 doi:10.20944/preprints202006.0299.v4

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; lan.umek@fu.uni-lj.si
* Correspondence: aleksander.aristovnik@fu.uni-lj.si

Abstract: The COVID-19 pandemic caused by the novel coronavirus emerged in Wuhan City, Hubei province of China at the end of 2019, has radically transformed the lives of people around the world. Due to its fast spreading, it is currently considered as a worldwide health, social and economic concern. The lack of knowledge on this area has encouraged academic sphere for extensive research, which is reflected in exponentially growing scientific literature in this area. However, current state of COVID-19 research reveals only early development of knowledge, while a comprehensive and in-depth overview remains neglected. Accordingly, the main aim of this paper is to fill the aforementioned gap in the literature and provide an extensive bibliometric analysis of COVID-19 research across science and social science research landscape, using innovative and sophisticated bibliometric approaches (e.g. Venn diagram, Biblioshiny descriptive statistics, VOSviewer co-occurrence network analysis, Jaccard distance cluster analysis, text mining based on logistic regression). The bibliometric analysis is based on the Scopus database including all relevant and latest information on COVID-19 related publications (n=16,866) in the first half of 2020. The empirical results indicate that there is still a lack of publications of COVID-19 and its implications in less-explored (non-health) sciences, especially in social sciences. Accordingly, the findings emphasize an importance of a comprehensive and in-depth approach considering different scientific disciplines in COVID-19 research. The understanding of the evolution of emerging scientific knowledge on COVID-19 is beneficial not only for scientific community but also for evidence-based policymaking in order to prevent and address the COVID-19 pandemic.

Keywords: COVID-19; coronavirus; pandemic; science; social science; bibliometric analysis

---

1. Introduction

Since 2000s, the world has witnessed two large-scale disease outbreaks. These are Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), which emerged in 2003 and 2012, respectively, and caused a worldwide threat that claimed thousands of human lives [1]. In December 2019, a new strain of coronavirus (COVID-19), not previously identified in humans, has emerged in Wuhan City, Hubei province of China. The virus has begun to spread exponentially across all inhabited continents and the number of cases and deaths related to COVID-19 has soon exceeded the numbers of other two coronaviruses (SARS and MERS). Due to the rapid spread of the COVID-19 around the world, the World Health Organization (WHO) declared the COVID-19 a pandemic on 11 March 2020 [2]. The outbreak of the COVID-19 pandemic is a typical public health emergency. Its high infection rate makes it a huge threat to global public health [3-5]. However, its rapid spread has not only affected the lives of many people around the world, but also disrupted the pattern of social and economic development, leading to incalculable social and economic losses [6]. In the last several months from the beginning of the COVID-19 pandemic, around 27 million cases and almost 900,000 deaths have been seen at the global level [7]. International institutions have therefore announced the global economy is now in a recession – as bad or worse than in the global financial crisis of 2009, arguing this recession will affect both developed and developing countries [8,9]. Therefore, it is not surprising, why the COVID-19 pandemic has attracted the attention of the academic sphere and spurred a new wave of research in this area.

The recent bibliometric studies considering broader aspect of coronavirus research over time emphasize that pandemics represent a major medical issue and provide some interesting findings. Taking into account previous coronavirus pandemics Hu et al. [10] establish that the highest research interest occurs in the first year after outburst. This is further confirmed by the study addressing coronavirus research trends during the last 20-years [11,12] and last 50-years period [13,14]. However, although the growth pattern was not uniform, China and the United States have played a major role in the contribution of coronavirus research [15]. Therefore, it is not surprising why recently COVID-19 has become the central topic in the recent scientific literature, since the research addressing various aspects of COVID-19 may be the key to mitigating the current COVID-19 pandemic as well as their consequences [16,17]. The current high-growing interest in COVID-19 and related coronaviruses has even led to the creation of so called the COVID-19 Open Research Dataset (CORD-19), which is a growing resource of scientific papers on COVID-19 and related historical coronavirus research and provides a solid basis for generating new insights in support of the ongoing fight against COVID-19 [18]. The overview of CORD-19 publications reveals that publications are mostly focused on a few and well-defined areas, including coronaviruses (primarily SARS, MERS and COVID-19), 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 and clinical trials). However, the review of latest CORD-19 publications from 2020 indicates a shift from health to other relevant scientific disciplines [19].

In the literature, there exist also several recent bibliographic studies, which are focusing only on COVID-19 research. Interestingly, never before in the history of academic publishing such a great volume of research focused on a single topic has been produced [20]. However, the rush for scientific evidence on the novel COVID-19 can inadvertently encourages dubious publications, which may have been published because the authors were not independent from the practices of the journals in which they appeared [21]. Nevertheless, the recent study on scientific globalism during the COVID-19 pandemic reveals that scientific globalism occurs differently when comparing COVID-19 publications with non-COVID-19 publications. More interestingly, although the COVID-19 pandemic is considered as a worldwide concern and countries indeed increased their proportion of international scientific collaboration during the COVID-19 pandemic, not all countries engaged more globally. The study reveals that countries that have been more affected by the COVID-19 pandemic and those with relatively lower GDPs tended to participate more in scientific globalism than their counterparts [22].

So far, the bibliometric approach examining COVID-19 related issues has been applied on different areas. Namely, some authors address a general overview of COVID-19 research (see Sa'ed and Al-Jabi [23]), while some of them consider a comparative approach, for example a comparison of COVID-19 research between English and Chinese studies (see Fan et al. [24]) or comparison between gender distribution of authors of medical papers related to the COVID-19 pandemic (see Andersen et al. [25]). Moreover, some of the existing bibliometric studies consider only a single country in their analysis (for Indian case see Vasantha Raju and Patil [26]) and some of them are focused on top cited COVID-19 publications (see ElHawary et al. [27]). Finally, some of the previous COVID-19 bibliometric studies provide in-depth analysis in the field 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 the existing COVID-19 bibliometric studies reveal that China and the United States have the largest COVID-19 scientific production [31-35]. The most relevant institutions involved in COVID-19 research are Huazhong University of Science and Technology, Wuhan University and University of Hong Kong. Moreover, the majority of published documents on COVID-19 are published in prestigious journals with high impact factors, including the Lancet, BMJ – Clinical Research Ed. and Journal of Medical Virology [31,35]. Furthermore, according to 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 the current research hotspots of COVID-19 [33,34].

Although, the absence of knowledge on the novel COVID-19 has grabbed the attention of the academic sphere, spurring a new wave of research into the virus [36], yet, the vast majority of recent studies chiefly consider health-related issues, leaving other aspects neglected, as indicated by the latest literature [31-35]. Moreover, COVID-19 research's current status is only of the early development of knowledge. Therefore, the literature stresses that greater research should be conducted in less-explored areas, including life, physical and social sciences & humanities [33]. Accordingly, the main aim of this paper is to provide an extensive bibliometric analysis on COVID-19 research in first half of 2020. Although there already exist several papers addressing bibliometric analysis of COVID-19 research, several research gaps are identified, which are carefully tackled by this paper. First, the existing bibliometric studies are predominantly focused on general analysis of COVID-19 research, showing the importance of health sciences in this area, while detailed insight considering different research landscapes remain neglected. Therefore, this paper provides in-depth bibliometric analysis by considering various science and social science research landscapes or subject areas, including corresponding subject area classifications and research fields. Second, the predominant part of the existing COVID-19 bibliometric studies is mostly addressing databases containing document information only. Accordingly, this paper extends the analysis on a comprehensive database including document and source information, allowing the bibliometric analysis in different research landscapes. Finally, recent COVID-19 research is neglecting the overlap across scientific disciplines as well as lacking innovative bibliometric approaches. Therefore, in addition to well-established approaches, this paper utilizes a wide range of innovative and sophisticated bibliometric approaches, including descriptive analysis, network analysis, cluster analysis based on the Jaccard distance and text mining based on logistic regression. These also allow showing all possible logical relations between different scientific disciplines.

