

## Article

# A Bibliometric Analysis of COVID-19 across Science and Social Science Research Landscape

Aleksander Aristovnik , Dejan Ravšelj  and Lan Umek 

Faculty of Public Administration, University of Ljubljana, 1000 Ljubljana, Slovenia;  
dejan.ravselj@fu.uni-lj.si (D.R.); lan.umek@fu.uni-lj.si (L.U.)

\* Correspondence: aleksander.aristovnik@fu.uni-lj.si

Received: 15 September 2020; Accepted: 30 October 2020; Published: 3 November 2020



**Abstract:** The lack of knowledge about the COVID-19 pandemic has encouraged extensive research in the academic sphere, reflected in the exponentially growing scientific literature. While the state of COVID-19 research reveals it is currently in an early stage of developing knowledge, a comprehensive and in-depth overview is still missing. Accordingly, the paper's main aim is to provide an extensive bibliometric analysis of COVID-19 research across the science and social science research landscape, using innovative bibliometric approaches (e.g., Venn diagram, Biblioshiny descriptive statistics, VOSviewer co-occurrence network analysis, Jaccard distance cluster analysis, text mining based on binary logistic regression). The bibliometric analysis considers the Scopus database, including all relevant information on COVID-19 related publications ( $n = 16,866$ ) available in the first half of 2020. The empirical results indicate the domination of health sciences in terms of number of relevant publications and total citations, while physical sciences and social sciences and humanities lag behind significantly. Nevertheless, there is an evidence of COVID-19 research collaboration within and between different subject area classifications with a gradual increase in importance of non-health scientific disciplines. The findings emphasize the great need for a comprehensive and in-depth approach that considers various scientific disciplines in COVID-19 research so as to benefit not only the scientific community but evidence-based policymaking as part of efforts to properly respond to the COVID-19 pandemic.

**Keywords:** COVID-19; coronavirus; pandemic; science; social science; bibliometric analysis; Jaccard distance; text mining

## 1. Introduction

The world has seen two large-scale outbreaks of disease since the 2000s. Respectively emerging in 2003 and 2012, these are Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), which posed a threat around the world and claimed thousands of lives [1]. In December 2019, a new strain of coronavirus (SARS-CoV-2), not previously identified in humans, emerged in Wuhan City, in the Hubei province of China. The virus soon spread across countries with the number of cases and deaths related to COVID-19 quickly exceeding the numbers of the two other coronaviruses (SARS-CoV-1 and MERS-CoV). This rapid spread of COVID-19 around the world led to the World Health Organization (WHO) to declare it a pandemic on 11 March 2020 [2]. The COVID-19 pandemic is a typical public health emergency. Its high infection rate means it is a huge threat to global public health [3–5]. However, its rapid proliferation has not only affected the lives of many people on the planet, but disrupted patterns of social and economic development, bringing incalculable social and economic losses [6]. Within just a six months of the outset of the COVID-19 pandemic (by 1 July 2020), some 10.3 million cases and 0.5 million deaths were registered at the global level [7]. International institutions have therefore announced the global economy is now in recession as bad or

worse than during the global financial crisis of 2009, stating the recession will affect both developed and developing countries [8,9]. It is thus no surprise that the COVID-19 pandemic has attracted the academic sphere's attention and spurred a new wave of research in this area.

Recent bibliometric studies considering broader aspects of coronavirus research over time stress that pandemics are a major medical issue and present some interesting findings. Taking previous coronavirus pandemics into account, Hu et al. [10] establish that the greatest research interest is seen in the first year after the outburst. This is confirmed by research addressing coronavirus research trends over the last 20-year [11,12] and 50-year periods [13,14]. Yet, although the growth pattern has not been uniform, China and the USA have been major contributors to coronavirus research [15]. This makes it unsurprising that COVID-19 has become a central topic of the recent scientific literature, given that research addressing various aspects of COVID-19 could hold the key to mitigating the COVID-19 pandemic and its consequences [16,17]. The fast-growing interest in COVID-19 and its coronaviruses seen today even led to the creation of the COVID-19 Open Research Dataset (CORD-19), an emerging resource offering scientific papers on COVID-19 and related historical coronavirus research and giving a solid basis for generating new insights in support of the ongoing fight against COVID-19 [18]. An overview of COVID-19 publications reveals they are chiefly focused on a few well-defined areas, including coronaviruses (primarily SARS-CoV-1, MERS-CoV, SARS-CoV-2), public health and viral epidemics, the molecular biology of viruses, influenza and other families of viruses, immunology and antivirals as well as methodology (testing, diagnosing, clinical trials). However, a review of the latest COVID-19 publications from 2020 indicates a shift has been underway from health to other relevant scientific disciplines [19].

In the literature one also finds several recent bibliographic studies that focus solely on COVID-19 research. Never before in the history of academic publishing has such a great volume of research concentrated on a single topic been produced [20]. Yet, the rush for scientific evidence on the novel COVID-19 might inadvertently encourage dubious publications, which make it into the scientific domain because the authors had some special relationship with the journals in which they appeared [21]. Nevertheless, a recent study on scientific globalism during the pandemic shows that scientific globalism manifests differently when COVID-19 publications are compared with non-COVID-19 publications. Interestingly, while the COVID-19 pandemic is obviously a worldwide concern and countries have indeed increased their extent of international scientific collaboration during the pandemic, not all countries have sought to engage more globally. The study reveals that countries more affected by the pandemic and those with relatively smaller GDP have tended to participate more strongly in scientific globalism than their counterparts [22].

The bibliometric approach examining COVID-19 related issues has thus far been applied to different areas. Namely, some authors aim for a general overview of COVID-19 research (see Sa'ed and Al-Jabi [23]), while others consider a comparative approach, for example a comparison of COVID-19 research between English and Chinese studies (see Fan et al. [24]) or of the gender distribution of authors of medical papers on the COVID-19 pandemic (see Andersen et al. [25]). Moreover, some bibliometric studies only consider a single country in their analysis (for the case of India, see Vasantha Raju and Patil [26]) and others on top-cited COVID-19 publications (see ElHawary et al. [27]). Finally, a few COVID-19 bibliometric studies provide in-depth analysis in the fields of traditional Chinese medicine (see Yang et al. [28]), economics (see Mahi et al. [29]) and business and management (see Verma and Gustafsson [30]).

Nevertheless, the vast majority of COVID-19 bibliometric studies reveal that China and the USA are producing the largest amount of COVID-19 associated scientific published work [31–35]. The most relevant institutions for COVID-19 research are the Huazhong University of Science and Technology, Wuhan University, and the University of Hong Kong. Moreover, most published documents on COVID-19 are found in prestigious journals with high impact factors, including *The Lancet*, *BMJ—Clinical Research Ed.*, and *Journal of Medical Virology* [31,35]. Further, by the number of publications, the most influential authors in COVID-19 research are Huang, C., Zhu, N. and

Chan, J.F. [33]. Finally, it is also established that virology, epidemiology, clinical features, laboratory examination, radiography, diagnosis and treatment are current hotspots in COVID-19 research [33,34].

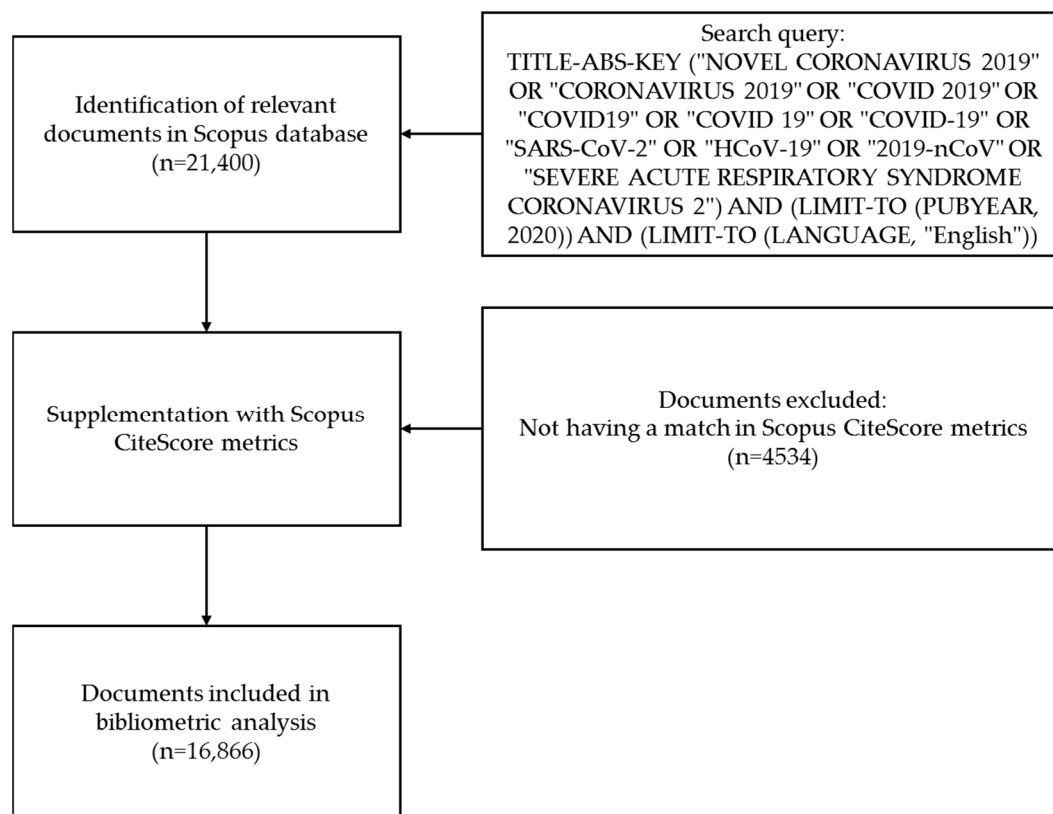
Still, while the absence of knowledge about the novel COVID-19 has grabbed the academic sphere's attention, triggering a new wave of research into the virus SARS-CoV-2 [36], the vast majority of recent studies chiefly consider health-related issues, leaving other aspects aside, as shown by the latest literature [31–35]. Moreover, COVID-19 research's present status is shown to be only the early development of knowledge. Therefore, the literature stresses that more research should be conducted in less-explored areas, including the life, physical and social sciences and the humanities [33]. Accordingly, this paper's attention is to provide extensive bibliometric analysis on COVID-19 research in the first half of 2020. Although several papers addressing the bibliometric analysis of COVID-19 research exist, a number of research gaps are identified and carefully tackled in this paper. First, existing bibliometric studies predominantly focus on a general analysis of COVID-19 research, showing the importance of the health sciences in this area, whereas a detailed insight encompassing different research landscapes remains missing. This paper thus presents in-depth bibliometric analysis by considering various science and social science research landscapes or subject areas, including the corresponding subject-area classifications and research fields. Second, the lion's share of COVID-19 bibliometric studies mostly considers databases that only contain document information. Accordingly, this paper extends the analysis by examining a comprehensive database that includes document and source information with a view to allowing bibliometric analysis of different research landscapes. Finally, recent COVID-19 research overlooks the overlap among scientific disciplines and lacks innovative bibliometric approaches. As a result, in addition to well-established approaches, this paper utilizes a wide range of innovative bibliometric approaches, including descriptive analysis, network analysis, cluster analysis based on the Jaccard distance, and text mining based on binary logistic regression. These also allow all possible logical relationships among different scientific disciplines to be shown.

Thus, the primary aim of this paper is to provide an unprecedented, comprehensive and in-depth examination of COVID-19 research across different research landscapes, which might suggest important guidelines for researchers with respect to avenues for future research. The remaining sections of this paper are structured as follows. The second section presents the materials and methods. The results are discussed in the third section. The paper ends with a conclusion in which the main findings are summarized.

## 2. Materials and Methods

Comprehensive bibliometric data on COVID-19-related research were obtained in two consecutive phases, as presented in Figure 1. The first phase involved identifying all relevant documents or publications from 1 January 2020 to 1 July 2020 in the Scopus database on document information, a database also widely recognized in previous research [10,14,31,35]. The applied search query extended previous narrowly-defined queries [33,34] by including a broad range of COVID-19 related keywords: "novel coronavirus 2019", "coronavirus 2019", "COVID 2019", "COVID19", "COVID 19", "COVID-19", "SARS-CoV-2", "HCoV-19", "2019-nCoV" and "severe acute respiratory syndrome coronavirus 2". The keyword search was set to include titles, abstract and keywords. In addition, the search period was limited to include documents published between 1 January 2020 and 1 July 2020. Finally, only documents in the English language were considered for the review process. According to the presented search query, a total of 21,400 documents was identified as relevant in COVID-19 research. Interestingly, the number of documents obtained by using an identical search query had increased by 58.8% since on 1 June 2020 the same search produced 13,480 documents. This implies that interest in COVID-19 research is growing exponentially. Due to the Scopus export limitation up to 20,000 records at a time, the unique academic work identifier assigned in Scopus bibliographic database (EID) was utilized to obtain basic citation metadata for 21,400 documents (author(s); document title; year; source title; volume, issue, pages; citation count; source and document type). Moreover, due to the additional Scopus export limitation up to 2000 records at a time on detailed document information (citation

information, bibliographical information, abstract and keywords, funding details, other information), the EID was also used to split the found relevant documents into smaller blocks of data. The data were exported in comma-separated values (csv) format. Finally, the mentioned blocks of data were merged to create a full dataset containing 21,400 documents.



