Winter 11-12-2020

# Coronavirus research output during 2001-2020: A Scientometrics Analysis

Lambodara Parabhoi
*Indian Institute of Advanced Study, Rashtrapati Nivas, Shimla*, suresh19871987@gmail.com

Dr. Manoj Kumar Verma
*Associate Professor Department of Library and Information Science and Coordinator, Career Counselling and Placement Cell Mizoram University (a Central University) Aizawl-796004*, manojdlis@mzu.edu.in

# Coronavirus research output during 2001-2020: A Scientometrics Analysis

Lambodara Parabhoi
Professional Assistant
Indian Institute of Advanced Study, Library, Shimla
HP, 171005, IN
Email: suresh19871987@gmail.com;lparabhoi@iias.ac.in


Co-Author
Dr. Manoj Kumar Verma
Associate Professor
Department of Library and Information Science and
Coordinator, Career Counselling and Placement Cell
Mizoram University (a Central University)
Aizawl-796004
Email: manojdlis@mzu.edu.in

**Abstract:**

COVID-19 virus originated from Wuhan city of China in December 2019. The emergence of COVID-19 the whole world and severely affected by The United States, China, Brazil, and India etc.. World Health Organization (WHO) declared it as a pandemic in March 2020. Due to COVID-19, a large number of literature published in early 2020. However, very few studies address the impact of published related to literature Coronavirus. In response to the current study conducted and reviewed 20 years' period from 2001- May 2020. A total of 14439 documents were found in the Scopus database, which was published during the study period i.e. 2001- May 2020. The study found that The United States 9973 contributed the highest number of published literature on Coronavirus followed by China. Overall, the USA, China, Germany, The UK, Canada, South Korea accounted for most of the Coronavirus research activity at the global level. Globally, the University of Hong Kong and the Chinese University of Hong Kong ranked with first and second positions in terms of the number of publications contributed to individual institutes. The large quantity of scholarly documents related to Coronavirus has considerably increased in early 2020.

**Keywords:** Bibliometrics, Coronavirus, COVID-19, Scientometrics, Research Output

## 1. Introduction:

In the recent few months, the whole world is a victim of Coronavirus infection and everyone was interested to know about Coronavirus. There are a large number of literature published in different journals, book chapters, and blogs, etc. and Indexed in different similar databases like Google Scholar, Dimensions, PubMed, Scopus, and Webb of Science(Kousha, Street, Thelwall, & Street, 2020) [7]. Now a day's various scholarly publisher has open the content related to coronavirus such as emerald, Wiley, Sage, Springer, and Elsevier, etc. COVID-19 was born in Wuhan city of China in December 2019. And World Health organization declared as Pandemic in the March 2020 WHO [19]. This virus has been spared out all most all the country and highly affected by the USA, China, Italy, the UK, India, and many more countries [21]. Due to the COVID-19 outbreak, many countries declared lockdown for many months and started research to invent vaccines. Recently, Oxford University has successfully conducted the first trial of the newly invented vaccine on the human body [14]. In meanwhile, various researches has been conducted and published in the form of research papers, clinical trials, newspaper articles, etc.

Scientometrics can be defined as "quantitative study of science, communication in science, and science policy" [4]. In the recent past, various studies have been conducted on Scientometrics aspect. It can be used to evaluate It also aims to understand the behavior of scientific citations as a mean of scholarly communication and map intellectual landscapes of a science [6].

As per the recent study, there were large 21,395 numbers of Coronavirus literatures indexed in the Dimensions database and early twenty. One notable study mentioned that more research contributed to the USA and countries followed China on Coronavirus [13]. Bibliometric is a method to analyze the impact of the research publications by applying statistical method and it is recognized worldwide. There are very few studies that address the impact of published literature coronavirus. In response to the current study conducted and reviewed 20 years' period from 2001- May 2020.

## 2. Related Works

More recent attention has focused on coronavirus and there have been large number literate published relater to COVID-19 [11],[18][15],[8],[3],[20]. Much more attention on Scientific study and coronavirus, antibody, prevalence and However, very few study published in social sciences and computer science [13],[18], [17], [7]. For instance, Shri Ram [13] made a study of over 50 years of coronavirus literature indexed in Scopus. It has been reported that the United States contributed a higher 32% proportion research into coronavirus while the University of Hong Kong contributed the highest number of publications. The further study noted that the majority of research work was published in the Journal of Virology. To address