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

2. Materials and Methods

A comprehensive bibliometric data on COVID-19 related research is obtained throughout two consecutive phases as presented in Figure 1. The first phase involves identification of all relevant documents or publications from January 1, 2020 to July 1, 2020 in the Scopus database on document information, which is widely recognised database also by the previous research [10,14,31,35]. The applied search query extends previously narrowly defined queries [33,34] by including a wide 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. Additionally, the search period was set to include documents published between January 1, 2020 and July 1, 2020. Finally, only documents in English language were considered for the review process. According to the presented search query, a total of 21.400 documents are identified as relevant in COVID-19 research. Interestingly, the number of documents obtained by using identical search query increased for 59.6%, as on June 1, 2020 the same search resulted to 13,480 documents. This implies that interest COVID-19 research is growing exponentially. The second phase involves supplementing the presented Scopus database on document information with Scopus CiteScore metrics containing source related information (e.g. citations, rankings, SNIP, etc.). The merging process revealed that some of the documents from Scopus had no match in Scopus CiteScore metrics (n=4,534), meaning that they are not considered in the bibliometric analysis. Accordingly, the screening process resulted in a unique database of 16,866 documents. The data preparation process, i.e. obtaining, merging and cleaning the relevant data is facilitated by Python programming language using the Pandas and Numpy libraries [37].

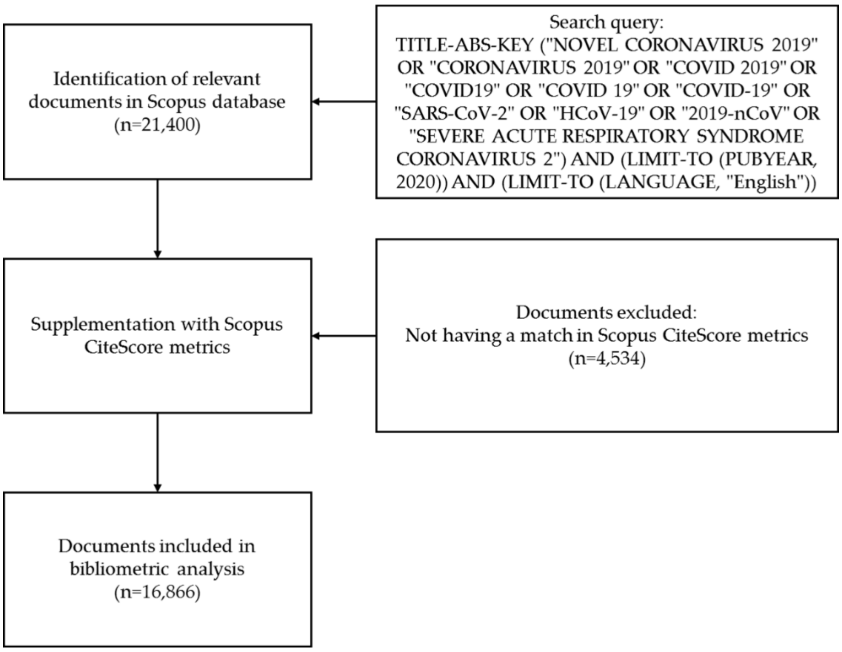


Figure 1. Flowchart of database determination (January-June 2020).
Source: Authors' elaboration based on Scopus database, July 2020.

Then, an in-depth bibliometric analysis followed, allowing for a sophisticated and innovative approach to literature review. Namely, the structured literature review represents a traditional approach to analyse and review scientific literature, providing an in-depth overview of the content. However, this approach suffers from several limitations related to subjective factors, time-consumption and efficiency. The application of modern bibliometric approaches reduces the aforementioned limitations and provide an effective way to handle extensive collections of scientific literature [38]. So far, the existing bibliometric studies on COVID-19 research applied some 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]). However, the existing bibliometric studies have been hardly neglecting the fact that scientific disciplines overlap strongly, resulting in similar findings and conclusions in these studies and lack of knowledge in less-explored areas [33]. Therefore, in order to supplement the existing research and assess the state of current COVID-19 research across different research landscapes (health sciences, life sciences, physical sciences and social sciences & humanities), innovative and sophisticated bibliometric approaches are utilised in this paper. The bibliometric analysis is performed by considering the Scopus hierarchical classification of documents based on All Science Journal Classification scheme (ASJC) and in-house experts' opinions. Accordingly, the documents are classified into three hierarchically arranged groups, namely: 1) subject area categories; 2) subject area classifications and 3) fields.

On this basis, the following bibliometric approaches are applied. First, for descriptive analysis, including Venn diagram for detecting overlap of scientific disciplines, Biblioshiny application [40] and Python library Pyvenn [41] are used. Second, in order to depict relations among keywords and fields, co-occurrence network analysis is made with VOSviewer, a software tool for constructing and visualizing bibliometric networks [42]. Furthermore, for the purposes of examining relationships between different subject area classifications within COVID-19 research, a cluster analysis based on the Jaccard distance (JD) (Jaccard index subtracted from 1), is performed. 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 clustermap is designed by using Python's most powerful visualisation libraries, i.e. Matplotlib and Seaborn [41, 45].

Finally, in order to predict document's subject area based on its abstract, text mining-based classification is used [46]. For this purposes, binary logistic regression is selected as a prediction model. Accordingly, four different binary logistic models are tested – each for individual subject area, with binary dependent variable, having value of 1 if document belongs to individual subject area and 0 if document belongs to other remaining subject areas. Based on the results of fitting model to the data, binary logistic regression provides us also with information which words are most characteristic for a particular subject area (which discriminate most between two subject areas). This approach requires documents, having a full abstract. Text mining is 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). Then, WordNet lemmatization is applied [48], the set of extracted words is further filtered with list from nltk.corpus and manually added stop words [49]. For construction of features (bag of words) "term frequency–inverse document frequency (tf-ifd)" method is used. The class TfidfVectorizer from sklearn.feature_extraction.text [50] is used with following parameters: sublinear tf scaling, smooth idf weights, utf-8 encoding, l2 norm regularization, min data frequency = 1, max data frequency = 10. For extraction of new features for classification the search for unigrams (single words) and bigrams (sequence of two words) is performed. Top 100 features are created and are further used as predictors (independent variables) in binary logistic model.

3. Results

An overview of scientific documents utilised in this study is presented in Table 1. A total of 16,866 documents written by 66,504 different authors and published in 2,548 journals were utilised in this study, whereby 7,422 (44.0%) of them 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 and the average authors per document were 3.94. A major proportion of the documents were articles (41.5%) and letters (26.5%). Much smaller proportion of the documents were reviews (10.2%), editorials (10.1%) and notes (9.4%). Finally, there was a negligible proportion of other documents (2.4%) such as short surveys, conference papers, errata and data papers. The presented characteristics of scientific documents on COVID-19 research are predominantly 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	2,548
Total citations	100,683
Cited documents	7,422
Average citations	13.57
Average authors	3.94
Document type	**Number (share)**
Article	6,998 (41.5%)
Letter	4,467 (26.5%)
Review	1,713 (10.2%)
Editorial	1,698 (10.1%)
Note	1,593 (9.4%)
Other	397 (2.4%)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

Source: Authors' elaboration based on Scopus database, July 2020.

The Scopus provides hierarchical classification of documents by considering ASJC (All Science Journal Classification scheme) and in-house experts' opinions. Accordingly, the documents are classified into three hierarchically arranged groups, namely: 1) subject area categories; 2) subject area classifications and 3) fields. The distribution of documents according to the mentioned groups is presented in Table 2.

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 Pharmaceutics (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 & 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 across subject areas, classifications and fields.
Source: Authors' elaboration based on Scopus database, July 2020.

It is evident that nearly two-thirds of documents are subject to health sciences (65.2%), with medicine (91.0%) being the most exposed, whereby the predominant focus is being put on infectious diseases (10.2%) and general medicine (9.7%). This is in line with previous bibliometric studies, emphasizing that COVID-19 research is the main domain for health-related sciences [31-35]. A much smaller number of documents is the subject of 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 recognised as the most important research fields within the life sciences. The smallest share of documents is found in physical sciences (7.5%). These are focused predominantly 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 subject to social sciences & humanities (8.3%). Nevertheless, social sciences (44.2%) and psychology (24.6%) are recognised 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 & humanities. The aforementioned confirms existing claims on the lack of knowledge in less-explored areas, including life, physical and social sciences [33]. Therefore, it is not surprising why there exist many calls for more extensive COVID-19 research in less-explored scientific disciplines.