**Figure 1.** Flowchart of database determination (January–June 2020).

The second phase involved supplementing the presented Scopus database on document information with Scopus CiteScore metrics exported in csv format from Scopus Sources page that contain source-related information (e.g., citations, rankings, source-normalized impact per paper (SNIP) etc.). These two data sets were merged by using the International Standard Serial Number (ISSN). The merging process revealed that some documents from Scopus had no match in Scopus CiteScore metrics ( $n = 4534$ ), meaning they were not considered in the bibliometric analysis. The biggest proportion of these documents were articles (61.1%), followed by letters (11.5%), reviews (10.0%), notes (8.3%), editorials (7.6) and other (1.5%). The screening process thus resulted in a database of 16,866 documents. The data preparation process, i.e., obtaining, merging and cleaning the relevant data, was facilitated by the Python programming language using the Pandas and Numpy libraries [37]. Python code used in the analysis is available and documented at GitHub repository: <https://github.com/covid-bib/bibliometric>.

An in-depth bibliometric analysis then followed, allowing for an innovative literature review approach and significantly upgrading traditional literature review techniques. Namely, a structured literature review is a traditional approach to analyzing and reviewing scientific literature, providing an in-depth overview of the content. However, this approach suffers from several limitations associated with subjective factors, time-consumption and efficiency. The application of modern bibliometric approaches reduces these limitations and entails an effective way of handling extensive collections of scientific literature [38]. Thus far, bibliometric studies on COVID-19 research applied well-established bibliometric approaches by utilizing VOSviewer (see Hamidah et al. [32]), SciMAT (see Herrera-Viedma et al. [13]) and basics of machine learning (see De Felice and Polimeni [39]).

Still, bibliometric studies mostly overlook the fact that scientific disciplines overlap strongly, resulting in these studies making similar findings and conclusions and producing a lack of knowledge in less-explored areas [33]. Therefore, in order to supplement existing research and assess the state of current COVID-19 research across different research landscapes (health sciences, life sciences, physical sciences and social sciences and humanities), innovative bibliometric approaches are relied on in this paper. The bibliometric analysis was performed by considering the Scopus hierarchical classification of documents based on the All Science Journal Classification scheme (ASJC) and in-house experts' opinions. Accordingly, the documents were classified in three hierarchically arranged groups: (1) subject area categories; (2) subject area classifications; and (3) fields.

On this basis, the following bibliometric approaches were applied. First, for descriptive analysis, including a Venn diagram for detecting the overlap of scientific disciplines, the Biblioshiny application [40] and the Python library Pyvenn [41] were used. Second, in order to depict relations among keywords and fields, a co-occurrence network analysis was performed with VOSviewer, a software tool for constructing and visualizing bibliometric networks [42]. Moreover, to examine relationships between different subject area classifications within COVID-19 research a cluster analysis was undertaken based on the Jaccard distance (JD) (Jaccard index subtracted from 1). The Jaccard distance measures dissimilarity between two fields (subject-area classifications). In other words, it counts the number of documents that belong to exactly one field and divides this number by the number of documents that belong to at least one field. In terms of measurement, Jaccard distance ranges from 0 to 1, with 0 suggesting perfect overlap and 1 indicating no overlap [43]. The Jaccard distance is calculated with Python library Scipy [44], while the clustermap is designed using Python's most powerful visualization libraries, i.e., Matplotlib and Seaborn [41,45]. Scopus database classifies its documents into 27 subject area classifications (SAC). Excluding multidisciplinary SAC, order remaining 26 SAC  $C_1, C_2, \dots, C_{26}$  are covered in our data set. We can define the similarity between two SAC  $C_i$  and  $C_j$  using the Jaccard coefficient (see Equation (1))

$$J(C_i, C_j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} \quad \forall i, j = 1, 2, \dots, 26 \quad (1)$$

where  $C_i = \{\text{documents that are classified to SAC } C_i\}$ . The Jaccard coefficient counts the number of documents that belong to both  $C_i$  and  $C_j$  (power of intersection  $|C_i \cap C_j|$ ) and divides this number by the number of documents that are classified to  $C_i$  or  $C_j$  (power of union  $|C_i \cup C_j|$ ). In the paper, the Jaccard coefficient is further used for clustering of SAC. Since clustering algorithms used dissimilarities (instead of similarities)  $J(C_i, C_j)$  is replaced by Jaccard distance  $JD(C_i, C_j)$ , i.e., by subtracting the Jaccard coefficient from 1 (see Equation (2)).

$$JD(C_i, C_j) = 1 - J(C_i, C_j) \quad \forall i, j = 1, 2, \dots, 26 \quad (2)$$

Finally, to predict a document's subject area based on its abstract, a text-mining-based classification was used [46]. For this purpose, binary logistic regression was selected as a prediction model. Accordingly, four different binary logistic models were tested for each individual subject area, with the binary dependent variable having the value of 1 if a document belongs to the individual subject area and 0 if the document belongs to other remaining subject areas. Based on the results of fitting the model to the data, the binary logistic regression also provides information on which words are most characteristic for a particular subject area (which discriminate the most between two subject areas). This approach requires documents to have a full abstract. Text mining was performed with the Natural Language Toolkit (NLTK), a Python package for natural language processing [47]. In the first phase, pre-processing is performed (abstracts are converted to lowercase, accents are removed, word punctuation is used as tokenization). WordNet lemmatization is then applied [48], the set of extracted words is further filtered with a list from nltk.corpus and manually-added stop words [49]. To construct features (bag of words), the "term frequency-inverse document frequency



(tf-idf) method was employed. The class TfidfVectorizer from sklearn.feature\_extraction.text [50] was used with the following parameters: sublinear term frequency (tf) scaling, smooth inverse document frequency (idf) weights, unicode transformation format (utf)-8 encoding, l2 norm regularization, min data frequency = 1, max data frequency = 10. To extract new features for classification, a search for unigrams (single words) and bigrams (sequence of two words) was performed. The top 100 features are created and then used as predictors (independent variables) in a binary logistic model. Binary logistic regression was used to empirically verify if it is possible to predict subject area of a document from its abstract. For every subject area ( $S_1$  = Health Sciences,  $S_2$  = Life Sciences,  $S_3$  = Physical Sciences,  $S_4$  = Social Sciences and Humanities) we define an indicator variable  $Y_i$  which takes values 1 (a document is classified to a subject area  $S_i$ ) and 0 (otherwise). The variables  $Y_1, Y_2, Y_3$  and  $Y_4$  are further treated as separate dependent variables for logistic regression models. For the predictor variables we used  $p = 300$  terms extracted from documents' abstracts. The values of the predictor variables ( $X_1$  = term "acute",  $X_2$  = term "admission",  $X_p = X_{300}$  = term "year") are TF-IDF statistics (top 300 terms were included). The models estimate the conditional probabilities  $P(Y_i = 1 | X_1, X_2, \dots, X_p)$  that a document is classified to subject area  $S_i$ ). The formula of binary logistic regressions used in the paper correspond to:

$$P(Y_i = 1 | X_1, X_2, \dots, X_p) = \frac{\exp(\beta_{0i} + \beta_{1i} \cdot X_1 + \beta_{2i} \cdot X_2 + \dots + \beta_{pi} \cdot X_p)}{1 + \exp(\beta_{0i} + \beta_{1i} \cdot X_1 + \beta_{2i} \cdot X_2 + \dots + \beta_{pi} \cdot X_p)} \text{ for } i = 1, 2, 3, 4 \quad (3)$$

### 3. Results

An overview of the scientific documents utilized in this study is presented in Table 1. A total of 16,866 documents written by 66,504 distinct authors and published in 2548 journals was relied on in this study, where 7422 (44.0%) have at least one citation in the Scopus database, providing a total of 100,683 citations. For these documents, the average citations per document were 13.57 while the average authors per document were 3.94. The biggest proportion of these documents were articles (41.5%) and letters (26.5%). A much smaller proportion of them were reviews (10.2%), editorials (10.1%) and notes (9.4%). Finally, there was a negligible share of other documents (2.4%) such as short surveys, conference papers, errata and data papers. The presented characteristics of these scientific documents on COVID-19 research are largely in line with previous research [32,33].

**Table 1.** Overview of scientific documents on COVID-19 research (January–June 2020).

Database Summary	Findings
Bibliometric Items	Number
Total documents	16,866
Total authors	66,504
Total journals	2548
Total citations	100,683
Cited documents	7422
Average citations	13.57
Average authors	3.94
Document Type	Number (Share)
Article	6998 (41.5%)
Letter	4467 (26.5%)
Review	1713 (10.2%)
Editorial	1698 (10.1%)
Note	1593 (9.4%)
Other	397 (2.4%)

Scopus provides a hierarchical classification of documents by considering the ASJC (All Science Journal Classification scheme) and in-house experts' opinions. The documents are hence classified

in three hierarchically arranged groups: (1) subject area categories; (2) subject-area classifications; and (3) fields. The distribution of documents according to these groups is presented in Table 2. Nearly two-thirds of documents are in the area of the health sciences (65.2%), with medicine (91.0%) being the most exposed, whereby the dominant focus is on infectious diseases (10.2%) and general medicine (9.7%). This is in harmony with earlier bibliometric studies which show that COVID-19 research is the main domain of the health-related sciences [31–35]. A much smaller number of documents is in the area of the life sciences (19.0%). Nevertheless, biochemistry, genetics and molecular biology (35.3%), as well as immunology and microbiology (31.4%) are identified as the most relevant subject-area classifications, while virology (11.6%) and immunology (10.2%) are recognized as the most important research fields within the life sciences. The smallest share of documents is found in the physical sciences (7.5%). These are focused on environmental science (31.4%) and engineering (15.4%), with the research field of pollution (10.7%) being the most exposed. Finally, a relatively small share of documents is found in the area of the social sciences and humanities (8.3%). Still, the social sciences (44.2%) and psychology (24.6%) are recognized as the most relevant subject-area classifications, while sociology and political science (9.2%) is identified as the most important research field within the social sciences and humanities. The aforementioned gives support for the claims of a lack of knowledge in less-explored areas, including the life, physical and social sciences [33]. Therefore, it is no surprise that many calls have been made for more extensive COVID-19 research in less-explored scientific disciplines.

Table 3 presents the most relevant (top 20) journals in COVID-19 research by number of documents. They contain almost one-fifth (17.6%) of total documents and cover a significant share (41.3%) of total citations. Regarding different scientific disciplines or subject areas (classifications), the most relevant journals mainly operate in the area of the health sciences (medicine), covering the following research fields: infectious diseases, general medicine, microbiology (medical), psychiatry and mental health, public health, environmental and occupational health, critical care and intensive care medicine, dermatology, endocrinology, diabetes and metabolism, epidemiology as well as internal medicine. Further, a smaller share of the most relevant journals operate in the area of the life sciences (immunology and microbiology as well as neuroscience), with a focus on biological psychiatry and virology. Some of these journals also publish on the physical sciences (environmental science, mathematics, physics and astronomy), focusing on the following research fields: applied mathematics, environmental chemistry, environmental engineering, general mathematics, general physics and astronomy, health, toxicology and mutagenesis, pollution, statistical and nonlinear physics, and waste management and disposal.

Finally, there is only one journal, which operates in the area of the social sciences (psychology), covering the research field of general psychology. There is also one journal classified as multidisciplinary. Most of these journals rank in the first quartile (Q1) and have a relatively high source-normalized impact per paper (SNIP), which is consistent with the existing research [31,35]. Further, most of these journals are from the UK, the Netherlands, and the USA. Similar findings are also made in previous COVID-19 bibliometric studies [33,34]. However, all of the current bibliometric studies overlook the large overlap that exists among scientific disciplines, leading to biased results and thus a lack of comprehensive understanding of COVID-19 research across different scientific disciplines [33].

**Table 2.** The distribution of COVID-19 related documents according to the Scopus hierarchical classification (January–June 2020).

Subject Area	Subject Area Classification (All)	Fields (Top 10)
Health Sciences (65.2%)	Medicine (91.0%); Nursing (4.9%); Health Professions (2.1%); Dentistry (1.2%); Veterinary (0.8%)	Infectious Diseases (10.2%); General Medicine (9.7%); Public Health, Environmental and Occupational Health (5.3%); Surgery (4.8%); Microbiology (medical) (4.4%); Cardiology and Cardiovascular Medicine (4.2%); Psychiatry and Mental Health (3.7%); Radiology, Nuclear Medicine and Imaging (3.1%); Neurology (clinical) (2.9%); Immunology and Allergy (2.9%)
Life Sciences (19.0%)	Biochemistry, Genetics and Molecular Biology (35.3%); Immunology and Microbiology (31.4%); Neuroscience (15.2%); Pharmacology, Toxicology and Pharmaceuticals (13.0%); Agricultural and Biological Sciences (5.1%)	Virology (11.6%); Immunology (10.2%); General Biochemistry, Genetics and Molecular Biology (5.9%); Pharmacology (5.3%); Cancer Research (4.9%); Neurology (4.6%); Molecular Biology (4.5%); Biochemistry (3.7%); Microbiology (3.6%); Biological Psychiatry (3.6%)
Physical Sciences (7.5%)	Environmental Science (31.4%); Engineering (15.4%); Computer Science (10.5%); Mathematics (9.4%); Chemical Engineering (8.6%); Physics and Astronomy (8.0%); Chemistry (6.9%); Energy (5.1%); Material Science (3.0%); Earth and Planetary Sciences (1.7%)	Pollution (10.7%); Health, Toxicology and Mutagenesis (6.8%); Environmental Engineering (6.1%); Environmental Chemistry (5.9%); Waste Management and Disposal (5.5%); Applied Mathematics (4.6%); General Physics and Astronomy (3.6%); Biomedical Engineering (3.4%); Statistical and Nonlinear Physics (3.0%); General Mathematics (2.9%)
Social Sciences and Humanities (8.3%)	Social Sciences (44.2%); Psychology (24.6%); Business, Management and Accounting (11.4%); Arts and Humanities (9.6%); Economics, Econometrics and Finance (8.8%); Decision Sciences (1.3%)	Sociology and Political Science (9.2%); Clinical Psychology (6.3%); Geography, Planning and Development (6.3%); Health (social science) (5.7%); Social Psychology (5.6%); Education (5.1%); Political Science and International Relations (5.0%); General Psychology (4.9%); Arts and Humanities (miscellaneous) (4.2%); Applied Psychology (3.7%)

Note: The calculations do not consider the overlapping of subject areas, classifications and fields.

