, top manuscripts per citation, corresponding author's countries, total citations per country, most relevant sources, most relevant keywords, and more. The analyzing methods are used to discover the most important knowledge from datasets such as bibliographic coupling, co-citation analysis, collaboration analysis, and co-word analysis. However, data visualization techniques are used to represent the map of the analyzed results. There are various relevant software tools are available for analysis the bibliometric data such as VantagePoint (<https://www.thevantagepoint.com>), Science of Science (Sci2) Tool [10], VOSviewer [11], CitNetExplorer [12], SciMAT [13] and CiteSpace [14].

We have used bibliometrix R-package (<https://bibliometrix.org>) and VOSviewer (<https://www.vosviewer.com>) for bibliometric analysis in this paper. The bibliometrix R-package is an open-source tool, enclosed bibliometric methods for analyzing the quantitative research in scientometrics and bibliometrics [9]. Moreover, the text mining algorithms can also be implemented with VOSviewer to draw the word map for the identification of the most used words.

3. Discussion and Results

The bibliometric dataset has acquired the records from the WoS database. The distribution of document types of datasets has already represented in figure 1. The analysis of the bibliometric dataset is unfolded using the following sub-sections.

3.1. Top 10 Most Productivity Authors

A total of 44,167 authors were contributed to all publication documents in which 828 documents are single-authorship and the rest of the documents are worked in collaboration. The collaboration index is reached at 05.95, measured on the basis of 01.24 documents situated per authors, and 06.45 Co-Authors per document.

Table 3. Most productive authors on COVID-19 research.

Author's Name	Total Citations	Total Publications	h-index	g-index
WANG Y	383	77	05	19
LI Y	133	66	04	11
ZHANG Y	397	64	08	19
LI J	285	60	05	16
LIU Y	130	51	04	11
WANG X	340	53	06	18
WANG J	486	50	05	22
LIU J	252	47	04	15
WANG L	452	47	05	21
CHEN Y	121	46	04	10

The parameters used in this literature to measure the most productive authors are total citations, total publications, h-index, and g-index in the dataset (Table 3). The author Wang Y has been awarded as a most productive author with 77 articles of total published articles followed by Li Y with 66 articles and Zhang Y with 64 articles. For total citations, Wang J has achieved the first rank with 486 citations, followed by ZHANG Y (64 publications and 397 citations) and WANG Y (77 publications and 383 citations). We have used the h-index and g-index parameter to measure the publication quality of authors. The h-index is the numerical value that is intended for both the productivity and researcher influence. On the basis of author quality, ZHANG Y ranked first with an index value of 08, followed authors are WANG X (h-index = 06). The g-index quantifies the scientific productivity based on publications and distribution of citations received by publications of researchers. The author WANG J scored first in g-index value i.e. 22, followed by WANG L with index value 21.

3.2. Top 10 Most Productive Corresponding author's country and Total Citations per Country

The affiliation of the first author's countries has been identified as the origin of published works [15]. Table 4 shows the corresponding author's country and analysis the articles with frequencies, the single country publication (SCP), multiple country publication (MCP), and Multiple Countries Publication (MCP) Ratio.

The USA has possessed as the most leading country of authors affiliation through the total 2,033 publications with a frequency of 0.263069, out of those 1684 were SCP and 349 were MCP with an MCP ratio of 0.1717. While the CHINA on the other hand with an MCP ratio of 0.2249, placed at second position showing although 679 and 197 publications were SCP and MCP respectively out of a total of 876 articles. The MCP ratio illustrates the working collaboration of a country with other countries. The highest value of the MCP ratio represents more group efforts among different countries. On the basis of the MCP ratio, CANADA has placed at the first position with a ratio of 0.3506, shows more country's affiliations.