Table 3 presents 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. As regards different scientific disciplines or subject areas (classifications), the most relevant journals are predominantly subject to 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, the smaller part of most relevant journals are subject to life sciences (immunology and microbiology as well as neuroscience) with the focus on biological psychiatry and virology. Some of these journals are also subject to 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 is subject to social sciences (psychology) covering the research field of general psychology. There is also one journal, which is classified as multidisciplinary. Most of these journals are ranked into the first quartile (Q1) and have a relatively high source normalized impact per paper (SNIP), which is in line with the existing research [31,35]. Furthermore, most of these journals are from Anglo-Saxon countries such as the United Kingdom, the Netherlands and the United States. Similar findings are provided also by previous COVID-19 bibliometric studies [33,34]. However, all of the existing bibliometric studies are neglecting a high overlap across scientific disciplines, which leads to biased results and consequently to a lack of comprehensive understanding of the COVID-19 research across different scientific disciplines [33].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

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
Journal of Medical Virology	293	3,657	Life Sciences (Immunology and Microbiology) Health Sciences (Medicine)	Virology (37/66, Q3) Infectious Diseases (108/283, Q2)	0.780	US
The BMJ	261	1,358	Health Sciences (Medicine)	General Medicine (21/529, Q1)	3.999	UK
The Lancet	239	13,755	Health Sciences (Medicine)	General Medicine (1/529, Q1)	21.313	UK
Medical Hypotheses	227	107	Health Sciences (Medicine)	General Medicine (99/529, Q1)	0.509	US
Science of the Total Environment	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	NL
International Journal of Environmental Research and Public Health	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	CH
Journal of Infection	155	1,049	Health Sciences (Medicine)	Microbiology (medical) (13/115, Q1) Infectious Diseases (21/238, Q1)	1.587	UK
International Journal of Infectious Diseases	148	1,503	Health Sciences (Medicine)	Microbiology (medical) (26/115, Q1) Infectious Diseases (59/283, Q1)	1.426	NL
Psychiatry Research	130	314	Health Sciences (Medicine) Life Sciences (Neuroscience)	Psychiatry and Mental Health (154/506, Q2) Biological Psychiatry (25/38, Q3)	0.968	IE
Journal of Clinical Virology	120	239	Life Sciences (Immunology and Microbiology) Health Sciences (Medicine)	Virology (19/66, Q2) Infectious Diseases (44/283, Q1)	1.238	NL
Diabetes and Metabolic Syndrome: Clinical Research and Reviews	119	462	Health Sciences (Medicine)	Internal Medicine (75/128, Q3) Endocrinology, Diabetes and Metabolism (135/217, Q3)	0.982	NL
Infection Control and Hospital Epidemiology	118	172	Health Sciences (Medicine)	Microbiology (medical) (39/115, Q2) Epidemiology (40/93, Q2) Infectious Diseases (91/283, Q2)	1.358	UK
Travel Medicine and Infectious Disease	113	621	Health Sciences (Medicine)	Public Health, Environmental and Occupational Health (73/516, Q1) Infectious Diseases (82/283, Q2)	1.184	NL
Critical Care	112	244	Health Sciences (Medicine)	Critical Care and Intensive Care Medicine (4/81, Q1)	2.508	UK
The Lancet Infectious Diseases	111	2,280	Health Sciences (Medicine)	Infectious Diseases (4/283, Q1)	7.234	UK
New England Journal of Medicine	106	11,768	Health Sciences (Medicine)	General Medicine (2/529, Q1)	13.212	US
Asian Journal of Psychiatry	101	433	Health Sciences (Medicine) Social Sciences & Humanities (Psychology)	Psychiatry and Mental Health (217/506, Q2) General Psychology (71/204, Q2)	1.022	NL
Dermatologic Therapy	100	153	Health Sciences (Medicine)	Dermatology (74/123, Q3)	0.883	UK
Chaos, Solitons and Fractals	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
Science	97	1,918	Multidisciplinary (Multidisciplinary)	Multidisciplinary (2/111, Q1)	7.521	US

Source: Authors' elaboration based on Scopus database, July 2020.

3.1. Bibliometric analysis across different subject area categories

According to the Scopus classification, documents can be classified into four different subject areas, namely: health sciences, life sciences, physical sciences and social sciences & humanities. However, these subject areas are strongly intersecting, meaning that individual document can be classified in several subject areas at the same time. Accordingly, for the purposes of addressing the comprehensiveness of COVID-19 research, Figure 2 shows the Venn diagram of the presented subject areas and all the possible sets that can be made from them. This also makes it possible to determine the so-called pure sciences, covering only those documents belonging exclusively to just one subject area (without intersecting with other subject areas). According to the number of documents obtained on July 1, 2020 (June 1, 2020), health sciences contain a total of 14,187 (8,896) documents of which 10,394 (6,575) documents are identified as pure health sciences. Further, life sciences encompass a total of 4,143 (2,549) documents of which 928 (599) documents are considered as pure life sciences. Moreover, physical sciences include a total of 1,625 (878) documents of which 568 (314) documents belongs to pure physical sciences.

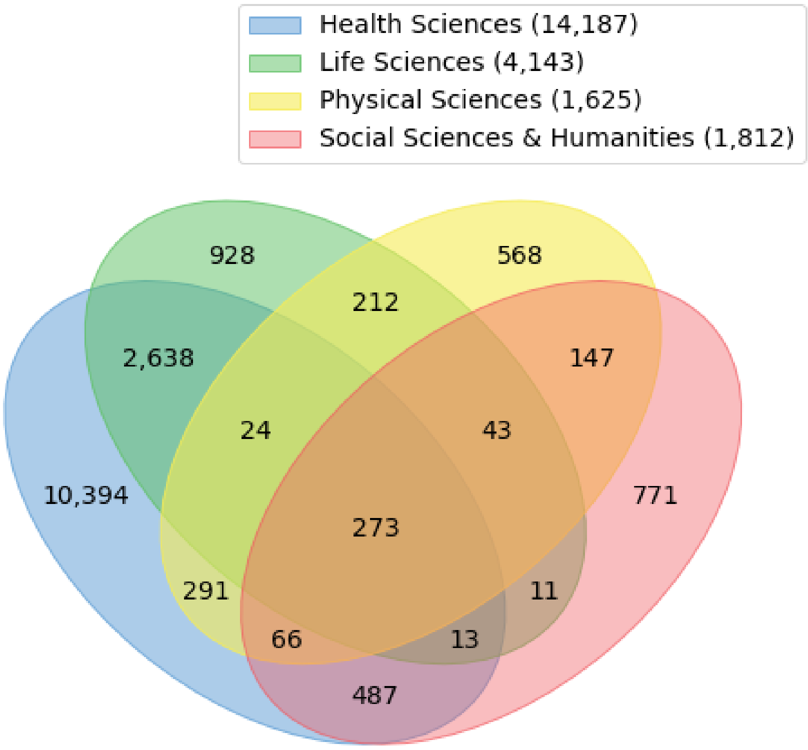


Figure 2. Venn diagram by number of documents on COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (all documents included), July 2020.

Finally, social sciences & humanities cover a total of 1,812 (977) documents of which 771 (323) are determined as pure social sciences & humanities. A comparison between different subject areas reveals that health sciences are the most relevant in COVID-19 research, while the second most relevant subject area is represented by life sciences. Moreover, physical sciences and social sciences & humanities seem to be the least popular so far, as also suggested by previous research [33]. However, according to the growth of number of documents in June 2020, social sciences seem to be the most growing scientific discipline, as the total number of documents in this subject area increased by 85.5% and even by 138.7% in pure social sciences. This is consistent with the expectations as well as with 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 COVID-19 pandemic is the protection of public health, while the real socio-economic consequences occur later.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

This path is also revealed by the recent scientific literature on COVID-19 published in the first half of 2020 and the review of the latest CORD-19 publications from 2020, indicating a shift from health to other relevant scientific disciplines [19]. Finally, some of the documents (273) are considered as multidisciplinary, making impossible to include them in the further bibliometric analysis.

Figure 3 presents most relevant countries of COVID-19 research by subject area. It shows the top 5 countries, providing the largest number of documents of corresponding author. The most relevant country is the United States, significantly dominating in all scientific disciplines, except in physical sciences, where it is ranked on the second place. In addition to the United States, which significantly outperform other countries, also China and Italy dominate in COVID-19 research as they are among top 3 countries in all scientific disciplines, except in social sciences, where Italy is replaced by India. These findings are consistent with the existing bibliometric studies (which do not consider scientific disciplines separately), arguing that the United States and China have world-leading position in COVID-19 research [31-35].