research literature published on different databases coverage and Altmetrics score a notable study made by Kousha et al. [7]. Authors reported that a large number of literature indexed in the Dimensions and Google Scholar while renowned databases like Web of Science and Scopus and PubMed failed. Where indexed 21392 publications relate to Coronavirus. When compared with citation Altmetrics. The study found statistically significant p=0.001 Spearman correlations between altmertics and citations. Thelwall & Thelwall [18] analyzed retweets in the English language related to COVID-19 on March 10-29, 2020. the study revealed that 87 succeeding's tweeted has found 14 million retweets and these tweets were related to lockdown life; safety messages, attitude towards social restrictions; politics. To address the gender difference in tweets related to COVID-19 Thelwall & Thelwall [18] reviewed 3,038,026 English tweets. The study result showed that females were more health consciousness tweets related to family, social distancing, and healthcare while male tweets were related to the sports, the spread of Coronavirus worldwide, and political responses. Rajkumar [12] study literature related to COVID-19 and Metal health. The study results suggest that symptoms of anxiety and depression (16–28%) and self-reported stress (8%) are common psychological reactions to the COVID-19 pandemic. Xu et al., [22] examined real-time cases related to COVID-19 using various sources like Government source, Official social media accounts, and news websites, etc during December 1, 2019, to February 5, 2020. The study confirmed that females confirm the case was lower than males globally. As this was the beginning of Coronavirus, China affected severely as compared to other countries. Khatri et al., [5] reviewed YouTube videos related to COVID-19 in both English and Mandarin languages. Further authors noted that 72 YouTube videos in English and 42 in Mandarin analyzed. It has been reported that a greater portion of English YouTube videos was useful information that Mandarin YouTube videos. Similar kinds of research made [8] on YouTube videos analyzed content. The study reviewed 75 from each groups 'coronavirus' and 'COVID-19 '. A further study reported that over 62 million views worldwide while the videos created by government and professional had information similarly 23%–26% of videos found mis-information related to COVID-19. There is no significant study found in the 2020 literature published in early 2020 on Coronavirus. In response, the current study has conducted.

### 3. Objectives of the study:

The objectives of the present study are:

1. To analyze the type and growth of the Coronavirus literature
2. To find out the prominent countries, institutions, authors, and sources involved in coronavirus research publications
3. To measure the annual citations growth of coronavirus literature
4. To identify the keywords used in research papers and growth in the frequency in use of these keywords

4. **Methodology**:

The present study is an exploration of trends of publications on Coronavirus research based on Scientometrics tools. The research data was collected from Scopus database which is the largest abstract and citation database from various disciplines with smart tools and tracks to analyze and visualize the research and published from Elsevier. The research data was extracted from Scopus database on 02nd May 2020 by using the following search terms as used to collect the data: (TITLE-ABS-KEY ("COVID-19 OR Coronavirus") AND (LIMIT-To (PUB YEAR, 2001-2020). A total of 14439 documents were found which was published during the study period i.e. 2001- May 2020. These documents are further scrutinized, tabulated, and analysed by using the Bibliometrix package through Biblioshiny [1].

5. **Data Analysis:**

**5.1 Overview of Coronavirus Literature**

Table.1 A-B summarise literature published between 2001 to 02 May 2020. A total of 14439 documents in 2545 different sources include a journal, books, conference, etc. It was found that more than one-third of the publications journal articles 10006. During the study period, 44616 keywords (authors and system generate keyword). A total 14439 documents received 24.14 average citations per document and 38113 authors found while authors of single-authored documents 1817, Documents per author was 0.379, 2.64 authors per document, 6.14 Co-Authors per documents and 2.92 collaboration Index was found in total 14439 documents.

| Description | Results |
|---|---|
| Documents | 14439 |
| Sources (Journals, Books, etc.) | 2545 |
| Keywords Plus (ID) | 31112 |
| Author's Keywords (DE) | 13504 |
| Period | 2001 – 2020 |
| Average citations per documents | 24.14 |
| Authors | 38113 |
| Single-authored documents | 1817 |
| Documents per Author | 0.379 |
| Authors per Document | 2.64 |
| Co-Authors per Documents | 6.14 |
| Collaboration Index | 2.92 |
| Table-1B Type of Documents | |
| Document types | |
| Article | 10006 |
| Article In Press | 2 |

4

| | |
|---|---:|
| Book | 8 |
| Book Chapter | 195 |
| Conference Paper | 255 |
| Data Paper | 1 |
| Editorial | 595 |
| Erratum | 5 |