**Table 3.** Most relevant journals by number of documents in COVID-19 research (January–June 2020).

Source Title	Number of Documents	Number of Citations	Subject Area (Classification)	Sub-Subject Area/Field (Ranking) 2019	SNIP 2019	Country
<i>Journal of Medical Virology</i>	293	3657	Life Sciences (Immunology and Microbiology) Health Sciences (Medicine)	Virology (37/66, Q3) Infectious Diseases (108/283, Q2)	0.780	USA
<i>The BMJ</i>	261	1358	Health Sciences (Medicine)	General Medicine (21/529, Q1)	3.999	UK
<i>The Lancet</i>	239	13,755	Health Sciences (Medicine)	General Medicine (1/529, Q1)	21.313	UK



Table 3. Cont.

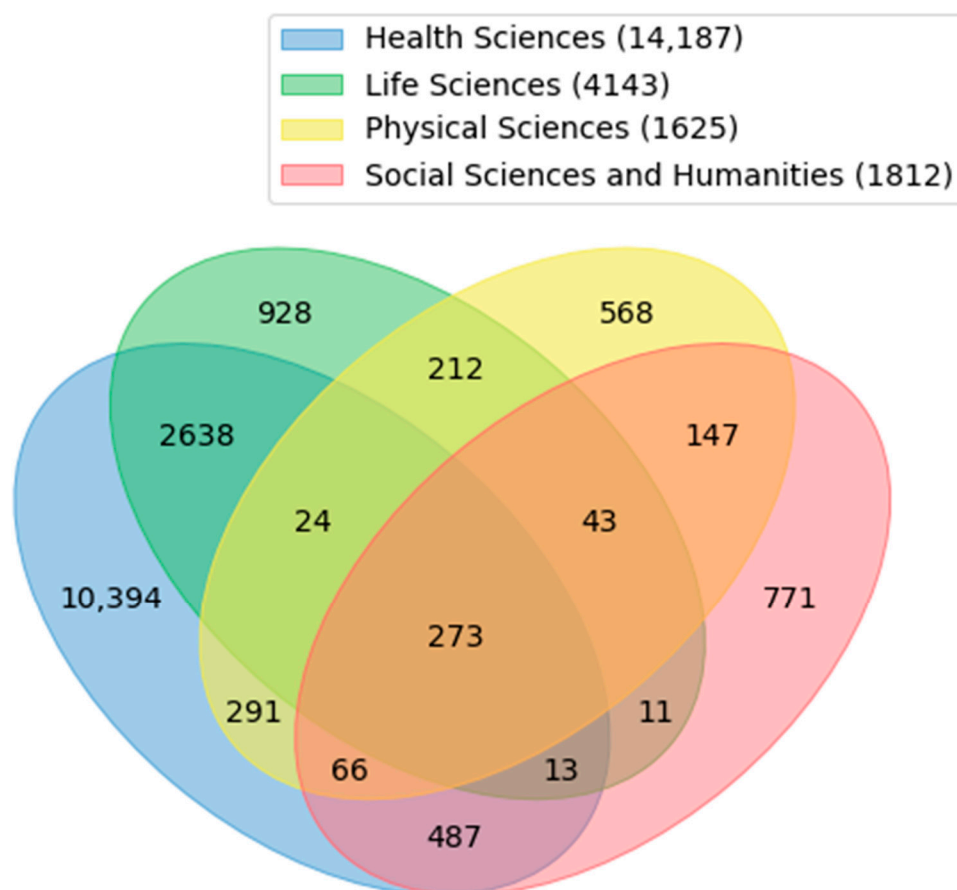
Source Title	Number of Documents	Number of Citations	Subject Area (Classification)	Sub-Subject Area/Field (Ranking) 2019	SNIP 2019	Country
<i>Medical Hypotheses</i>	227	107	Health Sciences (Medicine)	General Medicine (99/529, Q1)	0.509	USA
<i>Science of the Total Environment</i>	174	948	Physical Sciences (Environmental Science)	Environmental Engineering (10/132, Q1) Pollution (13/120, Q1) Waste Management and Disposal (10/100, Q1) Environmental Chemistry (17/115, Q1)	1.977	Netherlands
<i>International Journal of Environmental Research and Public Health</i>	155	490	Health Sciences (Medicine) Physical Sciences (Environmental Science)	Public Health, Environmental and Occupational Health (174/516, Q2) Health, Toxicology and Mutagenesis (68/128, Q3) Pollution (58/120, Q2)	1.248	Switzerland
<i>Journal of Infection</i>	155	1049	Health Sciences (Medicine)	Microbiology (medical) (13/115, Q1) Infectious Diseases (21/238, Q1)	1.587	UK
<i>International Journal of Infectious Diseases</i>	148	1503	Health Sciences (Medicine)	Microbiology (medical) (26/115, Q1) Infectious Diseases (59/283, Q1)	1.426	Netherlands
<i>Psychiatry Research</i>	130	314	Health Sciences (Medicine) Life Sciences (Neuroscience)	Psychiatry and Mental Health (154/506, Q2) Biological Psychiatry (25/38, Q3)	0.968	Ireland
<i>Journal of Clinical Virology</i>	120	239	Life Sciences (Immunology and Microbiology) Health Sciences (Medicine)	Virology (19/66, Q2) Infectious Diseases (44/283, Q1)	1.238	Netherlands
<i>Diabetes and Metabolic Syndrome: Clinical Research and Reviews</i>	119	462	Health Sciences (Medicine)	Internal Medicine (75/128, Q3) Endocrinology, Diabetes and Metabolism (135/217, Q3)	0.982	Netherlands
<i>Infection Control and Hospital Epidemiology</i>	118	172	Health Sciences (Medicine)	Microbiology (medical) (39/115, Q2) Epidemiology (40/93, Q2) Infectious Diseases (91/283, Q2)	1.358	UK
<i>Travel Medicine and Infectious Disease</i>	113	621	Health Sciences (Medicine)	Public Health, Environmental and Occupational Health (73/516, Q1) Infectious Diseases (82/283, Q2)	1.184	Netherlands
<i>Critical Care</i>	112	244	Health Sciences (Medicine)	Critical Care and Intensive Care Medicine (4/81, Q1)	2.508	UK
<i>The Lancet Infectious Diseases</i>	111	2280	Health Sciences (Medicine)	Infectious Diseases (4/283, Q1)	7.234	UK

Table 3. Cont.

Source Title	Number of Documents	Number of Citations	Subject Area (Classification)	Sub-Subject Area/Field (Ranking) 2019	SNIP 2019	Country
<i>New England Journal of Medicine</i>	106	11,768	Health Sciences (Medicine)	General Medicine (2/529, Q1)	13.212	USA
<i>Asian Journal of Psychiatry</i>	101	433	Health Sciences (Medicine) Social Sciences and Humanities (Psychology)	Psychiatry and Mental Health (217/506, Q2) General Psychology (71/204, Q2)	1.022	Netherlands
<i>Dermatologic Therapy</i>	100	153	Health Sciences (Medicine)	Dermatology (74/123, Q3)	0.883	UK
<i>Chaos, Solitons and Fractals</i>	97	132	Physical Sciences (Mathematics) Physical Sciences (Physics and Astronomy)	Applied Mathematics (25/510, Q1) General Mathematics (9/368, Q1) General Physics and Astronomy (27/224, Q1) Statistical and Nonlinear Physics (4/44, Q1)	1.380	UK
<i>Science</i>	97	1918	Multidisciplinary (Multidisciplinary)	Multidisciplinary (2/111, Q1)	7.521	USA

### 3.1. Bibliometric Analysis across Different Subject-Area Categories

According to the Scopus classification, the documents may be classified in four different subject areas: health sciences, life sciences, physical sciences, and social sciences and humanities. However, these subject areas strongly intersect, meaning that an individual document can be classified in several subject areas at one time. Therefore, to address the comprehensiveness of COVID-19 research, Figure 2 shows a Venn diagram of the presented subject areas and all possible sets that can be made from them. This also enables the so-called pure sciences to be determined by covering only those documents that exclusively belong to just one subject area (without intersecting with other subject areas). According to the number of documents obtained on 1 July 2020 (the number of documents obtained on 1 June 2020 is presented in parentheses), health sciences contain a total of 14,187 (8896) documents, of which 10,394 (6575) documents are identified as in the area of the pure health sciences. Further, life sciences encompass a total of 4143 (2549) documents, of which 928 (599) documents are to be in the area of the pure life sciences. Moreover, physical sciences include a total of 1625 (878) documents, of which 568 (314) documents belong to pure physical sciences.



**Figure 2.** Venn diagram by number of documents on COVID-19 research across subject areas (January–June 2020). All documents included.

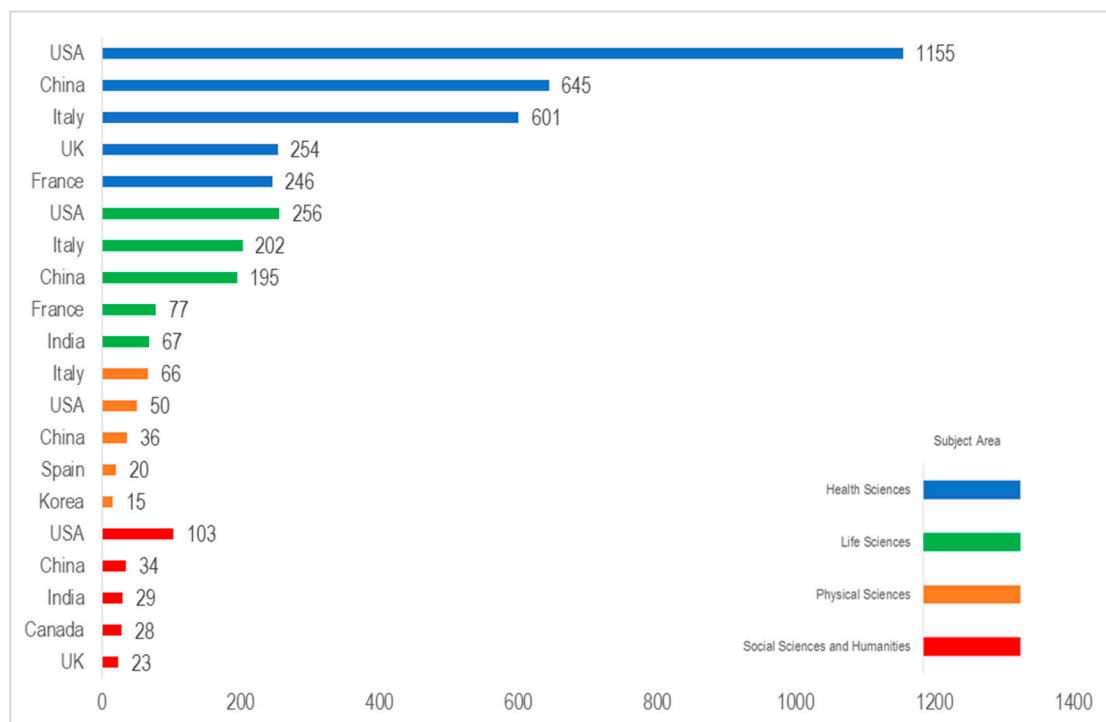
Lastly, the social sciences and humanities cover a total of 1812 (977) documents, of which 771 (323) are to be in the area of the pure social sciences and humanities. A comparison of different subject areas reveals that health sciences are the most relevant in COVID-19 research, while the second-most relevant subject area is life sciences. Moreover, physical sciences and social sciences and humanities seem to be the least popular thus far, as found by previous research [33]. However, considering growth in the number of documents in June 2020, the social sciences seem to be the most increasing scientific discipline as the total number of documents in this subject area rose by 85.5% and even by 138.7% in the

pure social sciences. This is consistent with the expectations and recent COVID-19 bibliometric studies on economics (see Mahi et al. [29]) and business and management (see Verma and Gustafsson [30]). Namely, the first immediate response to the COVID-19 pandemic has been to protect public health, while addressing the real socio-economic consequences may be expected to come later. This path is also revealed by the recent scientific literature on COVID-19 published in the first half of 2020 and a review of the latest COVID-19 publications from 2020, indicating that a shift is underway from health to other relevant scientific disciplines [19]. Moreover, some documents (273) are considered to be multidisciplinary, making it impossible to include them in the further bibliometric analysis across different subject-area categories.