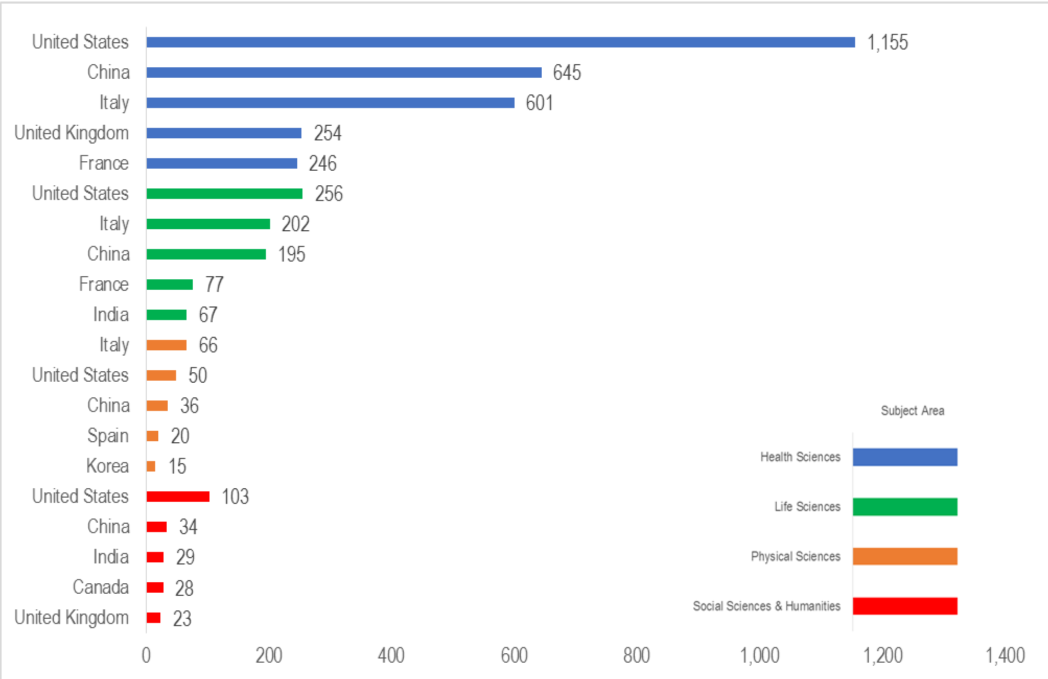


Figure 3. Most relevant countries by number of documents in COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (only documents with at least one citation included), July 2020.

Figure 4 shows most relevant institutions by number of documents in COVID-19 research across subject areas. Due to the strong overlap between individual scientific disciplines, they to some extent can share the same most relevant institutions. The most involved institution is Huazhong University of Science and Technology, providing a significantly higher number of documents in health sciences (n=1,380) and life sciences (n=420). Besides, Zhongnan Hospital of Wuhan University and Icahn School of Medicine at Mount Sinai also play an important role in these two scientific disciplines. Moreover, Fudan University is dominating in physical sciences (n=68), while providing an enviable number of publications also in life sciences (n=155). Finally, California Department of Public Health and Public Health – Seattle and King County are the most relevant institutions in social sciences & humanities, having an important role also in physical sciences. The findings are to some extend comparable with the existing bibliometric studies on COVID-19 research [33,35].

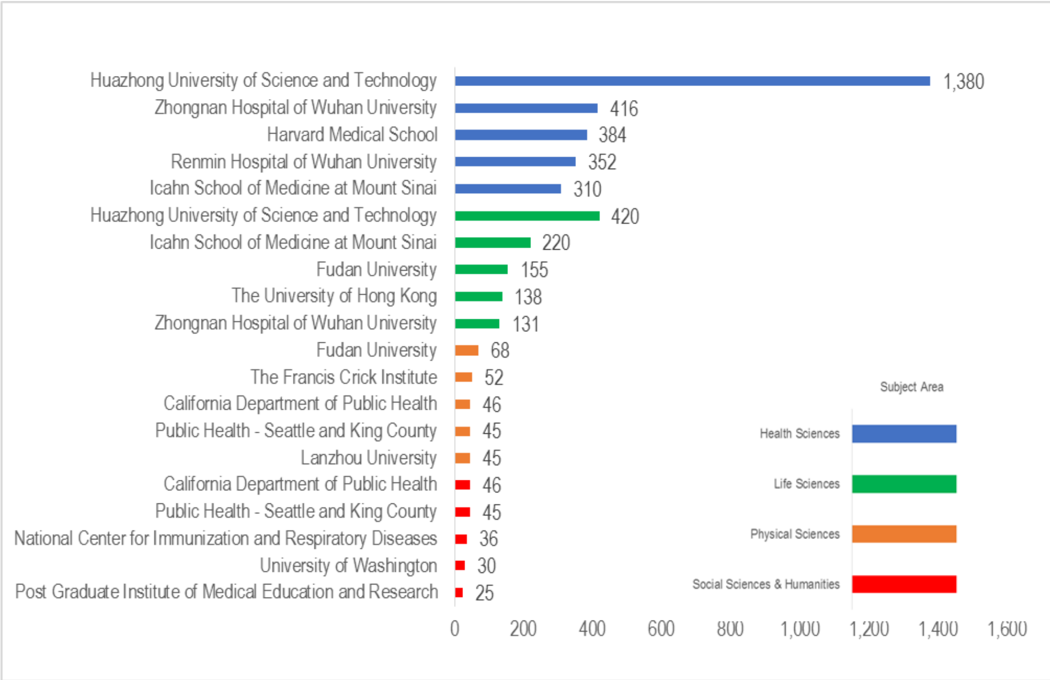


Figure 4. Most relevant institutions by number of documents in COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (only documents with at least one citation included), July 2020.

Figure 5 presents the most relevant journals in COVID-19 research across subject areas. It presents the number of documents provided in certain journal within individual subject area. In health sciences, Journal of Medical Virology has the most documents (n=293), which is followed by the BMJ (n=261), the Lancet (n=239), Medical Hypotheses (n=227), International Journal of Environmental Research and Public Health (n=155). These findings are to some extend 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 life sciences, due to strong interweaving with health sciences, the most relevant journal is also Journal of Medical Virology, having the most documents (n=293), which is followed by Psychiatry Research (n=130), Journal of Clinical Virology (n=120), Brain, Behaviour and Immunity (n=77) and Pharmacological Research (n=63). In 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 social sciences & 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).

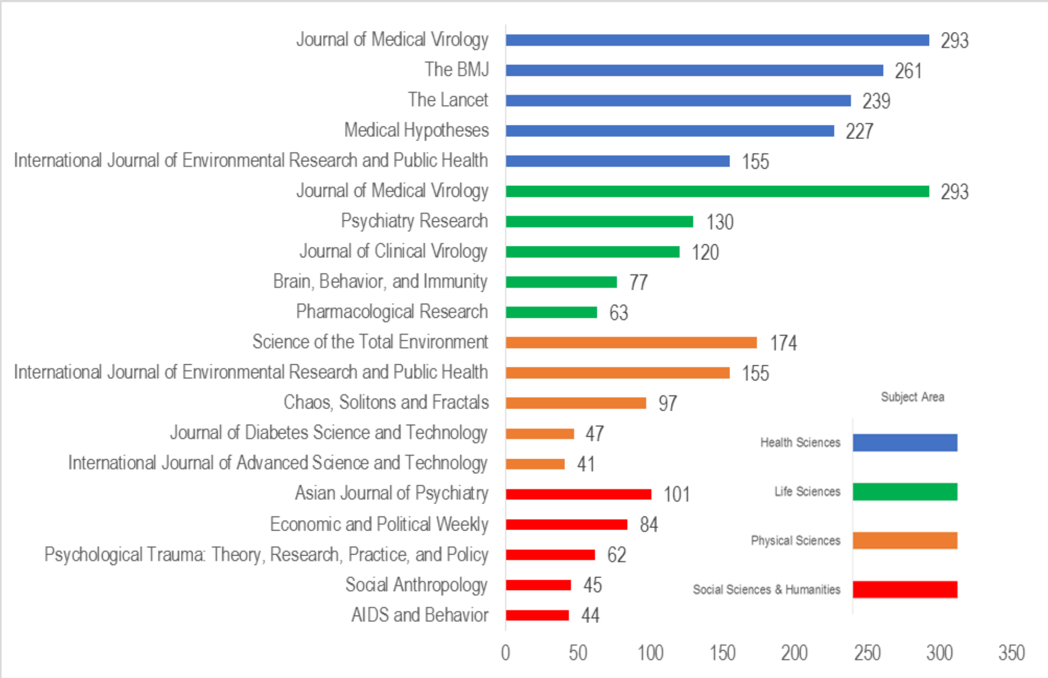


Figure 5. Most relevant journals by number of documents in COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (all documents included), July 2020.