Table.1 An Overview of Coronavirus Literature

## 5.2 Annual Growth of Coronavirus literature

Figure.1 shows the annual growth of the scientific productivity of Coronavirus literature between 2001 to early May 2020. It was noted that the growth rate was not stable and it fluctuated over time. In early 2020, coronavirus publications (2166) were pick top position. Due to the outbreak of COVID-19. In 2001 and 2002 the growth rate was very low. However, one burst can be seen in 2003 and it increased the publication rate from146 to 885 in 2003 and 948 publications in 2004. Again the growth rate was downward up to 2011 upward from 2012 to 2015 onwards.
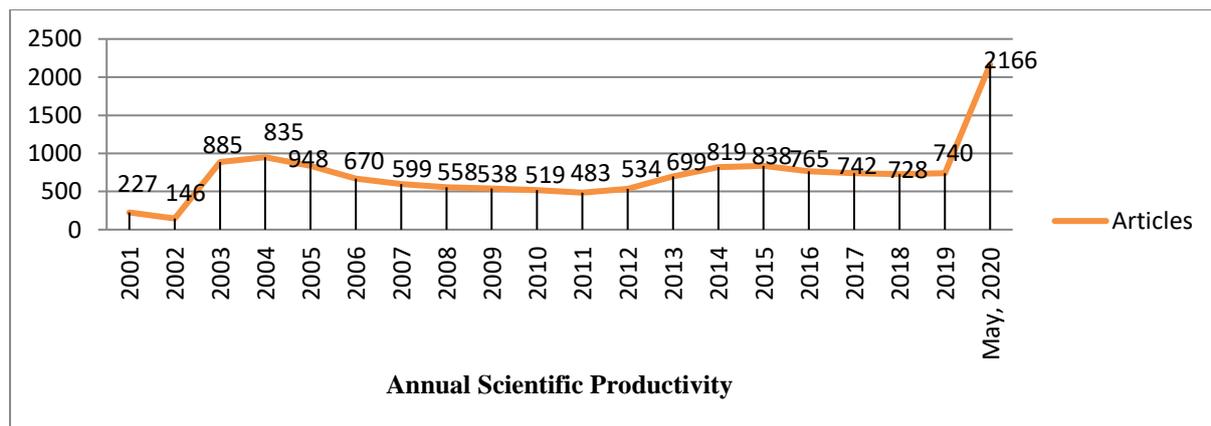


Fig.1 Annual Scientific Productivity of Coronavirus literature between 2001 and May, 2020

## 5.3 Most cited publications related to Coronavirus

Figure 2 shows the most cited documents published between 2001 to early May 2020. The top ten highly cited documents range citation between above 500 to below 2050 citations. It was found that the top-cited document was published in the New England Journal of Medicine by Ksiazek TG in 2003. The closer inspection of the graph depicted that majority of the top ten highly cited documents were published in 2003 while only one documents found in the figure that was published in 2020 in The Lancet journal.
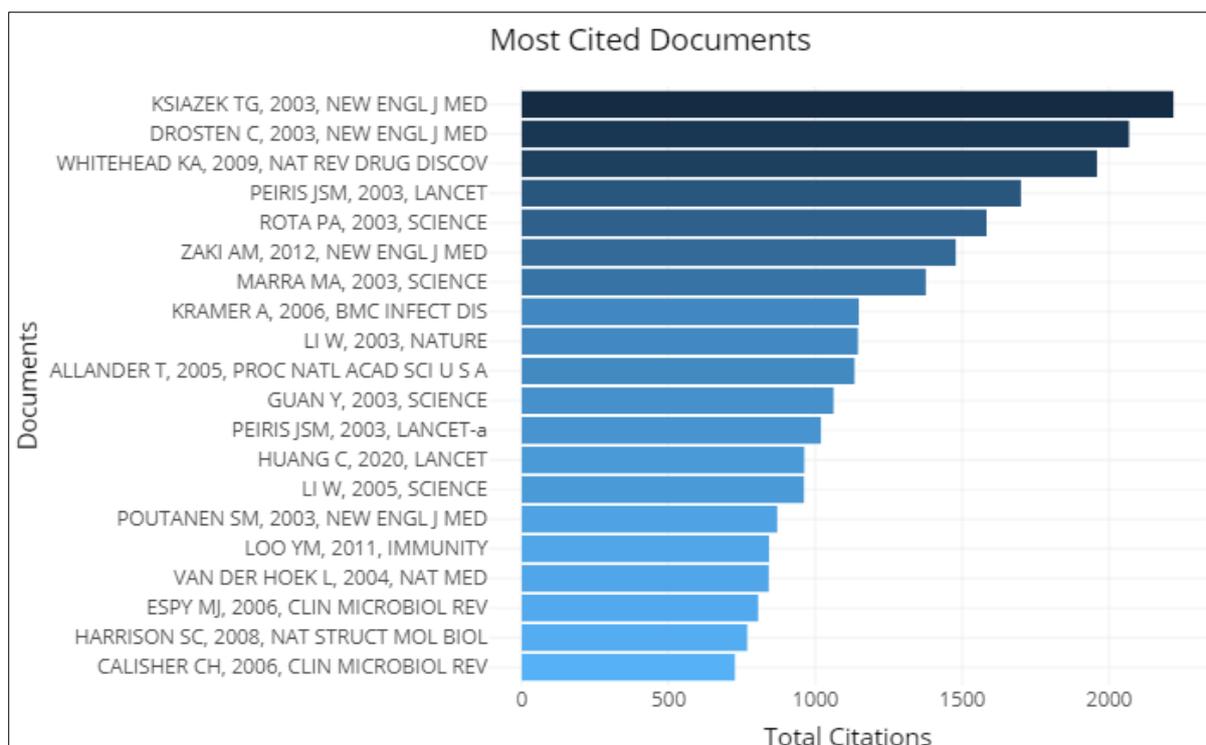
Fig. 2 Most Cited Documents on Coronavirus