Table 4. Most productive corresponding author's country

Country	Articles	Frequency	SCP	MC P	MCP Ratio
USA	2033	0.263069	4	349	0.171
CHINA	876	0.113354	679	197	0.224
ITALY	643	0.083204	526	117	0.182
UNITED KINGDOM	543	0.070264	371	172	0.316
INDIA	314	0.040631	255	59	0.187
SPAIN	304	0.039337	238	66	0.217
GERMAN	236	0.030538	165	71	0.300

Country	Articles	Frequency	SCP	MC P	MCP Ratio
CANADA	231	0.029891	150	81	0.350
FRANCE	220	0.028468	162	58	0.263
BRAZIL	213	0.027562	168	45	0.211

3.3. Top 10 Most Relevant Sources

Table 5 shown the top 10 sources (>50 documents), which contributed 11.94% of total documents (983 of 8,231 documents). International Journal of Environmental Research and Public Health Journal has published the highest number of articles on COVID-19 with 156 documents, followed by the BMJ-British Medical Journal. The third most productive source of publication was PLOS ONE with 111 documents.

Table 5. Top 10 most productive sources in coronavirus research

Sources	Documents
INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	156
BMJ-BRITISH MEDICAL JOURNAL	139
PLOS ONE	111
HEPATOLOGY	102
LANCET	99
SUSTAINABILITY	95
AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE	79
ANNALS OF EMERGENCY MEDICINE	70
JOURNAL OF CLINICAL MEDICINE	67
FRONTIERS IN PSYCHOLOGY	65

Table 6. Top 10 most cited articles on coronavirus research

Paper Title	Total Citations	Journal	Source
Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China	909	The Lancet	[16]
Clinical Characteristics of Coronavirus Disease 2019 in China	646	New England Journal of Medicine	[17]
Clinical course and risk factors for mortality of adult in patients with COVID-19 in Wuhan, China: a retrospective cohort study	554	The Lancet	[18]
A Novel Coronavirus from Patients with Pneumonia in China, 2019	524	New England Journal of Medicine	[19]
Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China	496	Journal of the American Medical Association	[20]
Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study	426	The Lancet	[21]
A pneumonia outbreak associated with a new coronavirus of probable bat origin	406	Nature	[22]

Table 7. Top 10 most occurred author's keyword occurred in CoV research

Keywords	Occurrences	Total Link Strength
COVID-19	3,066	15,161
SARS-CoV-2	1,078	6,257
Coronavirus	900	5,440
Pandemic	393	2,186
SARS	259	1,991
Infection	268	1,890
Pneumonia	259	1,769
Depression	257	1,351
Outbreak	177	1,252
Virus	155	1,239

4. Conclusion

This article has extremely dedicated to the researchers who currently working on the COVID-19 to shrink the effect of these outbreaks. In this study, bibliometric methods have been used to measure the research trends on CoV. We observed the markable growth in publishing the scientific literature related to CoV and the contribution of publishers for providing the literature open accessible. The International Journal of Environmental Research and Public Health Journal has ranked as the most productive journal. An article published in 2020 titled "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China" by C. Huang et al., recognized as the most cited article ($C_{2020} = 909$). Among the most productive authors, Wang Y published the most articles and the research was highly qualitative in terms of most citations and highest h-index and g-index value. The USA has been awarded as the most productive country in terms of the most productive corresponding author's country. Research trends on CoV rotate around the keywords COVID-19, SARS-CoV-2, Coronavirus, Pandemic, SARS, Infection, Pneumonia, Depression, Outbreak, and Virus, which were the most common keywords found.

References

[1] F.-X. F.-X. Lescure *et al.*, "Clinical and virological data of the first cases of COVID-19 in Europe: a case series," *LANCET Infect. Dis.*, vol. 20, no. 6, pp. 697–706, Jun. 2020, doi: 10.1016/S1473-3099(20)30200-0.

[2] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.

[3] "Coronavirus Update (Live): 62,086,291 Cases and 1,451,186 Deaths from COVID-19 Virus Pandemic - Worldometer." [Online]. Available: <https://www.worldometers.info/coronavirus/>. [Accessed: 28-Nov-2020].

[4] P. Howitt *et al.*, "Technologies for global health," *Lancet*, vol. 380, no. 9840, pp. 507–535, 2012, doi: 10.1016/S0140-6736(12)61127-1.

[5] W. Guan *et al.*, "Clinical characteristics of coronavirus disease 2019 in China," *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1708–1720, 2020, doi: 10.1056/NEJMoa2002032.