Figure 6 shows 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 as they are among top 5 cited authors in all four scientific disciplines. This finding is different according to the existing bibliometric studies, presumably due to different criteria applied [33].

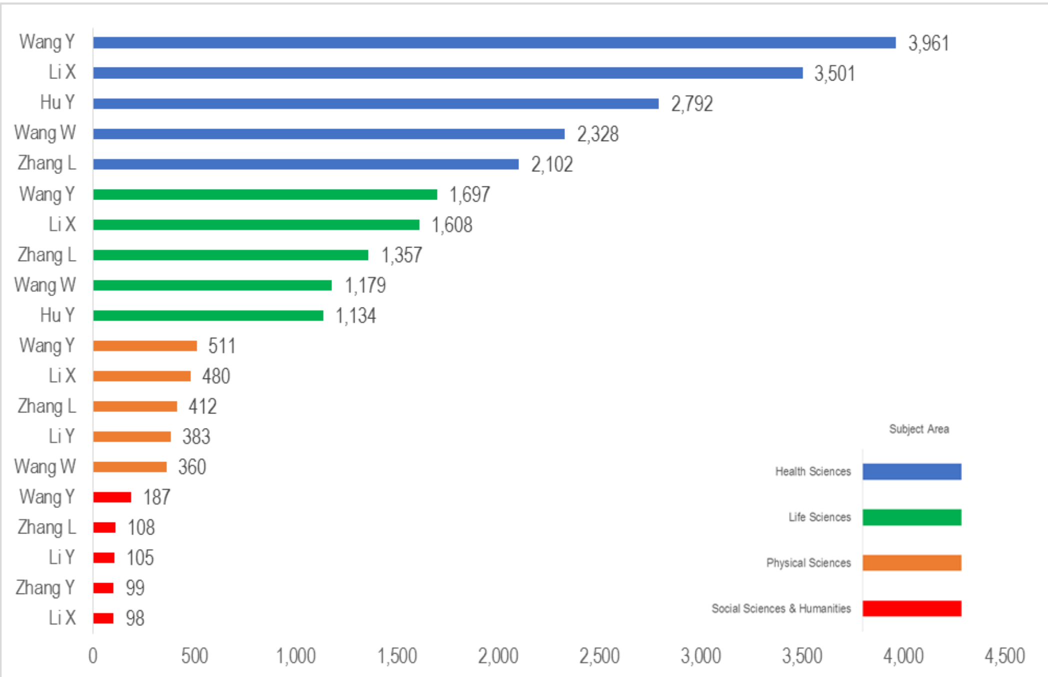
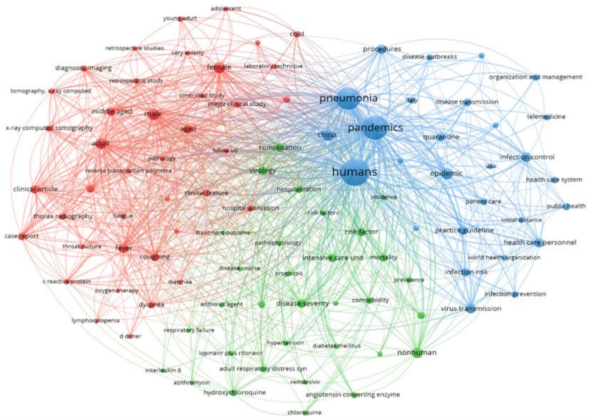


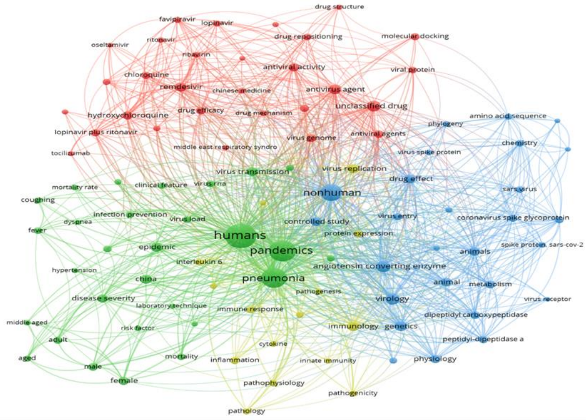
Figure 6. Most relevant authors by number of citations in COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (total citations included), July 2020.

 doi:10.20944/preprints202006.0299.v4

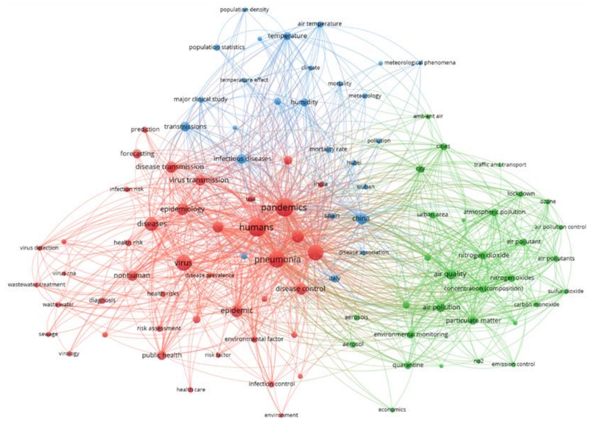
Figure 7 presents the keyword co-occurrence network for (a) health sciences, (b) life sciences, (c) physical sciences and (d) social sciences & humanities separately. In order to ensure greater distinction between individual subject areas, only pure sciences (without intersecting with other sciences)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



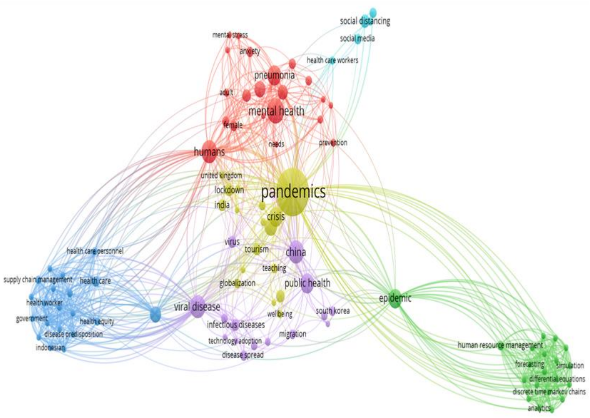
(a) Health Sciences



(b) Life Sciences



(c) Physical Sciences



(d) Social Sciences & Humanities

Figure 7. Keyword co-occurrence network in COVID-19 research across subject areas (January-June 2020).
Source: Authors' elaboration based on Scopus database (only documents on pure sciences included), July 2020.

The bibliometric analysis (keyword co-occurrence) reveals that research hotspots differ according to subject area. For health sciences, 3 clusters are identified, addressing the following topics: 1) pandemics; 2) risk factors and symptoms: and 3) mortality. Accordingly, health sciences deal predominantly with health related issues related to the COVID-19 pandemic. Next, in the life sciences, 3 clusters are found, which are dealing with: 1) pandemics; 2) virology; and 3) drug efficiency. The focus of life sciences seems to be more oriented towards the knowledge about spreading of the virus and ways how to prevent efficiently the disease with appropriate drugs. This corresponds to the findings from other recent bibliometric studies on COVID-19 research, emphasizing predominantly health related issues [33,34]. Moreover, the results for less-explored subject areas show the following. As regards physical sciences, 3 clusters are recognised, which are related to: 1) pandemics; 2) China and disease transmission and 3) air pollution. Physical sciences are focused on knowledge related to how fast the COVID-19 pandemic is spreading and environmental related issues. Finally, in social sciences & humanities, 6 clusters are identified, addressing the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

following topics: 1) pandemics; 2) epidemics; 3) viral disease and China; 4) respiratory disease; 5) social distancing; and 6) mental health. The detailed synopsis of the research hotspots, including top 10 keywords, related to COVID-19 in individual scientific discipline is presented in Table A1 in Appendix A.