## 5.4 Coronavirus Scientific Productivity worldwide

Figure.3 and table.2 shows the productivity of coronavirus literature worldwide. The map was created using the Bibliometrix package and the top twenty contributors calculated. The light blue and deep blue shows the density of the contributor. The deep blue colors represent the top twenty countries with publications while light blue color represents the contributors but not in the top twenty contributors list.  The data revealed that the majority of publications contributed by the USA (9973) followed by China (6851) and Germany (1744).
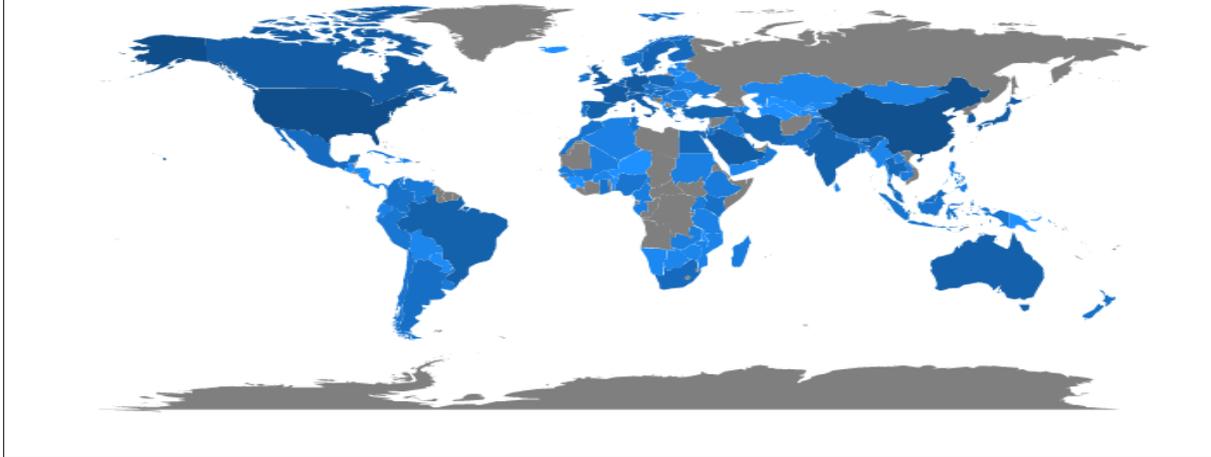
Fig.3 Worldwide Scientific production on Coronavirus

| Country | Freq |
|---|---|
| USA | 9973 |
| CHINA | 6851 |
| GERMANY | 1744 |
| UK | 1714 |
| CANADA | 1453 |
| SOUTH KOREA | 1389 |
| JAPAN | 1385 |
| FRANCE | 1329 |
| TAIWAN | 1327 |
| NETHERLANDS | 1199 |
| ITALY | 1190 |
| SAUDI ARABIA | 989 |
| SINGAPORE | 725 |
| AUSTRALIA | 692 |
| SPAIN | 656 |
| BRAZIL | 628 |
| SWITZERLAND | 593 |
| INDIA | 561 |
| SWEDEN | 343 |
| IRAN | 326 |

Table.2 Worldwide Scientific Productions on Coronavirus

## 5.5 Top Cited Countries

Figure.4 illustrates the top-cited countries globally. It was noted The USA and China got first and second positions while the 3rd position got by the Netherlands. The USA dominated in terms of received citations to the literature on Coronavirus.    By the closer inspection shows more than 10000 citations received while china got the second position in terms of the number of citations and received more than twenty-five thousand citations. While other countries received bellows 25000 citations.
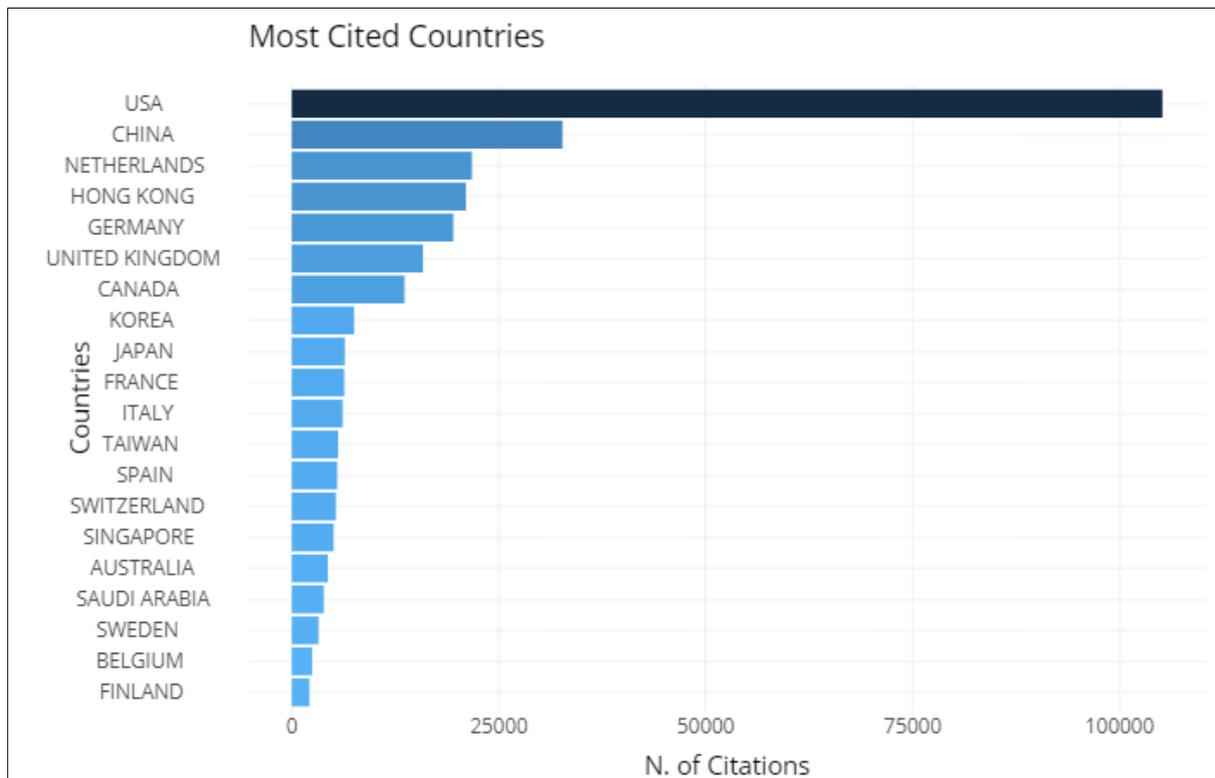