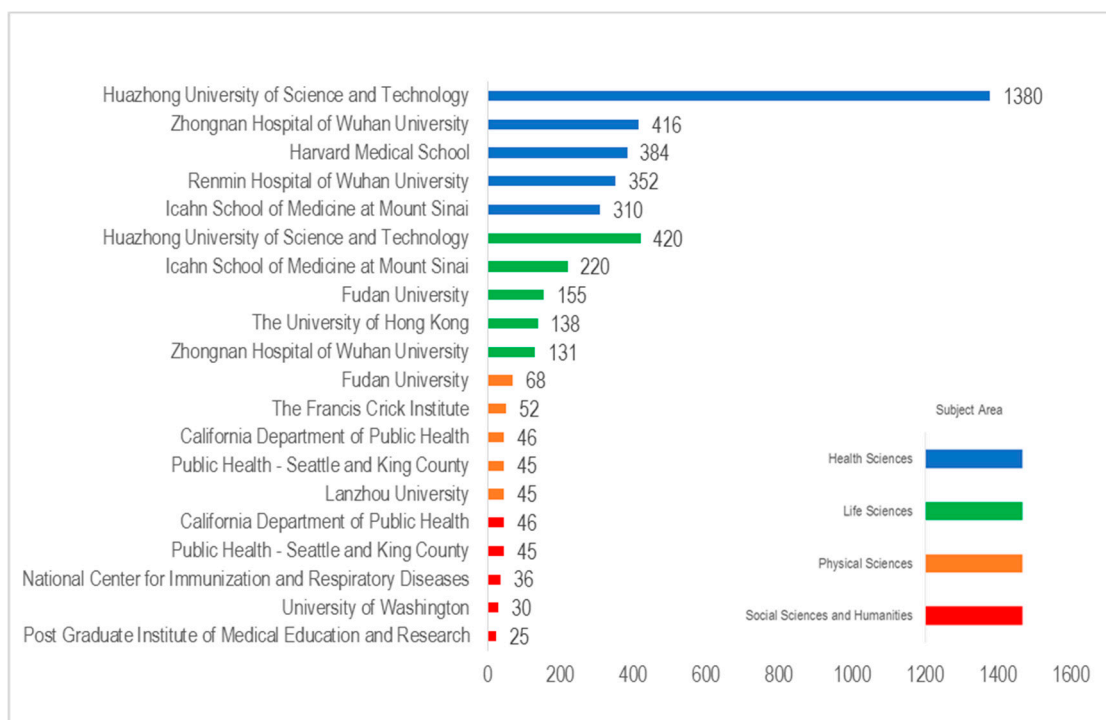
Finally, additional insight into openness of journals reveals that COVID-19 research is very open as 81.3% of total documents are published in open access journals. The highest openness of COVID-19 research is observed for health sciences (82.9% in general and 82.0% for pure science) and life sciences (85.9% in general and 86.7% for pure science), while lower openness is identified for physical sciences (67.9% in general and 50.0% for pure science) and social sciences and humanities (73.7% in general and 70.1% for pure science). In addition, the most relevant (top three) journals and authors (by number of citations) overlapping between at least three subject areas (excluding multidisciplinary subset) are identified. The highest overlap is identified for the following journals: *Morbidity and Mortality Weekly Report* with 619 citations (overlapping between health sciences, physical sciences and social sciences and humanities), *Journal of Pharmaceutical Analysis* with 114 citations and *International Journal of Radiation Oncology Biology Physics* with 64 citations (both overlapping between health, life and physical sciences). Furthermore, the most relevant authors overlapping across three subject areas are: Bialek, S. (CDC COVID-19 Response Team, the USA) with 146 citations, Chow, N. (CDC COVID-19 Response Team, Washington, the USA) with 89 citations (both overlapping between health sciences, physical sciences and social sciences and humanities) and Li, X. (Xi'an Jiaotong University Health Science, Xi'an, China) with 74 citations (overlapping between health, life and physical sciences).

Figure 3 presents the most relevant countries of COVID-19 research by subject area. It shows the top five countries, providing the largest number of documents by a corresponding author. The most relevant country is the USA, significantly dominating in all scientific disciplines, except the physical sciences where it ranks in second place. In addition to the USA, which significantly outperformed other countries, China and Italy also dominate in COVID-19 research since they are among the top three countries in all scientific disciplines, except in the social sciences where Italy is replaced by India. These findings are consistent with existing bibliometric studies (which do not consider scientific disciplines separately) that state that the USA and China are world leaders in COVID-19 research [31–35].

Figure 4 shows the most relevant institutions by the number of documents in COVID-19 research across subject areas. Due to the strong overlap among individual scientific disciplines, to some extent they may share the same most relevant institutions. The most involved institution is the Huazhong University of Science and Technology, providing a significantly higher number of documents in health sciences ( $n = 1380$ ) and life sciences ( $n = 420$ ). Moreover, the Zhongnan Hospital of Wuhan University and Icahn School of Medicine at Mount Sinai also play important roles in these two scientific disciplines. Moreover, Fudan University dominates in the physical sciences ( $n = 68$ ), while providing an enviable number of publications also in the life sciences ( $n = 155$ ). Finally, the California Department of Public Health as well as Public Health—Seattle and King County are the two most relevant institutions in the social sciences and humanities, also with an important role in the physical sciences. The findings are to some extent comparable with existing bibliometric studies on COVID-19 research [33,35].



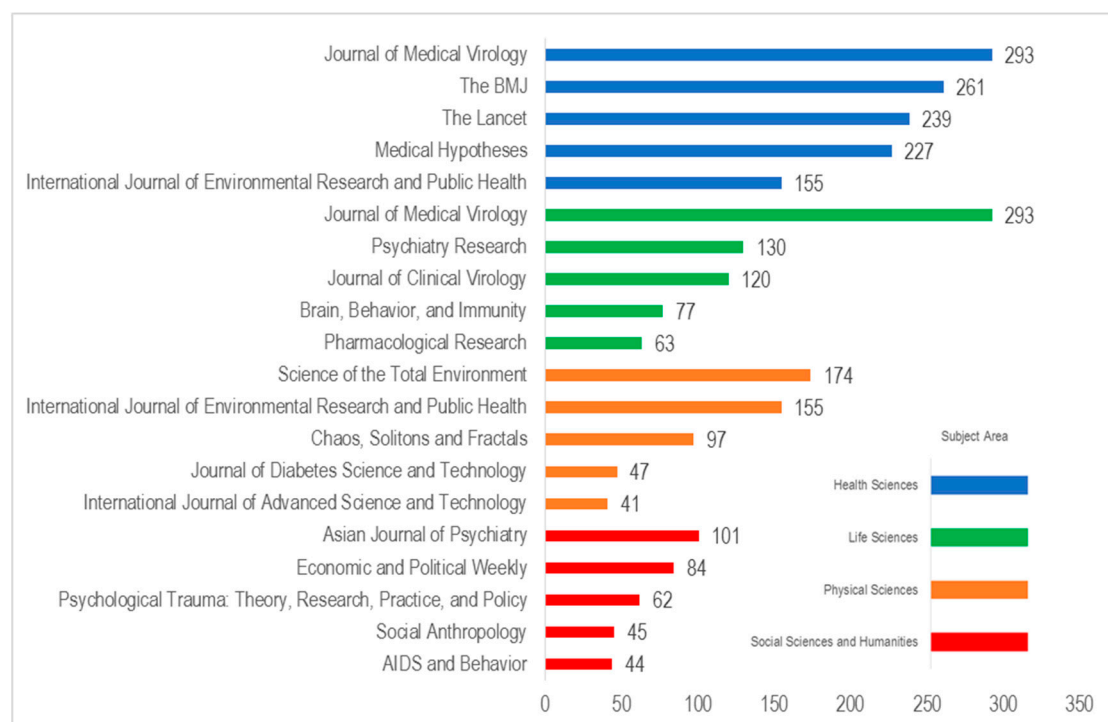
**Figure 3.** Most relevant countries by number of documents in COVID-19 research across subject areas (January–June 2020). Only documents with at least one citation included.



**Figure 4.** Most relevant institutions by number of documents in COVID-19 research across subject areas (January–June 2020). Only documents with at least one citation included.

Figure 5 presents the most relevant journals in COVID-19 research across subject areas. It presents the number of documents provided in certain journals within individual subject areas. In the health sciences, *Journal of Medical Virology* has the most documents ( $n = 293$ ), followed by the *BMJ* ( $n = 261$ ), the *Lancet* ( $n = 239$ ), *Medical Hypotheses* ( $n = 227$ ) and *International Journal of Environmental Research*

and Public Health ( $n = 155$ ). These findings are to some extent consistent with previous COVID-19 bibliometric research not distinguishing between individual scientific disciplines [31,35]. As far as other scientific disciplines are concerned, the results reveal the following. For the life sciences, due to the strong interweaving with the health sciences, the most relevant journal is also *Journal of Medical Virology*, with the most documents ( $n = 293$ ), followed by *Psychiatry Research* ( $n = 130$ ), *Journal of Clinical Virology* ( $n = 120$ ), *Brain, Behaviour and Immunity* ( $n = 77$ ) and *Pharmacological Research* ( $n = 63$ ). In the physical sciences, the most relevant journals are *Science of the Total Environment* ( $n = 174$ ), followed by *International Journal of Environmental Research* ( $n = 155$ ), *Chaos, Solitons and Fractals* ( $n = 97$ ), *Journal of Diabetes Science and Technology* ( $n = 47$ ) and *International Journal of Advanced Science and Technology* ( $n = 41$ ). Finally, for the social sciences and humanities the most relevant journals are *Asian Journal of Psychiatry* ( $n = 101$ ), followed by *Economic and Political Weekly* ( $n = 84$ ), *Psychological Trauma: Theory, Research, Practice, and Policy* ( $n = 62$ ), *Social Anthropology* ( $n = 45$ ) and *AIDS and Behavior* ( $n = 44$ ).



**Figure 5.** Most relevant journals by number of documents in COVID-19 research across subject areas (January–June 2020). All documents included.

Figure 6 shows the most relevant authors by number of citations in COVID-19 research across subject areas. According to the number of total citations, it is evident that Wang, Y. (China-Japan Friendship Hospital, Beijing, China) and Li, X. (Clinical and Research Centre of Infectious Diseases, Beijing, China) are the most important authors involved in COVID-19 research because they feature among the top five cited authors in all four scientific disciplines. This finding is inconsistent with existing bibliometric studies, presumably due to the different criteria applied [33].

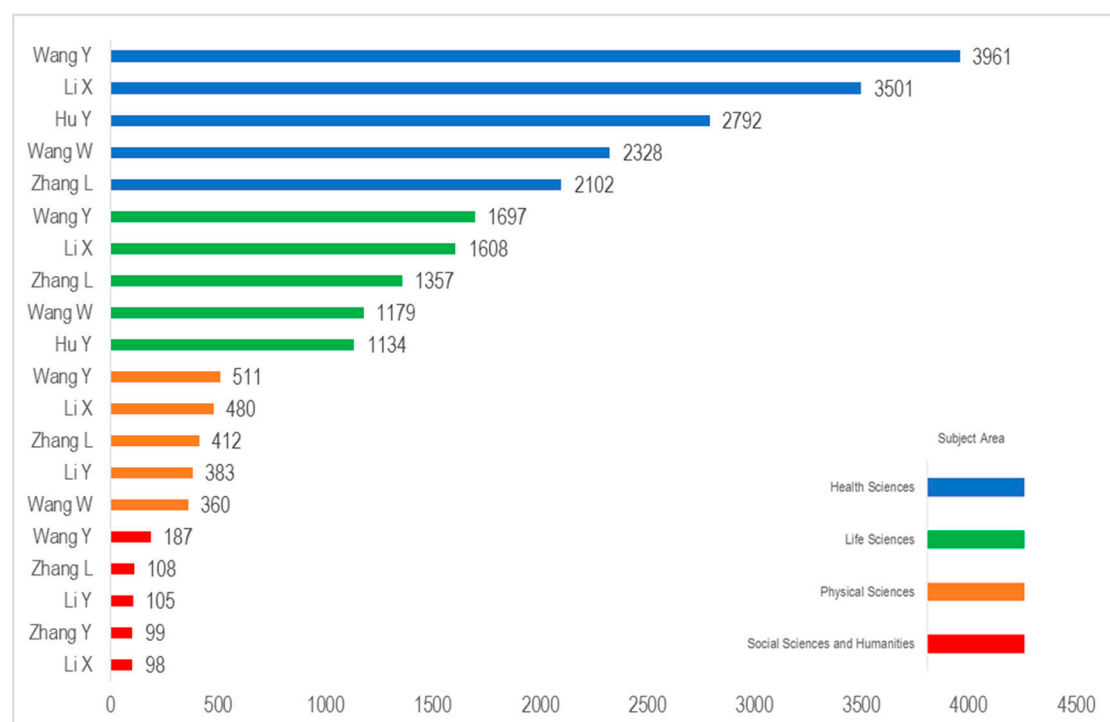
Figure 7 presents the keyword co-occurrence network for: (a) health sciences, (b) life sciences, (c) physical sciences, and (d) social sciences and humanities separately. To ensure a greater distinction between individual subject areas, only pure sciences (without intersecting with other sciences) are considered in the bibliometric analysis. Moreover, the bibliometric analysis is conducted on the 100 most frequent (author and index) keywords by considering the exclusion of the keywords used in the search query, elimination of stop words, and consolidation of keywords describing the same phenomenon.