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

The results of text mining-based classification (see Table A2 in Appendix A) show the following. The goodness-of-fit statistics for all of the estimated binary logistic models is proved to be adequate, as suggested by Pseudo R^2 value ranging from minimum 0.146 (health sciences) and maximum 0.403 (social sciences & humanities) and very low values of Log-Likelihood Ratio (LLR) P-value (<0.001) [51]. Moreover, also evaluation measures of models (AUC, CA, Precision and Recall) suggest very good discrimination (ability to classify documents belonging to individual subject area and documents belonging to other remaining subject areas) [52]. Table 4 presents the summary of the results of text mining-based classification of COVID-19 documents across subject areas. It shows the most discriminant words, (having significant and positive regression coefficient) for predicting corresponding subject area based on binary logistic regression. For health sciences, the top 3 most characteristic words are "patient", "health" and "healthcare". The regression coefficient for "patient" suggests that if a tf-idf of a word »patient« in a document increases by amount of t a probability of this document belonging to health sciences increases by exp(4,775). The same interpretation holds also for all of the regression coefficients. As regards other scientific areas, the top 3 most characteristic words are "protein", "human" and "vaccine" for life sciences, "factor", "lockdown" and "area" for physical sciences and "crisis", "pandemic" and "mental" for social sciences & humanities.

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

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

Note: Words in italics are identified as the most discriminant in more than one subject area.
Source: Authors' elaboration based on Scopus database (only documents with full abstract included), July 2020.

3.2. Bibliometric analysis across different subject area classifications and fields

In order to examine relationships between different subject area classifications within COVID-19 research, a cluster analysis based on the Jaccard distance (JD) (Jaccard index subtracted from 1), measuring dissimilarity is performed (see Figure 8). Jaccard distance ranges from 0 to 1, with 0 suggesting perfect overlap and 1 indicating no overlap [43]. Based on the results, the following clusters can be identified. The first and relatively pronounced cluster is engineering, bringing together: computer science, energy, materials science, chemistry, chemical engineering and engineering. A strong connection between these subject area classifications is further confirmed by relatively low Jaccard distance. This is reflected especially between engineering and chemical engineering (JD=0.69), meaning that 31% (1-0.69) COVID-19 related documents, belonging either to engineering or either to chemical engineering belong to both subject area classifications at the same time. One of the strongest (23%) overlap in this cluster can be found also for chemical engineering and chemistry. The second and most pronounced cluster concerns mathematics and physics, which is also suggested by the lowest Jaccard distance between mathematics and physics and astronomy (JD=0.58), meaning that there is a 42% overlap between these two subject area classifications.

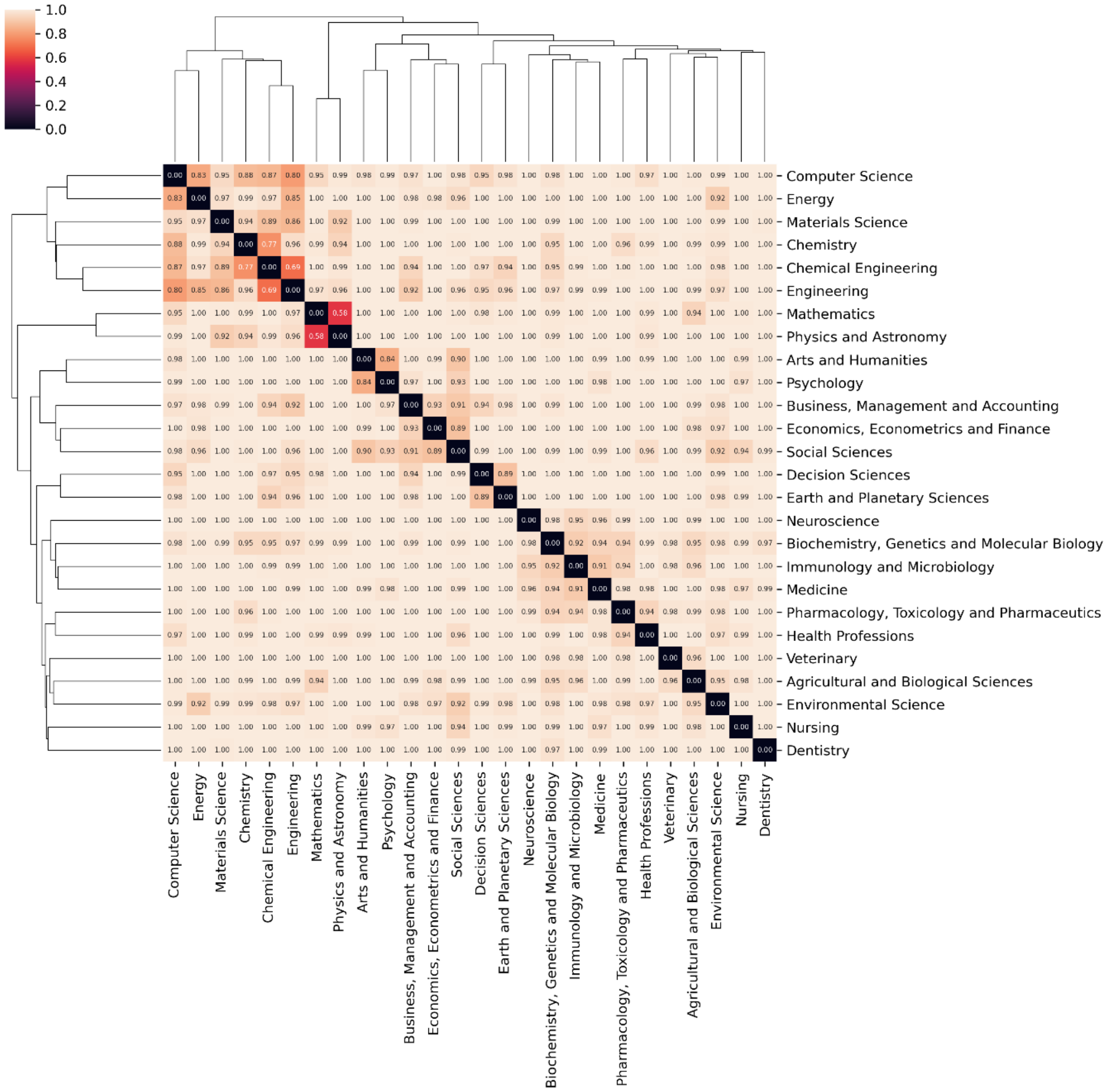
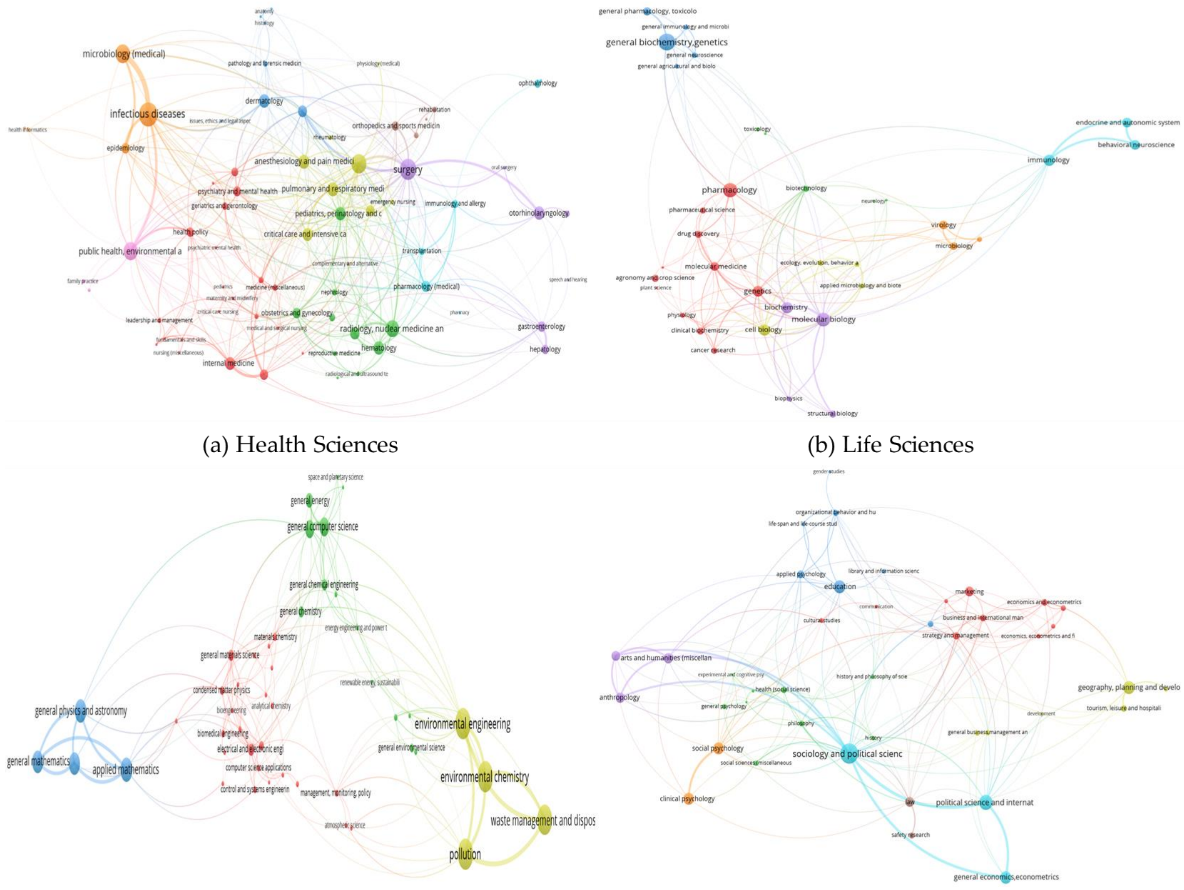


Figure 8. Clustermap of COVID-19 research based on Jaccard dissimilarities between subject area classifications (January-June 2020).
Source: Authors' elaboration based on Scopus database, July 2020.