Fig.4 Most Cited Countries

## 5.6 Top contributors by organizations

Table.3 summarize the top 15 contributors by organizations globally. The ranking was done by the highest number of publications contributed by the organizations. It can be seen from the data that slightly lower than one thousand publications contributed by The University of Hong Kong (949) and ranked as first positions it was followed by  Chinese University of Hong Kong (447) and ranked 2nd position while the University of California (337)  ranked 3rd positions globally. Closed inspection of the data revealed that the one and two ranked universities belong to Hong Kong. Data also indicates that most universities belong to The United States.

| Affiliations | Articles | Place | Rank |
|---|---|---|---|
| University of Hong Kong | 949 | Hong Kong | 1 |
| Chinese University Of Hong Kong | 447 | Hong Kong | 2 |
| University of California | 337 | United States | 3 |
| University of Iowa | 278 | United States | 4 |
| Utrecht University | 251 | Netherlands | 5 |
| University of Toronto | 188 | Canada | 6 |
| National Institute of Allergy and Infectious Diseases | 176 | United States | 7 |
| University of Texas Medical Branch | 171 | United States | 8 |
| The University of North Carolina at Chapel Hill | 170 | United States | 9 |
| Leiden University Medical Center | 165 | Netherlands | 10 |
| Vanderbilt University Medical Center | 159 | United States | 11 |
| Fudan University | 152 | China | 12 |
| National Institute of Infectious Diseases | 152 | United States | 13 |
| University of North Carolina | 149 | United States | 14 |
| Centers For Disease Control and Prevention | 148 | United States | 15 |

Table. 3 Top 15 contributors by Organizations

5.7 **Most Cited Sources**

Figure.5 represents the top twenty most cited sources where published on coronavirus content. It is cleary shows that first ranked by the Journal of Virology with more than 40,000 citations received followed by the Virology slightly lower than 10,000 citations and New Englan Journal of Medicine with slightly higher than 10,000 citations. Furthermore, the figure depicted that all 20 journals are reputed with high impact journals and ranked by Google Scholar.
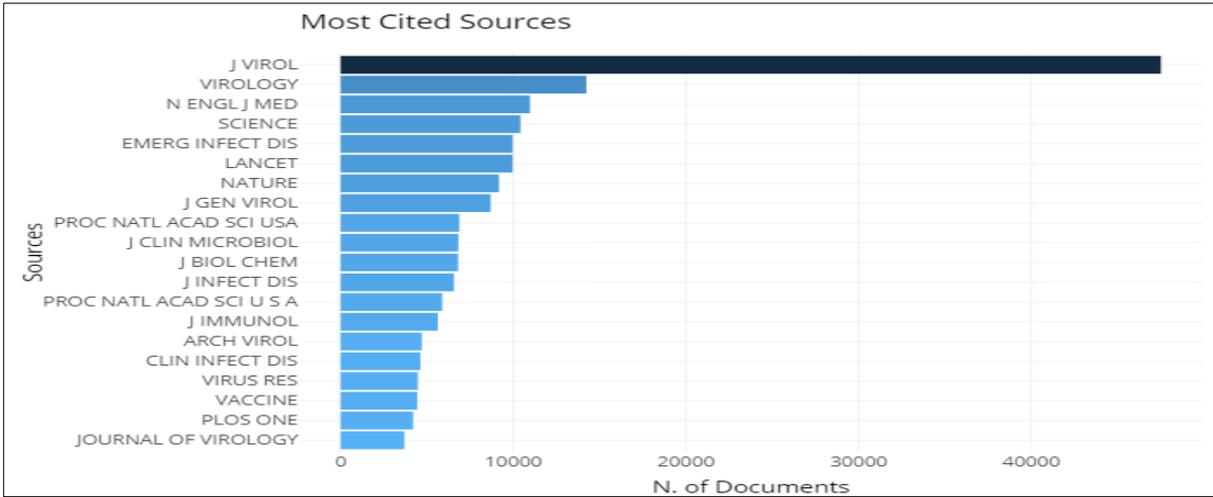
Fig.5 Most Cited Source on Coronavirus Literature