The bibliometric analysis (keyword co-occurrence) reveals the research hotspots by subject area. For the health sciences, three clusters are identified, addressing the following topics: (1)



pandemics; (2) risk factors and symptoms; and (3) mortality. Accordingly, the health sciences deal predominantly with health-related issues associated with the COVID-19 pandemic. Next, in the life sciences, four clusters are found, dealing with: (1) pandemics; (2) virology; (3) immunology; and (4) drug efficiency. The focus of the life sciences seems to be oriented more to knowledge about the spread of the virus and ways to efficiently prevent the disease with appropriate drugs. This corresponds with findings of other recent bibliometric studies on COVID-19 research, predominantly emphasizing health-related issues [33,34]. In addition, the results for less-explored subject areas are as follows. Regarding the physical sciences, three clusters are recognized, related to: (1) pandemics; (2) China and disease transmission; and (3) air pollution. The physical sciences focus on knowledge relating to how fast the COVID-19 pandemic is spreading and environmental-related issues. Finally, in the social sciences and humanities, six clusters are identified, addressing the following topics: (1) pandemics; (2) epidemics; (3) viral disease and China; (4) respiratory disease; (5) social distancing; and (6) mental health. A detailed synopsis of the research hotspots, including the top 10 keywords, related to COVID-19 in an individual scientific discipline is presented in Table A1 in Appendix A.

Moreover, in order to predict a document's subject area based on its abstract, a text-mining-based classification was used. For this purpose, binary logistic regression was selected as a prediction model. Accordingly, four different binary logistic models were tested for each individual subject area. Based on the results of fitting the model to the data, binary logistic regression also provides information on which words are most characteristic for a particular subject area (which discriminate the most between two subject areas). This approach requires documents with a full abstract, with 8347 documents meeting this criterion. To extract new features for classification, the search for the top 100 characteristic words resulted in 99 unigrams (single words) and 1 bigram (a sequence of two words). These features are further used as predictors (independent variables) in binary logistic models.



**Figure 6.** Most relevant authors by number of citations in COVID-19 research across subject areas (January–June 2020). Total citations included.



**Table 4.** The most discriminant words (with a significant and positive regression coefficient) for predicting a corresponding subject area based on a binary logistic regression (January–June 2020).

Health Sciences	Life Sciences	Physical Sciences	Social Sciences and Humanities
patient, health, healthcare, infection, acute, hospital, child, method, surgery, symptom, disease, medicine, guideline, woman, <i>risk</i> , diabetes, recommendation, <i>clinical</i> , medical, procedure, diagnosis, pneumonia, cancer, surgical, <i>service</i> , <i>experience</i> , therapy, emergency, <i>immune</i> , laboratory, December	protein, human, vaccine, <i>immune</i> , laboratory, RNA, therapeutic, <i>clinical</i> , cancer, drug, testing, worldwide	factor, lockdown, area, transmission, epidemic, infectious, condition, global, spread, virus	crisis, pandemic, mental, government, <i>service</i> , group, <i>experience</i> , <i>risk</i> , people, social, public

Note: Words in italics are identified as the most discriminant in more than one subject area. Only documents with a full abstract were included.

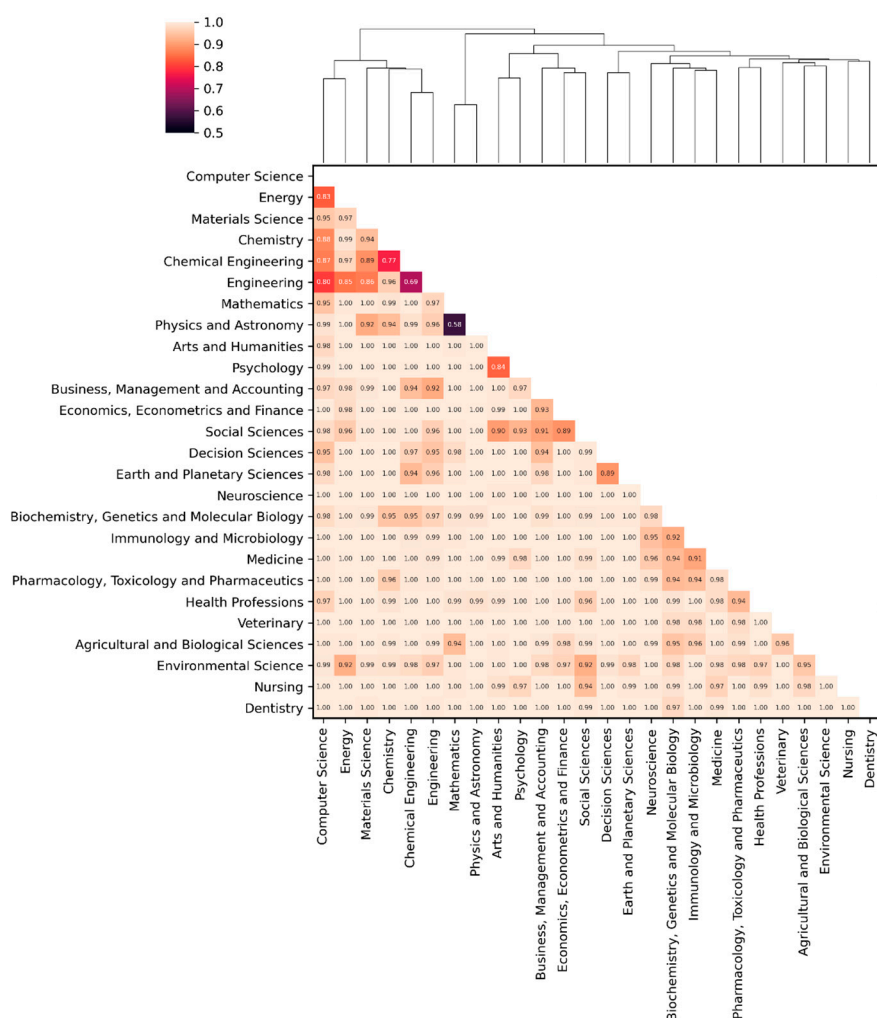
According to the presented results, some interesting relationships between different bibliometric aspects can be identified. Namely, Table 3 shows the most relevant journals in COVID-19 research and consequently the most productive subject areas by number of documents (particularly health sciences), while Table 2 supplements these findings by describing additional lagging subject areas (especially physical sciences and social sciences and humanities [33]. Finally, Table 4 complements these two tables in the sense of highlighting the most discriminant words regardless of subject area dominance and relevance of the journal. Since COVID-19 research is obviously a relatively new field, with the science still evolving, it is important to understand the issue from different perspectives. Nevertheless, the field of COVID-19 research will certainly continue to develop in the future, presumably making a shift from health to other relevant scientific disciplines [19].

### 3.2. Bibliometric Analysis across Different Subject-Area Classifications and Fields

To examine the relationships between different subject-area classifications within COVID-19 research, a cluster analysis based on the Jaccard distance (JD) (Jaccard index subtracted from one), measuring dissimilarity is performed (see Figure 8). Jaccard distance ranges from zero to one, with zero suggesting perfect overlap and one indicating no overlap [43]. Based on the results, the following clusters may be identified. The first and most relatively pronounced cluster is engineering, bringing together: computer science, energy, materials science, chemistry, chemical engineering and engineering. The strong connection between these subject-area classifications is further confirmed by the relatively low Jaccard distance. This is reflected especially between engineering and chemical engineering (JD = 0.69), meaning that 31% (1–0.69) of COVID-19 related documents belonging to either engineering or chemical engineering belong to both subject-area classifications at the same time. One of the strongest (23%) overlaps in this cluster is also found for chemical engineering and chemistry. The second and most pronounced cluster concerns mathematics and physics, as suggested by the lowest Jaccard distance between mathematics and physics and astronomy (JD = 0.58), meaning there is a 42% overlap between these two subject-area classifications.

Furthermore, according to the results, the other subject-area classifications are not very different from each other (the Jaccard distance is equal to or very close to one), making it difficult to identify meaningful or homogeneous clusters. Nevertheless, some further potential or emerging clusters can be identified. Accordingly, the third cluster is the humanities and psychology, grouping the individual subject-area classifications of the arts and humanities and psychology with a 16% overlap. The fourth cluster is business, management and economics, covering business, management and accounting,

economics, econometrics and finance and social sciences, where the most connected subject-area classifications are social sciences and economics, econometrics and finance with an 11% overlap, then social sciences and business, management and accounting with a 9% overlap. The fifth cluster is about decision and earth sciences, grouping individual subject-area classifications of decision sciences and earth and planetary sciences with an 11% overlap. Finally, the sixth cluster concerns health and the environment, covering neuroscience, biochemistry, genetics and molecular biology, immunology and microbiology, medicine, pharmacology, toxicology and pharmaceuticals, health professions, veterinary, agricultural and biological sciences, environmental science, nursing and dentistry. The biggest overlap in this cluster is identified between medicine and immunology and microbiology (9%) and immunology and microbiology and biochemistry, genetics and molecular biology (8%).

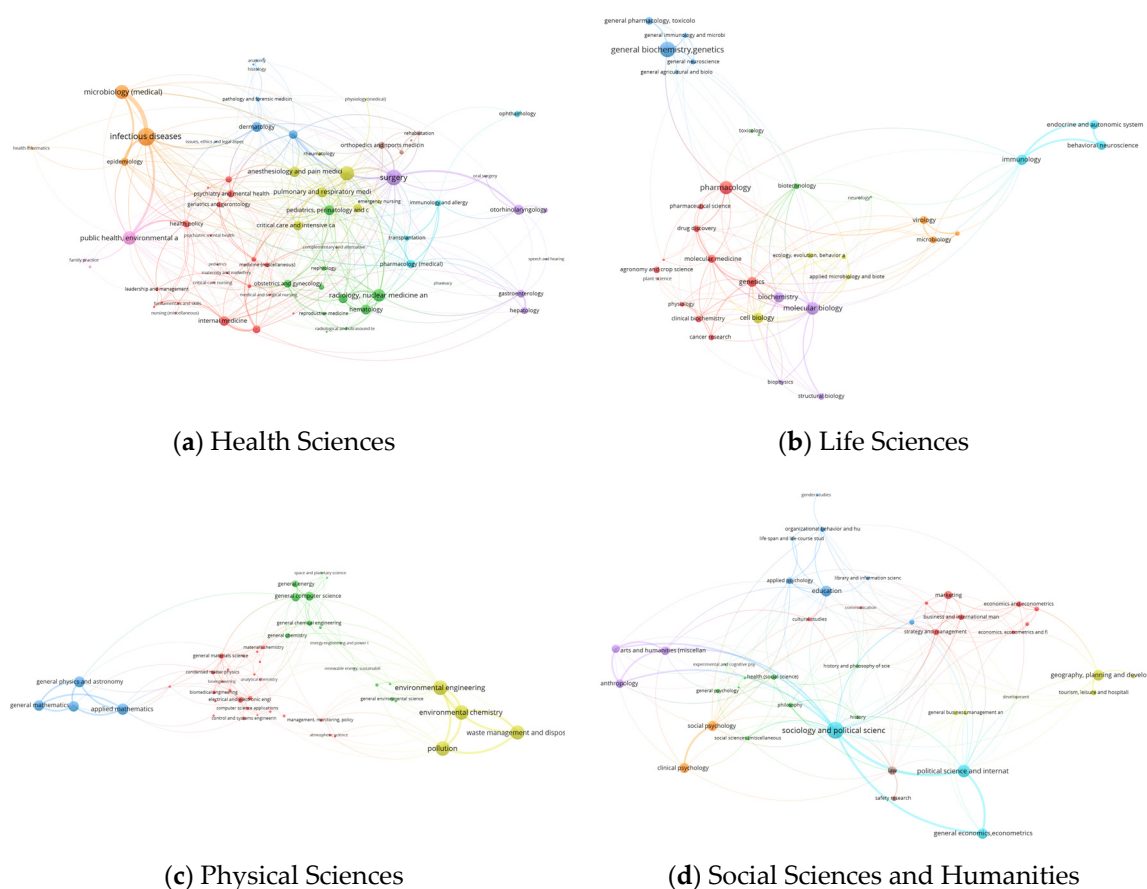


**Figure 8.** Clustermap of COVID-19 research based on Jaccard dissimilarities between subject-area classifications (January–June 2020).

Regarding the overlap of COVID-19 research among different subject-area classifications outside of the identified clusters, the strongest connection is identified between environmental science and energy, physics and astronomy and material science and environmental science and social sciences (8%). This is followed by the overlap between the social sciences and psychology (7%) as well as the connection between the agricultural and biological sciences and mathematics and decision sciences and business, management and accounting (6%). These results provide additional evidence on COVID-19 research collaboration occurring within and between different subject-area classifications [22].