Moreover, according to the results, the other subject area classifications are not very different from each other (Jaccard distance is equal or very close to 1), making difficult to identify meaningful or homogeneous clusters. Nevertheless, some further potential or emerging clusters can be identified. Accordingly, the third cluster is humanities and psychology, grouping individual subject area classifications of arts and humanities and psychology with 16% overlap. The fourth cluster is business, management and economics, covering business, management and accounting, economics, econometrics and finance and social sciences, whereby the most connected subject area classifications are social sciences and economics, econometrics and finance with 11% overlap and social sciences and business, management and accounting with 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 11% overlap. Finally, the sixth cluster concerns health and environment, covering neuroscience, biochemistry, genetics and molecular biology, immunology and microbiology, medicine, pharmacology, toxicology and pharmaceutics, health professions, veterinary, agricultural and biological sciences, environmental science, nursing and dentistry. The highest overlap in this cluster is identified especially between medicine and immunology and microbiology (9%) and immunology and microbiology and biochemistry, genetics and molecular biology (8%).

Regarding the overlap of COVID-19 research between different subject area classifications outside 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 social sciences and psychology (7%) as well as the connection between agricultural and biological sciences and mathematics and decision sciences and business, management and accounting (6%). The presented results provide additional evidence on COVID-19 research collaboration within and between different subject area classifications [22].

Figure 9 presents the field co-occurrence network for (a) health sciences, (b) life sciences, (c) physical sciences and (d) social sciences & humanities separately. In order to ensure 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 297 research fields, which are distributed among these four main subject areas. The bibliometric analysis (field co-occurrence) reveals different clusters related to COVID-19 within individual subject area. For health sciences, 9 clusters are identified, namely: 1) internal medicine; 2) radiology and haematology; 3) dermatology and neurology; 4) cardiology, pulmonary and anaesthesiology; 5) surgery; 6) pharmacology; 7) epidemiology; 8) sports medicine and rehabilitation; and 9) public health. Next, in the life sciences, 7 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. As regards physical sciences, 4 clusters are recognised, which are 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 social sciences & humanities, 8 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; 7) social and clinical psychology; and law and safety. The detailed synopsis of the clusters, including top 5 fields, related to COVID-19 in individual scientific discipline is presented in Table A3 in Appendix A.



(a) Health Sciences

(b) Life Sciences

(c) Physical Sciences

(d) Social Sciences & Humanities

Figure 9. Field co-occurrence network in COVID-19 research by subject area (January-June 2020).
Source: Authors' elaboration based on Scopus database, July 2020.

4. Discussion and Conclusion

The outbreak of COVID-19 is a typical public health emergency, which due to its high infection rate makes it a huge threat not only to global public health but also to economic and social development. In order to be able to solve such kind of emergencies, it is necessary to fully understand the problem, its implications for different areas as well as the solutions that may be effective and efficient in addressing potential devastating consequences. Therefore, the scientific knowledge on COVID-19 is very important as it facilitates answering real-life questions. However, the extent of the current COVID-19 pandemic calls for in-depth knowledge allowing identification of numerous issues in different areas. Therefore, it is not surprising, why there is an unprecedented increase in the COVID-19 research since the pandemic started [36,53]. The COVID-19 pandemic resulted in generation of large amount of scientific publications, which can consequently present potential problems regarding the information velocity, availability, and scientific collaboration, especially in the early stages of the pandemic [54]. The current state of COVID-19 research, therefore, needs a comprehensive analysis to help guide an agenda for further research, especially from the perspective of cooperation between different scientific disciplines in different stages of pandemic prevention and control, by applying innovative and sophisticated scientific approaches [55,56].

Accordingly, this paper provides an extensive bibliometric analysis of COVID-19 research across science and social science research landscape by using a wide variety of different bibliometric approaches, including descriptive analysis, network analysis, cluster analysis based on the Jaccard distance and text mining based on logistic regression. In general, the results show that a total of 21,400 documents related to COVID-19 research were published in Scopus database in the first half of 2020. Interestingly, the number of the documents has increased by 59.6% in June 2020, suggesting exponential interest in COVID-19 research. The database suitable for the review process includes a total of 16,866 documents. They were written by 66,504 different authors, published in different 2,548 journals and together provide a total of 100,683 citations. A major proportion of the documents were articles (41.5%) and letters (26.5%), which is in line with previous bibliometric studies [32,23]. Moreover, the distribution of COVID-19 related documents according to the Scopus hierarchical classification reveals that nearly two-thirds (65.2%) of documents are subject to health sciences, confirming existing claims in COVID-19 research on the lack of knowledge in less-explored subject areas, including life, physical and social sciences & humanities [33]. Furthermore, the most relevant journals in COVID-19 research cover almost one-fifth (17.6%) of total documents and a significant share (41.3%) of total citations. As regards different scientific disciplines or subject areas (classifications), the most relevant journals are predominantly subject to health sciences (medicine), while other scientific disciplines (life sciences, physical sciences and social sciences & humanities) remain in the background. Most of these journals are ranked into the first quartile (Q1) and have a relatively high source normalized impact per paper (SNIP), which is in line with the existing research [31,35]. Finally, most of these journals are from Anglo-Saxon countries such as the United Kingdom, the Netherlands and the United States. Similar findings are provided also by previous COVID-19 bibliometric studies [33,34].

A more detailed comparison of COVID-19 research between four scientific disciplines reveals that subject areas are strongly intersecting, which calls for an in-depth analysis of individual subject area separately. The results of bibliometric analysis across different subject area categories show the following. According to the number of documents health science 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 & humanities seem to be the least popular so far. However, during the June 2020 social sciences seem to be the most growing scientific discipline, as the total number of documents in this subject area increased by 85.5% and even by 138.7% in pure social sciences. A shift from health to other relevant scientific disciplines can be observed in the review of the latest CORD-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 suggest that the United States significantly dominates in all scientific disciplines, except in physical sciences. Besides the United States, which significantly outperform other countries, also China and Italy dominate in COVID-19 research. As regards the most relevant institutions, Huazhong University of Science and Technology, Zhongnan Hospital of Wuhan University and Icahn School of Medicine at Mount Sinai play an important role in health sciences and life sciences. Moreover, Fudan University is dominating in physical sciences, while having a crucial role also in life sciences. Finally, California Department of Public Health and Public Health – Seattle and King County are the most relevant institutions in social sciences & humanities, while having an important role also in physical sciences. The results regarding journals reveal that Journal of Medical Virology is the most relevant journal for health sciences and life sciences, Science of the Total Environment for physical sciences and Asian Journal of Psychiatry for social sciences & humanities. As regards most important authors, Wang, Y. (China-Japan Friendship Hospital, Beijing, China) and Li, X. (Clinical and Research Center of Infectious Diseases, Beijing, China) are the most important authors involved in COVID-19 research. The presented results are to some extent comparable with the previous bibliometric studies on COVID-19 research [31-35]. Moreover, the results of keyword co-occurrence analysis by main subject areas reveal different research hotspots for individual scientific disciplines, with a common point of pandemics. Health sciences are more focused on health consequences (see Hossain [33] and Lou et al. [34]), while life sciences are more oriented towards drug efficiency. Finally, physical sciences are more focused on environmental consequences, while social sciences are more oriented towards socio-economic consequences. Furthermore, the results of text mining-based classification based on binary logistic regression reveal the most characteristic words for predicting corresponding area. For health sciences, the top 3 most characteristic words are “patient”, “health” and “healthcare”. As regards other scientific areas, the top 3 most characteristic words are “protein”, “human” and “vaccine” for life sciences, “factor”, “lockdown” and “area” for physical sciences and “crisis”, “pandemic” and “mental” for social sciences & humanities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

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 6 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 between different subject area classifications outside 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. These results provide additional evidence on COVID-19 research collaboration within and between different subject area classifications [22]. Moreover, the results of field co-occurrence analysis by main subject areas reveal different research clusters of fields for individual scientific disciplines, providing in-depth segmentation of different scientific disciplines.