## 5.8 Most Relevant Sources

The topmost favorite source was calculated using the Rbibliometrix package. Only the top twenty sources were calculated. From the figure.6, it was noted that the majority of publications published in the Journal of Virology. It was slightly over 800 documents while second favorite sources ranked by Emerging Infectious Diseases and published over 200 documents and thrid position ranked to Virology journal with slightly lower than Emerging Infectious Diseases. The third-ranked journal was Virology which was slightly higher than 200. Futher figure informs that all the journals were reputed and published in the areas of Virus and diseases.
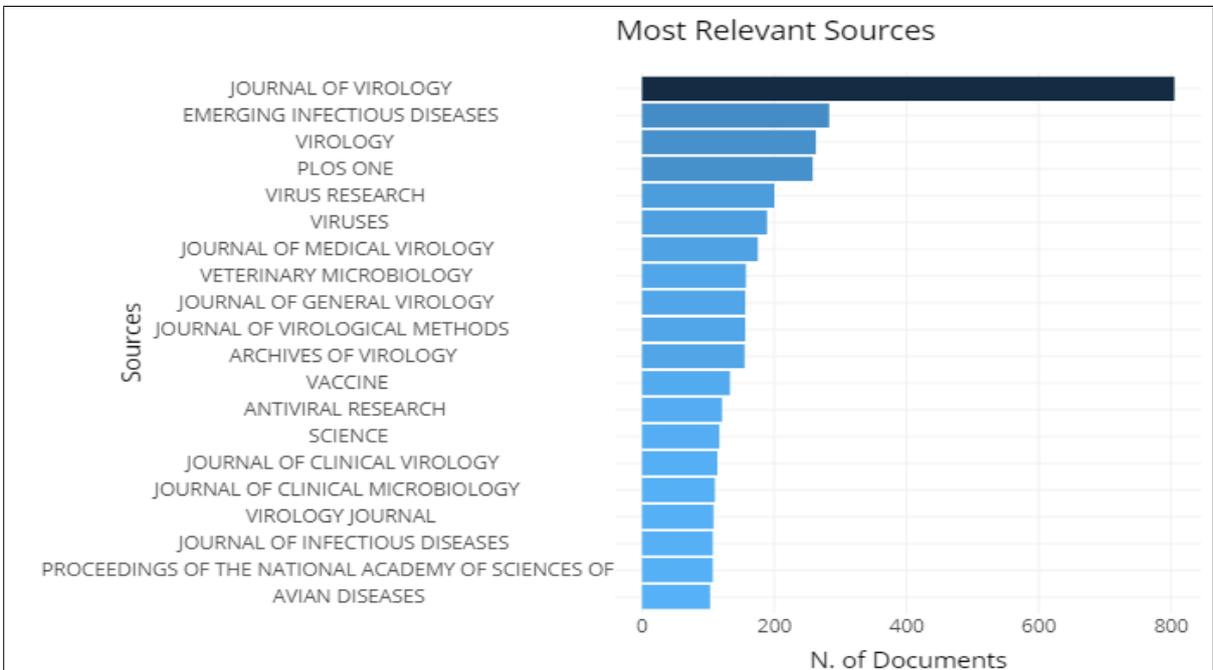


Fig.6 Most Relevant Sources

## 5.9 Word Cloud:

Word cloud of Coronavirus was created using a bibliometrix package using biblioshiny tool of most frequent words used in coronavirus literature. It was shown from the below figure.7 that COVID-19 word more frequently used than other keywords like Sars cov, Sars, Sars Cov 2.Svere acute respiratory syndrome.



Fig.7 Word Cloud on Coronavirus Literature

## 5.10 Growth of Word on Coronavirus Literature

Figure.8 shows the trend of the top ten frequently used words in Coronavirus publications between 2001 and early May 2020. It showed the words COVID-19 , Coronavirus and Sars Cov-2 found at the top of the graph. While other words moderate growth rate found. Moreover, the graph shows that, Coronavirus words upwards from 2001 to 2005 and again downward from 2007 onwards. Since 2016 onwards there has been considerably increased of Coronavirus, COVID-19 , and Sars Cov-2 words.