Figure 9 presents the field co-occurrence network for the: (a) health sciences, (b) life sciences, (c) physical sciences, and (d) social sciences and humanities separately. To ensure a greater distinction between individual subject areas, only pure sciences (without intersecting with other sciences) are considered in the bibliometric analysis. Moreover, the bibliometric analysis is conducted on the 297 research fields distributed among these four main subject areas. The bibliometric analysis (field co-occurrence) reveals different clusters related to COVID-19 within an individual subject area. For the health sciences, nine clusters are identified, namely: (1) internal medicine; (2) radiology and hematology; (3) dermatology and neurology; (4) cardiology, pulmonary and anesthesiology; (5) surgery; (6) pharmacology; (7) epidemiology; (8) sports medicine and rehabilitation; and (9) public health. Next, in the life sciences, seven clusters are found, addressing: (1) pharmacology and genetics; (2) biotechnology and toxicology; (3) biochemistry and pharmacology; (4) microbiology and ecology; (5) molecular biology and biochemistry; (6) immunology, neuroscience and endocrine systems; and (7) virology and microbiology. Regarding the physical sciences, four clusters are recognized, related to: (1) electrical/electronic and mechanical engineering; (2) general computer science and engineering (3) mathematics and physics; and (4) environment and pollution. Finally, in the social sciences and humanities, eight clusters are identified, addressing the following topics: (1) business, management and economics; (2) health, philosophy and psychology; (3) education and applied psychology; (4) geography and tourism; (5) humanities and anthropology; (6) sociology and economics; and (7) social and clinical psychology; and law and safety. A detailed synopsis of the clusters, including the top five fields, related to COVID-19 in an individual scientific discipline is presented in Table A3 in Appendix A.



**Figure 9.** Field co-occurrence network in COVID-19 research by subject area (January–June 2020).

#### 4. Discussion and Conclusions

The outbreak of COVID-19 is a typical public health emergency where the high infection rate poses a huge threat to not only global public health but economic and social development. In order to be able to solve such emergencies, it is vital to fully understand the problem, its implications for different areas, and the solutions that may be effective and efficient in addressing potential devastating consequences. Therefore, scientific knowledge on COVID-19 is essential because it leads to answers to real-life questions. However, the extent of the COVID-19 pandemic calls for in-depth knowledge so as to allow numerous issues in different areas to be identified. It is hence not surprising that COVID-19 research has seen such an unprecedented rise since the pandemic started [36,53]. The COVID-19 pandemic has led to the generation of a large amount of scientific publications, which might engender possible problems with the velocity and availability of information and scientific collaboration, particularly in the early stages of the pandemic [54]. The current state of COVID-19 research therefore needs a comprehensive analysis to help guide the agenda for further research, especially from the perspective of cooperation among different scientific disciplines at varying stages of pandemic prevention and control, by applying innovative scientific approaches [55–57].

Accordingly, this paper provides extensive bibliometric analysis of COVID-19 research across the science and social science research landscape by relying on a wide variety of bibliometric approaches, including descriptive analysis, network analysis, cluster analysis based on the Jaccard distance and text mining based on binary logistic regression. The results generally show that a total of 21,400 documents related to COVID-19 research were published in the Scopus database in the first half of 2020. Interestingly, the number of documents had risen by 58.8% in June since May 2020, suggesting an exponential interest in COVID-19 research. The database suitable for the review process includes a total of 16,866 documents, written by 66,504 different authors and published in 2548 different journals, that together provide a total of 100,683 citations. The biggest share of the documents were articles (41.5%) and letters (26.5%), which agrees with previous bibliometric studies [23,32]. Moreover, the distribution of the COVID-19 related documents according to the Scopus hierarchical classification reveals that nearly two-thirds (65.2%) of them are found in the area of the health sciences, supporting the claims that COVID-19 research is lacking knowledge in less-explored subject areas, including the life, physical and social sciences and the humanities [33]. Furthermore, the most relevant journals in COVID-19 research account for almost one-fifth (17.6%) of total documents and a significant share (41.3%) of total citations. With regard to different scientific disciplines or subject areas (classifications), the most relevant journals mainly publish in the health sciences (medicine), while other scientific disciplines (life sciences, physical sciences and social sciences and humanities) remain in the background. Most of these journals rank in the first quartile (Q1) and have a relatively high source-normalized impact per paper (SNIP), which is in line with existing research [31,35]. Finally, most of these journals come from the UK, the Netherlands and the USA. Similar findings have also been made in earlier COVID-19 bibliometric studies [33,34].

A more detailed comparison of COVID-19 research between four scientific disciplines shows that subject areas strongly intersect, which calls for an in-depth analysis of individual subject areas separately. The results of bibliometric analysis across different subject-area categories show the following. According to the number of documents, health sciences is the most relevant subject area in COVID-19 research, the second-most relevant subject area is life sciences, while physical sciences and social sciences and humanities seem to be the least popular hitherto. However, during June 2020 the social sciences seem to be the fastest-growing scientific discipline, with the total number of documents in this subject area rising by 85.5% and even by 138.7% in the pure social sciences. A shift from health to other relevant scientific disciplines is observable in the review of the latest COVID-19 publications as well as in recent COVID-19 bibliometric studies on economics (see Mahi et al. [29]) and business and management (see Verma and Gustafsson [30]). Moreover, the results reveal that most published documents on COVID-19 (81.3%) are found in open access journals. The highest openness of COVID-19 research is observed for health sciences and life sciences, while lower openness is identified for physical



sciences and social sciences and humanities. In addition, the most relevant journals and authors overlapping between at least three subject areas (excluding the multidisciplinary subset) are identified. As regards to journals, the highest overlap is identified for *Morbidity and Mortality Weekly Report* (overlapping between health sciences, physical sciences and social sciences and humanities), *Journal of Pharmaceutical Analysis* and *International Journal of Radiation Oncology Biology Physics* (both overlapping between health, life and physical sciences). Furthermore, the most relevant authors overlapping across three subject areas are: Bialek, S. (CDC COVID-19 Response Team, the USA), Chow, N. (CDC COVID-19 Response Team, Washington, the USA) (both overlapping between health sciences, physical sciences and social sciences and humanities) and Li, X. (Xi'an Jiaotong University Health Science, Xi'an, China) (overlapping between health, life and physical sciences).

The results also suggest that the USA significantly dominates in all scientific disciplines, except physical sciences. Besides the USA, which significantly outperforms other countries, China and Italy dominate in COVID-19 research. Regarding the most relevant institutions, the Huazhong University of Science and Technology, Zhongnan Hospital of Wuhan University, and Icahn School of Medicine at Mount Sinai are playing important roles in health sciences and life sciences. Moreover, Fudan University is dominating in physical sciences while also having a crucial role in life sciences. Finally, the California Department of Public Health along with Public Health—Seattle and King County are the most relevant institutions in the social sciences and humanities, while also playing an important role in the physical sciences. The results regarding journals reveal that *Journal of Medical Virology* is the most relevant journal for the health sciences and life sciences, *Science of the Total Environment* for the physical sciences, and *Asian Journal of Psychiatry* for the social sciences and humanities. Concerning the most important authors, Wang, Y. (China–Japan Friendship Hospital, Beijing, China) and Li, X. (Clinical and Research Center of Infectious Diseases, Beijing, China) stand out in COVID-19 research. The presented results are somewhat comparable with previous bibliometric studies on COVID-19 research [31–35]. Furthermore, the results of keyword co-occurrence analysis by main subject areas reveal different research hotspots for individual scientific disciplines, with the common point of pandemics. The health sciences are focused more on health consequences (see Hossain [33] and Lou et al. [34]), the life sciences are more strongly oriented to drug efficiency, the physical sciences are more focused on environmental consequences, whereas the social sciences are more oriented to socio-economic consequences. In addition, the results of text-mining-based classification based on binary logistic regression reveal the most characteristic words for predicting a corresponding area. For the health sciences, the top three most characteristic words are “patient”, “health” and “healthcare”. As regards to other scientific areas, the top three most characteristic words are “protein”, “human” and “vaccine” for the life sciences, “factor”, “lockdown” and “area” for the physical sciences, and “crisis”, “pandemic” and “mental” for the social sciences and humanities.

Further bibliometric analysis on COVID-19 research across different subject-area classifications and fields provides additional in-depth insights. Namely, a cluster analysis based on the Jaccard distance reveals six different clusters: engineering, mathematics and physics, humanities and psychology, business management and economics, decision and earth sciences and health and environment. Regarding the overlap of COVID-19 research among different subject-area classifications outside of the identified clusters, the strongest connection is seen between environmental science and energy, physics and astronomy and material science and environmental science and social sciences. These results provide further evidence about COVID-19 research collaboration occurring within and between different subject-area classifications [22]. The results of field co-occurrence analysis by main subject areas also reveal different research clusters of fields, providing a detailed segmentation of different scientific disciplines.

Several limitations of the present study should be noted. First, the bibliometric analysis is only based on COVID-19 related documents retrieved from the Scopus database and published in journal with available Scopus CiteScore metrics. Although Scopus is considered to be one of the largest abstract and citation databases of peer-reviewed literature, it might not cover the complete collection

of COVID-19 research. Therefore, the inclusion of other databases, especially the expanding body of preprints available in the Google Scholar database, could have provided additional insights not available in this study. Second, this study is based on a short time period (first half of 2020). Although this limitation cannot be solved at this stage, a repeated study with a longer period would yield further time-dimensional insights. This would also be beneficial in terms of achieving a higher number of publications in some under-represented disciplines, especially the social sciences and humanities. Another limitation is that only titles, abstracts and keywords in the English language were included in this study, which might cause some publication bias. Future studies should therefore address this issue. Finally, another study limitation is the lack of citation and collaboration networks that could be identified using sophisticated methodological approaches due to the small number of studies and continuously changing citations metrics. Accordingly, future bibliometric studies should address these limitations and further examine the evolution of scientific knowledge about COVID-19 across different scientific disciplines over time.

Notwithstanding the above limitations, the findings of the paper highlight the importance of a comprehensive and in-depth approach that considers different scientific disciplines in COVID-19 research. In order to address the economic, socio-cultural, political, environmental and other (non-medical) consequences of the COVID-19 pandemic, in the near future COVID-19 must appear higher up the research agenda of non-health sciences, particularly the social sciences and humanities. Namely, understanding of the evolution of emerging scientific knowledge on COVID-19 is not only beneficial for the scientific community, but for evidence-based policymaking with a view to fully addressing the implications of the COVID-19 pandemic.

**Author Contributions:** A.A. supervised the work on the paper and revised it. D.R. wrote the paper. L.U. performed the analysis. All authors read and agreed to the published version of the manuscript.

**Funding:** This research and the APC were funded by the Slovenian Research Agency grant number P5-0093.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Research hotspots based on keyword co-occurrence network in COVID-19 research across subject areas (January–June 2020).

Subject Area	Research Hotspots	Keywords
Health Sciences	Pandemics	Humans, Pandemics, Pneumonia, Epidemic, China, Infection Control, Virus Transmission, Health Care Personnel, Procedures, Practice Guideline
	Risk Factors and Symptoms	Female, Male, Adult, Fever, Middle Aged, Aged, Clinical Article, Coughing, Case Report, Computer Assisted Tomography
	Mortality	Nonhuman, Disease Severity, Virology, Complication, Risk Factor, Intensive Care Unit, Mortality, Mortality Rate, Hospitalization, Comorbidity
Life Sciences	Pandemics	Humans, Pandemics, Pneumonia, China, Epidemic, Virus Transmission, Disease Severity, Female, Male, Adult
	Virology	Nonhuman, Angiotensin Converting Enzyme 2, Virology, Genetics, Controlled Study, Animals, Animal, Drug Effect, Physiology, Metabolism
	Immunology	Immunology, Virus Replication, Immune Response, Inflammation, Protein Expression, Interleukin 6, Pathophysiology, Pathology, Signal Transduction, Pathogenicity
	Drug Efficiency	Unclassified Drug, Antivirus Agent, Remdesivir, Hydroxychloroquine, Antiviral Activity, Antiviral Agents, Virus Genome, Drug Efficacy, Chloroquine, Lopinavir Plus Ritonavir

Table A1. Cont.

Subject Area	Research Hotspots	Keywords
Physical Sciences	Pandemics	Pandemics, Humans, Pneumonia, Virus, Viral Disease, Diseases, Epidemic, Respiratory Disease, Epidemiology, Disease Transmission
	China and Disease Transmission	China, Infectious Diseases, Transmissions, Temperature, Humidity, Italy, Environmental Temperature, Population Statistics, Major Clinical Study, Air Temperature
	Air Pollution	Air Quality, Air Pollution, Particulate Matter, Nitrogen Dioxide, Concentration (Composition), Nitrogen Oxides, Quarantine, Atmospheric Pollution, City, Environmental Monitoring
Social Sciences and Humanities	Pandemics	Pandemics, Crisis, Resilience, Inequality, Lockdown, India, Tourism, Globalization, Learning, Teaching
	Epidemics	Epidemic, Human Resource Management, Analytics, Critical Care, Differential Equations, Discrete Time Markov Chains, Forecasting, Forecasting Models, Hubei Province, Intensive Care Units
	Viral Disease and China	Viral Disease, China, Public Health, Infectious Diseases, Virus, Disease Spread, Australia, Disease Control, Migration, South Korea
	Respiratory Disease	Respiratory Disease, Health Care, Health Care Personnel, Health Equity, Supply Chain Management, Vulnerability, Disease, Predisposition, Government, Health Care Availability, Health Care Planning
	Social Distancing *	Social Distancing, Consumer Behavior, Social Media, Digital Technology, Health Care Workers
	Mental Health	Mental Health, Humans, Pneumonia, Trauma, Psychology, PTSD, Anxiety, Female, Male, Stress

Note: \* Only 5 keywords are identified for this cluster.