[6] M. L. Holshue *et al.*, "First case of 2019 novel coronavirus in the United States," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 929–936, 2020, doi: 10.1056/NEJMoa2001191.

[7] C.-C. Lai *et al.*, "COVID-19 in long-term care facilities: An upcoming threat that cannot be ignored," *J. Microbiol. Immunol. Infect.*, vol. 53, no. 3, pp. 444–446, Jun. 2020, doi: 10.1016/j.jmii.2020.04.008.

[8] R. N. Broadus, "Toward a definition of 'bibliometrics,'" *Scientometrics*, vol. 12, no. 5–6, pp. 373–379, 1987, doi: 10.1007/BF02016680.

[9] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, pp. 959–975, 2017, doi: 10.1016/j.joi.2017.08.007.

[10] "Sci2 Tool." [Online]. Available: <https://sci2.cns.iu.edu/user/index.php>. [Accessed: 30-Jun-2020].

[11] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010, doi: 10.1007/s11192-009-0146-3.

[12] N. J. van Eck and L. Waltman, "CitNetExplorer: A new software tool for analyzing and visualizing citation networks," *J. Informetr.*, vol. 8, no. 4, pp. 802–823, Oct. 2014, doi: 10.1016/j.joi.2014.07.006.

[13] and F. H. M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, "SciMAT," *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. July, pp. 1852–1863, 2012, doi: 10.1002/asi.

[14] M. B. Synnestvedt, C. Chen, and J. H. Holmes, "CiteSpace II: visualization and knowledge discovery in bibliographic databases," *AMIA Annu. Symp. Proc.*, pp. 724–728, 2005.

[15] Y. S. Ho, "A bibliometric analysis of highly cited articles in materials science," *Curr. Sci.*, vol. 107, no. 9, pp. 1565–1572, 2014.

[16] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020, doi: 10.1016/S0140-6736(20)30183-5.

[17] W. Guan *et al.*, "Clinical Characteristics of Coronavirus Disease 2019 in China," *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1708–1720, 2020, doi: 10.1056/nejmoa2002032.

[18] F. Zhou *et al.*, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054–1062, 2020, doi: 10.1016/S0140-6736(20)30566-3.

[19] N. Zhu *et al.*, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," *N. Engl. J. Med.*, vol. 382, no. 8, pp. 727–733, 2020, doi: 10.1056/nejmoa2001017.

[20] D. Wang *et al.*, "Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 11, pp. 1061–1069, 2020, doi: 10.1001/jama.2020.1585.

[21] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *Lancet*, vol. 395, no. 10223, pp. 507–513, 2020, doi: 10.1016/S0140-6736(20)30211-7.

[22] P. Zhou *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020, doi: 10.1038/s41586-020-2012-7.

[23] M. Hoffmann *et al.*, "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor," *Cell*, vol. 181, no. 2, pp. 271–280.e8, 2020, doi: 10.1016/j.cell.2020.02.052.

[24] Q. Li *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020, doi: 10.1056/nejmoa2001316.

[25] Z. Wu and J. M. McGoogan, "Characteristics of and Important Lessons from the Coronavirus Disease 2019

- (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention,” *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 13, pp. 1239–1242, 2020, doi: 10.1001/jama.2020.2648.
- [26] E. Garfield, “Citation frequency as a measure of research activity and performance,” *Essays an Inf. Sci.*, vol. 1, no. 1, pp. 406–408, 1973.
- [27] S. Galiani and R. H. GGlvez, “The Life Cycle of Scholarly Articles Across Fields of Research,” *SSRN Electron. J.*, 2017, doi: 10.2139/ssrn.2964565.
- [28] D. R. Smith, “Citation indexing and highly cited articles in the Australian Veterinary Journal,” *Aust. Vet. J.*, vol. 86, no. 9, pp. 337–339, 2008, doi: 10.1111/j.1751-0813.2008.00330.x.
- [29] H. Z. Fu and Y. S. Ho, “Top cited articles in thermodynamic research,” *J. Eng. Thermophys.*, vol. 24, no. 1, pp. 68–85, 2015, doi: 10.1134/S1810232815010075.