Fig.8 Word growth on Coronavirus Literature

11

## 5.11 Scientific Collaboration Network on Coronavirus

Figure .9 illustrates the collaborations network of the country. Color represents the cluster while nodes represent contributions an edge represents the relation between the nodes. It clearly shows that there were strong relations on coronavirus literature between the USA and China, the United Kingdom, Canada, and Hong Kong. Since more funding is availed in these countries; similarly, their countries are directly affected by COVID-19 . It is also funded that, these countries have been working on various projects in collaboration with countries.
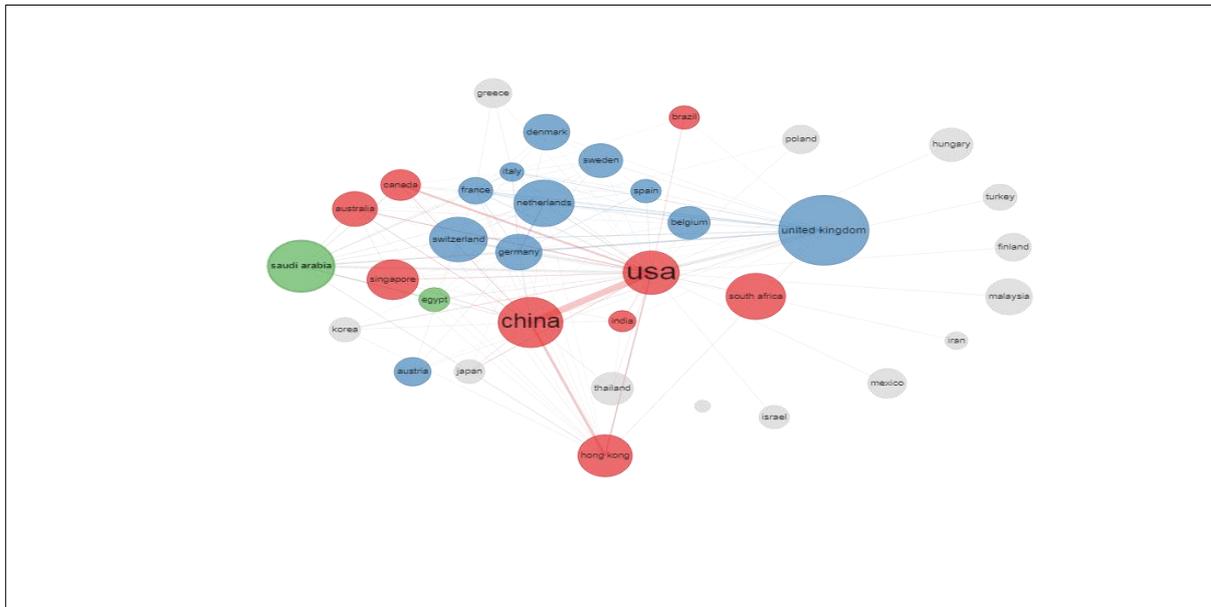


Fig.9 Scientific Collaboration Network on Coronavirus

## 6. Discussion and Conclusion

This study presents an inclusive Scientometrics review of the publications related to Coronavirus. A total of 14439 publications reviewed from multiple perspectives: Characteristics of literature, prominent authors, journals, country, and highly cited publications, favorite's keywords, and growth of keywords, etc. In the recent few months' large numbers of publications were found on Coronavirus due to pandemic COVID-19 . While The United States was the leading country in terms of publications on Coronavirus followed by China. Although COVID-19  and SARS virus origin from China (WHO, 2020) however China contributions rate was less than the USA on Coronavirus literature.  Overall, the USA, China, Germany, The UK, Canada, South Korea accounted for most of the Coronavirus research activity at the global level. A possible explanation for these results may be severely affected by Coronavirus these countries. Another possible explanation may a large number of research and funding institutes found in these countries. As far as the type of publications most of the publications journal articles and reviews articles on Coronavirus rather than conference publications, letters, notes, and book chapters, etc. Regarding the growth of Coronavirus literature, the growth rate of publications was not stable over the period it fluctuated. While the highest growth rate was found in early 2020. This may happen

due to the pandemic COVID-19 outbreak in December 2019. The word cloud shows that COVID-19 , SARS, MERS CoV dominated over other keywords and similar cases found on the growth of the top ten Coronavirus words. Coronavirus and COVID-19, SARS Cov-2 words found impact in recent years. While COVID-19 found in top rank and highest growth found in early 2020. Hence, the COVID-19 outbreak in December 2019. When concerning with the top 15 institutes at the global level. The result shows that the University of Hong Kong and the Chinese University of Hong Kong ranked with first and second positions. It was also found Chinese and the USA institutes dominated over other institutes worldwide on Coronavirus work. All the documents included in the present analysis were published in 2545 different sources. When compared with strong collaborations the countries the USA, China, and the UK strong collocations. This may have happened due to the pandemic COVID-19 and SARS virus outbreak in China. More funding and collaboration with the USA and the UK

**References:**

1. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

2. Bornmann, L., &Haunschild, R. (2018). Do altmetrics correlate with the quality of papers ? A large-scale empirical study based on F1000Prime data. *PLoS Biology*, *13*(5), 1–12. https://doi.org/. https://doi.org/ 10.1371/journal.pone.0197133