**Table A2.** The results of binary logistic models (betas and *p*-values) for the classification of COVID-19 documents across subject areas (January–June 2020).

Binary Logistic Model	Health Sciences		Life Sciences		Physical Sciences		Social Sciences & Humanities	
	(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)	
	Beta	<i>p</i> -Value	Beta	<i>p</i> -Value	Beta	<i>p</i> -Value	Beta	<i>p</i> -Value
acute	1.682	0.000	0.426	0.107	−2.123	0.000	−3.483	0.000
admission	0.758	0.234	0.089	0.811	−1.206	0.188	−1.055	0.445
age	0.673	0.131	0.324	0.365	−1.339	0.018	−0.648	0.299
antiviral	−0.459	0.172	0.415	0.180	−0.133	0.765	−1.134	0.226
april	0.243	0.529	−0.878	0.022	0.120	0.779	−1.234	0.025
area	−0.683	0.051	−0.618	0.086	1.666	0.000	−0.769	0.098
cancer	1.369	0.006	0.992	0.001	−1.129	0.091	−1.745	0.049
cell	−0.145	0.591	0.073	0.760	−1.538	0.000	−2.561	0.005
challenge	−0.322	0.287	−1.007	0.002	−0.605	0.095	0.178	0.606
change	−1.022	0.001	−0.835	0.008	0.016	0.964	0.669	0.060
characteristic	0.041	0.925	0.277	0.429	−0.162	0.760	−0.341	0.630
chest	0.627	0.247	−1.417	0.000	−0.282	0.660	−1.338	0.319
child	1.693	0.000	−0.966	0.001	−1.996	0.000	−0.734	0.100
china	0.202	0.576	−0.193	0.582	−0.395	0.322	−1.075	0.046
clinical	1.262	0.000	0.928	0.000	−2.986	0.000	−2.940	0.000
community	0.441	0.155	−0.277	0.392	−0.564	0.128	−0.089	0.810
compared	−0.141	0.713	0.372	0.253	0.418	0.341	−1.564	0.015
concern	0.132	0.718	−0.056	0.870	−0.239	0.587	−0.034	0.943

Table A2. Cont.

Binary Logistic Model	Health Sciences		Life Sciences		Physical Sciences		Social Sciences & Humanities	
	(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)	
	Beta	p-Value	Beta	p-Value	Beta	p-Value	Beta	p-Value
condition	−0.737	0.034	−0.431	0.186	0.958	0.015	0.274	0.569
confirmed	−0.209	0.530	−0.633	0.038	0.191	0.604	−1.899	0.004
country	0.434	0.090	−0.838	0.003	−0.291	0.306	−0.642	0.053
crisis	−1.847	0.000	−1.789	0.000	−1.659	0.000	2.022	0.000
death	−0.306	0.301	−0.473	0.097	−0.124	0.713	−0.287	0.536
december	0.960	0.046	0.699	0.089	−0.941	0.095	−1.074	0.243
diabetes	1.701	0.000	0.194	0.511	−1.101	0.045	−1.634	0.027
diagnosis	1.281	0.004	0.224	0.455	−0.938	0.090	−1.712	0.072
disease	1.182	0.000	−0.767	0.004	−0.843	0.021	−4.145	0.000
drug	−1.286	0.000	0.648	0.007	0.166	0.629	−1.387	0.033
emergency	0.824	0.023	−0.548	0.099	−0.903	0.043	−0.849	0.066
epidemic	0.049	0.844	−1.169	0.000	0.734	0.005	−0.780	0.028
experience	0.871	0.012	−0.537	0.110	−1.950	0.000	1.131	0.004
factor	−0.990	0.001	0.305	0.281	1.791	0.000	0.064	0.878
february	−0.072	0.870	0.124	0.752	−0.089	0.855	−1.157	0.125
finding	−0.312	0.358	−0.978	0.002	−0.534	0.206	−0.414	0.413
global	−1.355	0.000	−0.290	0.312	0.723	0.018	−0.001	0.998
government	−1.147	0.000	−1.592	0.000	0.014	0.967	1.442	0.000
group	0.077	0.789	−0.302	0.227	−2.053	0.000	1.129	0.003
guideline	1.860	0.000	−0.723	0.069	−1.242	0.039	−0.713	0.219
health	2.374	0.000	−0.805	0.014	−1.108	0.004	−2.080	0.000
healthcare	2.292	0.000	−0.816	0.010	−1.054	0.009	−1.688	0.000
hospital	1.935	0.000	−1.115	0.000	−1.513	0.003	−2.350	0.000
human	−1.028	0.000	1.463	0.000	0.254	0.411	−0.591	0.189
illness	0.621	0.160	−0.319	0.354	−1.083	0.061	0.135	0.833
immune	0.766	0.026	1.410	0.000	−1.450	0.004	−1.787	0.038
individual	−0.465	0.110	−0.504	0.097	−0.102	0.765	0.136	0.721
infected	−0.507	0.153	−0.146	0.636	0.651	0.112	−1.188	0.078
infection	1.416	0.000	0.127	0.575	−1.750	0.000	−2.919	0.000
infectious	−0.035	0.923	−0.080	0.812	1.031	0.010	−0.821	0.220
information	−0.527	0.061	−0.784	0.010	0.557	0.069	−0.664	0.067
international	−0.176	0.593	−0.264	0.435	−1.361	0.002	0.752	0.065
intervention	−0.106	0.760	−0.736	0.044	−0.447	0.282	−0.420	0.388
laboratory	1.041	0.026	1.294	0.000	−1.128	0.062	−1.371	0.138
lockdown	−1.602	0.000	−1.198	0.000	1.298	0.000	−0.802	0.010
lung	−0.020	0.955	−0.758	0.008	−0.761	0.141	−1.745	0.122
march	0.282	0.383	−1.068	0.001	−0.288	0.429	−0.875	0.052
mechanism	−0.222	0.509	0.039	0.897	−0.487	0.267	−1.097	0.128
medical	1.023	0.001	−1.302	0.000	−1.453	0.000	−0.991	0.013
medicine	1.542	0.000	−0.302	0.387	−1.839	0.001	−1.671	0.008
mental	0.171	0.593	0.511	0.112	−2.091	0.000	1.615	0.000
method	1.561	0.000	−1.008	0.002	−0.307	0.444	−2.900	0.000
mortality	0.468	0.193	0.221	0.427	−0.827	0.067	−2.068	0.005
organization	−0.575	0.158	0.094	0.824	−0.296	0.533	−0.201	0.689
outbreak	−0.533	0.046	−0.087	0.741	0.294	0.318	−0.380	0.312
outcome	0.227	0.551	−0.239	0.415	−0.326	0.515	−0.631	0.292
pandemic	−1.071	0.000	−1.643	0.000	−1.439	0.000	1.610	0.000
patient	4.775	0.000	−0.323	0.154	−5.349	0.000	−6.197	0.000
people	0.207	0.452	−0.694	0.019	−0.682	0.034	0.737	0.026
pneumonia	1.144	0.005	−0.872	0.003	−0.894	0.077	−1.727	0.073
procedure	1.678	0.002	−1.587	0.000	−0.986	0.101	−1.336	0.068
protective	0.495	0.238	−0.680	0.075	−0.489	0.321	−1.232	0.033
protein	−0.366	0.178	1.866	0.000	−0.571	0.087	−2.245	0.005

Table A2. Cont.

Binary Logistic Model	Health Sciences		Life Sciences		Physical Sciences		Social Sciences & Humanities	
	(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)		(1-Yes, 0-No)	
	Beta	p-Value	Beta	p-Value	Beta	p-Value	Beta	p-Value
public	−0.629	0.137	−0.260	0.589	−0.051	0.911	1.054	0.032
public health	0.886	0.102	0.282	0.626	−0.211	0.727	−0.655	0.315
recommendation	1.746	0.000	−0.937	0.015	−1.270	0.034	−1.035	0.082
resource	0.311	0.416	−0.932	0.017	−0.716	0.118	0.187	0.666
risk	1.089	0.000	−0.648	0.015	−0.541	0.122	0.998	0.008
rna	−0.973	0.003	1.103	0.000	−0.143	0.714	−1.569	0.098
service	0.913	0.008	−0.751	0.036	−0.825	0.048	1.163	0.001
social	−0.261	0.287	−1.589	0.000	−0.363	0.198	0.589	0.032
society	0.407	0.169	−1.848	0.000	−1.129	0.002	−0.469	0.198
spread	−0.728	0.012	−0.808	0.007	0.718	0.023	−0.150	0.717
strategy	0.083	0.771	−0.583	0.049	0.048	0.882	−0.167	0.648
surgery	2.642	0.000	−2.302	0.000	−2.106	0.005	−2.433	0.001
surgical	1.701	0.008	−1.420	0.006	−1.240	0.092	−1.955	0.020
symptom	1.454	0.000	−0.439	0.075	−1.855	0.000	−1.266	0.044
testing	0.303	0.391	0.784	0.007	−0.790	0.073	−1.005	0.069
therapeutic	−0.750	0.025	1.103	0.000	−0.538	0.264	−1.638	0.067
therapy	0.897	0.021	−0.191	0.502	−1.433	0.017	−1.322	0.113
transmission	−0.143	0.608	−1.139	0.000	1.010	0.001	−1.949	0.000
treatment	0.078	0.802	0.314	0.210	−0.993	0.021	−2.274	0.001
trial	0.254	0.524	−0.281	0.375	−1.240	0.088	−1.038	0.278
vaccine	0.456	0.126	1.618	0.000	−0.211	0.561	−1.318	0.017
viral	0.281	0.366	0.048	0.854	−0.591	0.134	−2.356	0.005
virus	−0.184	0.454	−0.009	0.968	0.574	0.041	−0.700	0.101
woman	1.509	0.000	−1.243	0.001	−1.847	0.001	−0.766	0.125
worker	0.493	0.251	−0.108	0.776	−0.846	0.093	0.405	0.403
world	−0.548	0.071	−0.356	0.253	0.210	0.537	0.497	0.191
worldwide	0.465	0.190	0.694	0.032	−0.357	0.399	−0.753	0.147
wuhan	0.612	0.161	0.467	0.228	−0.757	0.122	−1.402	0.076
year	0.025	0.948	−0.965	0.007	−0.537	0.224	−0.360	0.462
Pseudo R <sup>2</sup>	0.256		0.146		0.217		0.403	
LLR p-value	<0.001		<0.001		<0.001		<0.001	
AUC	0.824		0.750		0.822		0.910	
CA	0.807		0.761		0.881		0.912	
Precision	0.793		0.740		0.858		0.900	
Recall	0.807		0.761		0.881		0.912	

Note: Unadjusted *p*-values are presented. Only documents with a full abstract included.

**Table A3.** Clusters based on the field co-occurrence network in COVID-19 research across different subject areas (January–June 2020).

Subject Area	Clusters	Fields
Health Sciences	Internal Medicine	Internal Medicine; Endocrinology, Diabetes and Metabolism; Psychiatry and Mental Health; Health Policy; General Nursing
	Radiology and Hematology	Radiology, Nuclear Medicine and Imaging; Hematology; Pediatrics, Perinatology and Child Health; Oncology; Obstetrics and Gynecology
	Dermatology and Neurology	Dermatology; Neurology (Clinical); Pathology and Forensic; Medicine; Histology; Anatomy
	Cardiology, Pulmonary and Anesthesiology	Cardiology and Cardiovascular Medicine; Pulmonary and Respiratory Medicine; Anesthesiology and Pain Medicine; Critical Care and Intensive Care Medicine; Emergency Medicine
	Surgery	Surgery; Otorhinolaryngology; Gastroenterology; Hepatology; General Dentistry
	Pharmacology	Pharmacology (Medical); Ophthalmology; Immunology and Allergy; Transplantation; Optometry
	Epidemiology	Infectious Diseases; Microbiology (Medical); Epidemiology; Health Informatics; Health Information Management
	Sports Medicine and Rehabilitation	Orthopedics and Sports Medicine; Physical Therapy, Sports Therapy and Rehabilitation; Rehabilitation; Complementary and Alternative Medicine; Occupational Therapy
Life Sciences	Public Health *	Public Health, Environmental and Occupational Health; Family Practice; Community and Home Care
	Pharmacology and Genetics	Pharmacology; Genetics; Molecular Medicine; Drug Discovery; Clinical Biochemistry
	Biotechnology and Toxicology	Biotechnology; Toxicology; Food Science; Neurology; Aging
	Biochemistry and Pharmacology	General Biochemistry, Genetics and Molecular Biology; General Pharmacology, Toxicology and Pharmaceutics; General Neuroscience; General Immunology and Microbiology; General Agricultural and Biological Sciences
	Microbiology and Ecology *	Cell Biology; Ecology, Evolution, Behavior and Systematics; Applied Microbiology and Biotechnology; Developmental Biology
	Molecular Biology and Biochemistry *	Molecular Biology; Biochemistry; Structural Biology; Biophysics
	Immunology, Neuroscience and Endocrine Systems *	Immunology; Behavioral Neuroscience; Endocrine and Autonomic Systems
	Virology and Microbiology *	Virology; Microbiology; Parasitology



Table A3. Cont.

Subject Area	Clusters	Fields
Physical Sciences	Electrical/Electronic and Mechanical Engineering	Electrical and Electronic Engineering; General Materials Science; Mechanical Engineering; Condensed Matter Physics; Materials Chemistry
	General Computer Science and Engineering	General Computer Science; General Engineering; General Energy; General Chemistry; General Chemical Engineering
	Mathematics and Physics *	Applied Mathematics; General Physics and Astronomy; Statistical and Nonlinear Physics; General Mathematics
	Environment and Pollution *	Environmental Chemistry; Pollution; Environmental Engineering; Waste Management and Disposal
Social Sciences and Humanities	Business, Management and Economics	Marketing; Strategy and Management; Business and International Management; Economics and Econometrics; Finance
	Health, Philosophy and Psychology	Health (Social Science); Philosophy; Social Sciences (Miscellaneous); General Psychology; History
	Education and Applied Psychology	Education; Applied Psychology; Organizational Behavior and Human Resource Management; Public Administration; Library and Information Sciences
	Geography and Tourism	Geography, Planning and Development; Tourism, Leisure and Hospitality Management; General Business, Management and Accounting; General Social Sciences; Urban Studies
	Humanities and Anthropology *	Arts and Humanities (Miscellaneous); Anthropology; Developmental and Educational Psychology
	Sociology and Economics *	Sociology and Political Science; Political Science and International Relations; General Economics, Econometrics and Finance
	Social and Clinical Psychology *	Social Psychology; Clinical Psychology
	Law and Safety *	Law; Safety Research

Note: \* Less than 5 fields are identified for this cluster.

## References

1. Fan, Y.; Zhao, K.; Shi, Z.L.; Zhou, P. Bat Coronaviruses in China. *Viruses* **2019**, *11*, 210. [CrossRef]
2. World Health Organization. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19. Available online: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed on 1 September 2020).
3. Bogoch, I.I.; Watts, A.; Thomas-Bachli, A.; Huber, C.; Kraemer, M.U.; Khan, K. Pneumonia of unknown aetiology in Wuhan, China: Potential for international spread via commercial air travel. *J. Travel Med.* **2020**, *27*, 1–3. [CrossRef]
4. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **2020**, *93*, 211–216. [CrossRef]
5. Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* **2020**, *395*, 689–697. [CrossRef]
6. Gao, X.; Yu, J. Public governance mechanism in the prevention and control of the COVID-19: Information, decision-making and execution. *J. Chin. Gov.* **2020**, *5*, 178–197. [CrossRef]
7. ECDC. COVID-19 Situation Update Worldwide, as of 1 July 2020. Available online: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases> (accessed on 1 July 2020).
8. IMF. *World Economic Outlook, April 2020: The Great Lockdown*; IMF: Washington, DC, USA, 2020.
9. OECD. *OECD Economic Outlook, June 2020*; OECD: Paris, France, 2020.
10. Hu, Y.; Chen, M.; Wang, Q.; Zhu, Y.; Wang, B.; Li, S.; Xu, Y.; Zhang, Y.; Liu, M.; Wang, Y.; et al. From SARS to COVID-19: A bibliometric study on emerging infectious diseases with natural language processing technologies. *Res. Sq.* **2020**. [CrossRef]
11. Zhai, F.; Zhai, Y.; Cong, C.; Song, T.; Xiang, R.; Feng, T.; Liang, Z.; Zeng, Y.; Yang, J.; Yang, J.; et al. Research Progress of Coronavirus Based on Bibliometric Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3766. [CrossRef]
12. Zhou, Y.; Chen, L. Twenty-Year Span of Global Coronavirus Research Trends: A Bibliometric Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3082. [CrossRef] [PubMed]
13. Herrera-Viedma, E.; López-Robles, J.R.; Guallar, J.; Cobo, M.J. Global trends in coronavirus research at the time of Covid-19: A general bibliometric approach and content analysis using SciMAT. *El Profesional de la Información* **2020**, *29*. [CrossRef]
14. Ram, S. Coronavirus Research Trends: A 50-Year Bibliometric Assessment. *Sci. Tech. Libr.* **2020**, *39*, 210–266. [CrossRef]
15. Joshua, V.; Sivaprakasam, S. Coronavirus: Bibliometric analysis of scientific publications from 1968 to 2020. *Med. J. Islam. Repub. Iran* **2020**, *34*, 456–463. [CrossRef]
16. Chahrour, M.; Assi, S.; Bejjani, M.; Nasrallah, A.A.; Salhab, H.; Fares, M.; Khachfe, H.H. A bibliometric analysis of Covid-19 research activity: A call for increased output. *Cureus* **2020**, *12*. [CrossRef]
17. Tao, Z.; Zhou, S.; Yao, R.; Wen, K.; Da, W.; Meng, Y.; Yang, K.; Liu, H.; Tao, L. COVID-19 will stimulate a new coronavirus research breakthrough: A 20-year bibliometric analysis. *Ann. Transl. Med.* **2020**, *8*, 528. [CrossRef]
18. CORD-19. COVID-19 Open Research Dataset. Available online: <https://www.semanticscholar.org/cord19> (accessed on 28 June 2020).
19. Colavizza, G.; Costas, R.; Traag, V.A.; Van Eck, N.J.; Van Leeuwen, T.; Waltman, L. A scientometric overview of CORD-19. *BioRxiv* **2020**. [CrossRef]
20. Odone, A.; Salvati, S.; Bellini, L.; Bucci, D.; Capraro, M.; Gaetti, G.; Amerio, A.; Signorelli, C. The runaway science: A bibliometric analysis of the COVID-19 scientific literature. *Acta Biomed* **2020**, *91*, 34–39. [CrossRef] [PubMed]
21. Locher, C.; Moher, D.; Cristea, I.; Florian, N. Publication by association: The Covid-19 pandemic reveals relationships between authors and editors. *MetaArXiv* **2020**. [CrossRef]
22. Lee, J.J.; Haupt, J.P. Scientific Globalism during A Global Crisis: Research Collaboration and Open Access Publications on COVID-19. *Res. Sq.* **2020**. [CrossRef]

23. Sa'ed, H.Z.; Al-Jabi, S.W. Mapping the situation of research on coronavirus disease-19 (COVID-19): A preliminary bibliometric analysis during the early stage of the outbreak. *BMC Infect. Dis.* **2020**, *20*, 561. [\[CrossRef\]](#)
24. Fan, J.; Gao, Y.; Zhao, N.; Dai, R.; Zhang, H.; Feng, X.; Shi, G.; Tian, J.; Chen, C.; Hambly, B.D.; et al. Bibliometric analysis on COVID-19: A comparison of research between English and Chinese studies. *Front. Public Health* **2020**, *8*, 477. [\[CrossRef\]](#)
25. Andersen, J.P.; Nielsen, M.W.; Simone, N.L.; Lewiss, R.E.; Jagsi, R. Meta-Research: COVID-19 medical papers have fewer women first authors than expected. *Elife* **2020**, *9*. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Vasantha Raju, N.; Patil, S.B. Indian Publications on SARS-CoV-2: A bibliometric study of WHO COVID-19 database. *Diabetes Metab. Syndr.* **2020**, *14*, 1171. [\[CrossRef\]](#)
27. ElHawary, H.; Salimi, A.; Diab, N.; Smith, L. Bibliometric Analysis of Early COVID-19 Research: The Top 50 Cited Papers. *Infect. Dis. Res. Treat.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Yang, K.L.; Jin, X.Y.; Gao, Y.; Xie, J.; Liu, M.; Zhang, J.H.; Tian, J.H. Bibliometric analysis of researches on traditional Chinese medicine for coronavirus disease 2019 (COVID-19). *Integr. Med. Res.* **2020**, *9*, 100490. [\[CrossRef\]](#)
29. Mahi, M.; Mobin, M.A.; Habib, M.; Akter, S. Knowledge Mapping of Pandemic and Epidemic Studies in Economics: Future Agenda for COVID-19 Research. *SSRN* **2020**. [\[CrossRef\]](#)
30. Verma, S.; Gustafsson, A. Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *J. Bus. Res.* **2020**, *118*, 253–261. [\[CrossRef\]](#)
31. Dehghanbanadaki, H.; Seif, F.; Vahidi, Y.; Razi, F.; Hashemi, E.; Khoshmirsafa, M.; Aazami, H. Bibliometric analysis of global scientific research on Coronavirus (COVID-19). *Med. J. Islam. Repub. Iran* **2020**, *34*, 354–362. [\[CrossRef\]](#)
32. Hamidah, I.; Sriyono, S.; Hudha, M.N. A Bibliometric Analysis of Covid-19 Research using VOSviewer. *Indones. J. Sci. Technol.* **2020**, *5*, 34–41. [\[CrossRef\]](#)
33. Hossain, M.M. Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. *F1000Research* **2020**. [\[CrossRef\]](#)
34. Lou, J.; Tian, S.J.; Niu, S.M.; Kang, X.Q.; Lian, H.X.; Zhang, L.X.; Zhang, J.J. Coronavirus disease 2019: A bibliometric analysis and review. *Eur. Rev. Med. Pharmacol. Sci.* **2020**, *24*, 3411–3421. [\[CrossRef\]](#)
35. Nasab, F.R. Bibliometric Analysis of Global Scientific Research on SARSCoV-2 (COVID-19). *MedRxiv* **2020**. [\[CrossRef\]](#)
36. Kambhampati, S.B.; Vaishya, R.; Vaish, A. Unprecedented surge in publications related to COVID-19 in the first three months of pandemic: A bibliometric analytic report. *J. Clin. Orthop. Trauma* **2020**, *11*, S304–S306. [\[CrossRef\]](#)
37. McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2012.
38. Wang, C.; Lim, M.K.; Zhao, L.; Tseng, M.L.; Chien, C.F.; Lev, B. The evolution of Omega-The International Journal of Management Science over the past 40 years: A bibliometric overview. *Omega* **2020**, *93*, 102098. [\[CrossRef\]](#)
39. De Felice, F.; Polimeni, A. Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis. *In Vivo* **2020**, *34*, 1613–1617. [\[CrossRef\]](#)
40. Moral-Muñoz, J.A.; Herrera-Viedma, E.; Santisteban-Espejo, A.; Cobo, M.J. Software tools for conducting bibliometric analysis in science: An up-to-date review. *El profesional de la información* **2020**, *29*. [\[CrossRef\]](#)
41. VanderPlas, J. *Python Data Science Handbook: Essential Tools for Working with Data*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016.
42. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [\[CrossRef\]](#)
43. Levandowsky, M.; Winter, D. Distance between sets. *Nature* **1971**, *234*, 34–35. [\[CrossRef\]](#)
44. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010; van der Walt, S., Millman, J., Eds.; pp. 62–96.
45. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [\[CrossRef\]](#)
46. Ye, Z.; Tafti, A.P.; He, K.Y.; Wang, K.; He, M.M. Sparktext: Biomedical text mining on big data framework. *PLoS ONE* **2016**, *11*, e0162721. [\[CrossRef\]](#)

47. Loper, E.; Bird, S. NLTK: The natural language toolkit. *arXiv* **2002**, arXiv:cs/0205028.
48. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [[CrossRef](#)]
49. Perkins, J. *Python Text Processing with NLTK 2.0 Cookbook*; Packt Publishing Ltd.: Birmingham, UK, 2010.
50. Lilleberg, J.; Zhu, Y.; Zhang, Y. Support vector machines and word2vec for text classification with semantic features. In Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC), Beijing, China, 6–8 July 2015; pp. 136–140. [[CrossRef](#)]
51. Windmeijer, F.A. Goodness-of-fit measures in binary choice models. *Econom. Rev.* **1995**, *14*, 101–116. [[CrossRef](#)]
52. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240.
53. Darsono, D.; Rohmana, J.A.; Busro, B. Against COVID-19 Pandemic: Bibliometric Assessment of World Scholars' International Publications related to COVID-19. *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia* **2020**, *5*, 75–89. [[CrossRef](#)]
54. Homolak, J.; Kodvanj, I.; Virag, D. Preliminary analysis of COVID-19 academic information patterns: A call for open science in the times of closed borders. *Preprints* **2020**. [[CrossRef](#)]
55. Tran, B.X.; Ha, G.H.; Nguyen, L.H.; Vu, G.T.; Hoang, M.T.; Le, H.T.; Latkin, C.A.; Ho, C.S.; Ho, R.C. Studies of Novel Coronavirus Disease 19 (COVID-19) Pandemic: A Global Analysis of Literature. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4095. [[CrossRef](#)]
56. Hossain, M.M.; Sarwar, S.A.; McKyer, E.L.J.; Ma, P. Applications of Artificial Intelligence Technologies in COVID-19 Research: A Bibliometric Study. *Preprints* **2020**. [[CrossRef](#)]
57. Helmy, Y.A.; Fawzy, M.; Elawad, A.; Sobieh, A.; Kenney, S.P.; Shehata, A.A. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *J. Clin. Med.* **2020**, *9*, 1225. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).