3. Chih Cheng, L., Tzu Ping, S., Wen Chien, K., Hung Jen, T., & Po Ren, H. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19 ): The epidemic and the challenges. *International Journal of Antimicrobial Agents*, *55*(January), 1–9. Retrieved from https://pubmed.ncbi.nlm.nih.gov/32081636/

4. Hess, D. J. (1997). *Science studies: An advanced introduction*. New York University Press.

5. Khatri, P., Singh, S. R., Belani, N. K., Yeong, Y. L., Lohan, R., Lim, Y. W., &Teo, W. Z. (2020). YouTube as a source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content. *Travel Medicine and Infectious Disease*, (February), 101636. https://doi.org/10.1016/j.tmaid.2020.101636

6. Kim, M. C., & Zhu, Y. (2018). Scientometrics of Scientometrics: Mapping Historical Footprint and Emerging Technologies in Scientometrics. In M. Jibu & Y. Osabe (Eds.), *Scientometrics*. IntechOpen. https://doi.org/10.5772/intechopen.77951

7. Kousha, K., Street, W., Thelwall, M., & Street, W. (2020). COVID-19 publications : Database coverage , citations , readers , tweets , news , Facebook walls , Reddit posts, 1–24.

8. Li, H., Liu, S.-M., Yu, X.-H., Tang, S.-L., & Tang, C.-K. (2020). Coronavirus disease 2019 (COVID-19 ): current status and future perspectives. *International Journal of Antimicrobial Agents*, (January).

9. Li, M., Yan, S., Yang, D., Li, B., & Cui, W. (2019). YouTube™ as a source of information on food poisoning. *BMC Public Health*, *19*(1). https://doi.org/10.1186/s12889-019-7297-9

10. Moon, J. Y., Yun, E. J., Yoon, D. Y., Seo, Y. L., Cho, Y. K., Lim, K. J., & Hong, J. H. (2020). Analysis of the altmetric top 100 articles with the highest altmetric attention scores in medical imaging journals. *Japanese Journal of Radiology*, (0123456789), 1–6. https://doi.org/10.1007/s11604-020-00946-0

11. Perricone, C., Triggianese, P., Bartoloni, E., Cafaro, G., Bonifacio, A. F., Bursi, R., …Gerli, R. (2020). The anti-viral facet of anti-rheumatic drugs: Lessons from COVID-19 . *Journal of Autoimmunity Journal*, (January). https://doi.org/10.1016/j.jaut.2020.102468

12. Rajkumar, R. P. (2020). COVID-19  and mental health: A review of the existing literature. *Asian Journal of Psychiatry*, *52*(January), 1–5. https://doi.org/10.1016/j.ajp.2020.102066

13. Shri Ram. (2020). Coronavirus Research Trends: A 50–Year Bibliometric Assessment. *Science and Technology Libraries*, *00*(00), 1–17. https://doi.org/10.1080/0194262X.2020.1742270

14. Smout, A. (2020). COVID-19 : Oxford Vaccine Successful in Early Human Trials. *The Wire*. Retrieved from https://thewire.in/health/COVID-19 -oxford-vaccine-successful-in-early-human-trials

15. Tan, W., &Aboulhosn, J. (2020). The cardiovascular burden of coronavirus disease 2019 (COVID-19 ) with a focus on congenital heart disease. *International Journal of Cardiology*, *309*(January), 70–77.

16. Thelwall, M. (2018). Altmetric Prevalence in the Social Sciences, Arts and Humanities: Where are the Online Discussions? *Journal of Altmetrics*, *1*(1), 1–12. https://doi.org/10.29024/joa.6

17. Thelwall, M., & Levitt, J. M. (2020). Retweeting COVID-19  disability issues: Risks, support and outrage. *El Profesional de La Información*, *29*(2), 1–6. https://doi.org/10.3145/epi.2020.mar.16

18. Thelwall, M., &Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics*, *12*(1), 237–248. Retrieved from http://10.0.3.248/j.joi.2018.01.008%0Ahttp://search.ebscohost.com/login.aspx?direct=true&db=llf&AN=128275248&site=ehost-live

19. Valencia, D. N. (2020). Brief Review on COVID-19 : The 2020 Pandemic Caused by SARS-CoV-2. *Cureus*, *12*(3), 1–13. https://doi.org/10.7759/cureus.7386

20. Wilcox, M. A., Khan, N. R., McAbee, J. H., Boop, F. A., & Jr, P. K. (2013). Highly cited publications in pediatric neurosurgery. *Childs NervSyst*, *29*, 2201–2213. Retrieved from https://link.springer.com/article/10.1007%2Fs00381-013-2228-z

21. World Health Organization. (2020). *Coronavirus disease (COVID-19 ) pandemic*. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

22. Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., … Kraemer, M. U. G. (2020). Epidemiological data from the COVID-19  outbreak, real-time case information. *Scientific Data*, *7*(1), 1–6. https://doi.org/10.1038/s41597-020-0448-0