Several limitations of the present study should also be noted. First, the bibliometric analysis is based COVID-19 related documents retrieved from the Scopus database only. Although the Scopus is considered as one of the largest abstract and citation database of peer-reviewed literature, it may not cover a complete collection of COVID-19 research. Therefore, the inclusion of other databases, especially a growing body of preprints available in the Google Scholar database, could have provided additional insights that are not available in this study. Second, this study is based on short period of time (first half of 2020). Although this limitation cannot be solved so far, a repeated study with extended period would give additional time-dimensional insights. This would be beneficial also in terms of achieving higher number of publications in some under-represented disciplines, especially social sciences & humanities. Another limitation is that only titles, abstracts and keywords in English language are included in this study, which might cause publication bias to some extent. Further studies may therefore address this issue. Finally, another limitation of this study is a lack of citations and collaborations networks by using sophisticated methodological approaches due to a low number of studies and continuously changing citations metrics. Accordingly, future bibliometric studies should address these limitations and further examine the evolution of scientific knowledge on 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 considering different scientific disciplines in COVID-19 research. In order to address the economic, socio-cultural, political, environmental and other (non-medical) consequences of the current COVID-19 pandemic, COVID-19 must be higher on research agenda of non-health sciences, in particular social sciences, in the near future. Namely, the understanding of the evolution of emerging scientific knowledge on COVID-19 is beneficial not only for scientific community but also for evidence-based policymaking in order to prevent and address the COVID-19 pandemic implications.

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

Acknowledgments: The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P5-0093).

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

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
	Drug Efficiency	Unclassified Drug, Antivirus Agent, Remdesivir, Hydroxychloroquine, Antiviral Activity, Antiviral Agents, Virus Genome, Drug Efficacy, Chloroquine, Lopinavir Plus Ritonavir
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 & 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 Behaviour, 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.
Source: Authors' elaboration based on Scopus database, July 2020.

Table A2. The results of binary logistic models (coefficients and P-values) for classification of COVID-19 documents across subject areas (January-June 2020).

Binary dependent variable	Health Sciences		Life Sciences		Physical Sciences		Social Sciences & Humanities	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-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
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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

Table A2. Cont.

Binary dependent variable	Health Sciences		Life Sciences		Physical Sciences		Social Sciences & Humanities	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
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
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 R2	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.

Source: Authors' elaboration based on Scopus database (only documents with full abstract included), July 2020.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 September 2020 doi:10.20944/preprints202006.0299.v4

Table A3. Clusters based on 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 Haematology	Radiology, Nuclear Medicine and Imaging; Haematology; Paediatrics, Perinatology and Child Health; Oncology; Obstetrics and Gynaecology
	Dermatology and Neurology	Dermatology; Neurology (Clinical); Pathology and Forensic; Medicine; Histology; Anatomy
	Cardiology, Pulmonary and Anaesthesiology	Cardiology and Cardiovascular Medicine; Pulmonary and Respiratory Medicine; Anaesthesiology 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	Orthopaedics and Sports Medicine; Physical Therapy, Sports Therapy and Rehabilitation; Rehabilitation; Complementary and Alternative Medicine; Occupational Therapy
	Public Health*	Public Health, Environmental and Occupational Health; Family Practice; Community and Home Care
Life Sciences	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, Behaviour and Systematics; Applied Microbiology and Biotechnology; Developmental Biology
	Molecular Biology and Biochemistry*	Molecular Biology; Biochemistry; Structural Biology; Biophysics
	Immunology, Neuroscience and Endocrine Systems*	Immunology; Behavioural Neuroscience; Endocrine and Autonomic Systems
	Virology and Microbiology*	Virology; Microbiology; Parasitology
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 & 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 Behaviour 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.

Source: Authors' elaboration based on Scopus database, July 2020.

References

1 Fan, Y.; Zhao, K.; Shi, Z.L.; Zhou, P. Bat Coronaviruses in China. *Viruses* **2019**, *11*, 210, doi: https://doi.org/10.3390/v11030210.

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, doi: https://doi.org/10.1093/jtm/taaa008.

4 Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; He, D. 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, doi: https://doi.org/10.1016/j.ijid.2020.02.058.

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, doi: https://doi.org/10.1016/S0140-6736(20)30260-9.

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, doi: https://doi.org/10.1080/23812346.2020.1744922.

7 ECDC. COVID-19 situation update worldwide, as of 17 July 2020. Available online: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases (accessed on 6 September 2020).

8 IMF. *World Economic Outlook, April 2020: The Great Lockdown*; IMF: Washington, 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.; Hu, Y. From SARS to COVID-19: A bibliometric study on emerging infectious diseases with natural language processing technologies. *Preprints* **2020**, doi: https://doi.org/10.21203/rs.3.rs-25354/v1.

11 Zhai, F.; Zhai, Y.; Cong, C.; Song, T.; Xiang, R.; Feng, T.; Liang, Z.; Zeng, Y.; Yang, J.; Yang, J.; Liang, J. Research Progress of Coronavirus Based on Bibliometric Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3766, doi: https://doi.org/10.3390/ijerph17113766.

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, doi: https://doi.org/10.3390/ijerph17093082.

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*, doi: https://doi.org/10.3145/epi.2020.may.22.

14 Ram, S. Coronavirus Research Trends: A 50–Year Bibliometric Assessment. *Sci. Tech. Libr.* **2020**, *39*, 210-266, doi: https://doi.org/10.1080/0194262X.2020.1742270.

15 Joshua, V.; Sivaprakasam, S. Coronavirus: Bibliometric analysis of scientific publications from 1968 to 2020. *Med. J. Islam. Repub. Iran* **2020**, *34*, 456-463, doi: https://doi.org/10.34171/mjiri.34.64.

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, doi: https://doi.org/10.7759/cureus.7357.

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, doi: https://doi.org/10.21037/atm.2020.04.26.

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**, doi: https://doi.org/10.1101/2020.04.20.046144.

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, doi: https://doi.org/10.23750/abm.v91i9-S.10121.

21 Locher, C.; Moher, D.; Cristea, I.; Florian, N. Publication by association: the Covid-19 pandemic reveals relationships between authors and editors. *MetaArXiv* **2020**, doi: https://doi.org/10.31222/osf.io/64u3s.
22 Lee, J.J.; Haupt, J.P. Scientific Globalism During A Global Crisis: Research Collaboration and Open Access Publications on COVID-19. *Research Square* **2020**, doi: https://doi.org/10.21203/rs.3.rs-33977/v2.
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, doi: https://doi.org/10.1186/s12879-020-05293-z.
24 Fan, J.; Gao, Y.; Zhao, N.; Dai, R.; Zhang, H.; Feng, X.; Shi, G.; Tian, J.; Chen, C.; Hambly, B.D.; Bao, S. Bibliometric analysis on COVID-19: A comparison of research between English and Chinese studies. *Front. Public Health* **2020**, *8*, 477, doi: https://doi.org/10.3389/fpubh.2020.00477.
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*, doi: https://doi.org/10.7554/eLife.58807.
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, doi: https://doi.org/10.1016/j.dsx.2020.07.007.
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**.
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, doi: https://doi.org/10.1016/j.imr.2020.100490.
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**, doi: https://doi.org/10.2139/ssrn.3597812.
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, doi: https://doi.org/10.1016/j.jbusres.2020.06.057.
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, doi: https://doi.org/10.34171/mjiri.34.51.
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, doi: https://doi.org/10.17509/ijost.v5i2.24522.
33 Hossain, M.M. Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. *Preprints* **2020**, doi: https://doi.org/10.12688/f1000research.23690.1.
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, doi: https://doi.org/10.26355/eurrev_202003_20712.
35 Nasab, F.R., 2020. Bibliometric Analysis of Global Scientific Research on SARSCoV-2 (COVID-19). *Preprints* **2020**, doi: https://doi.org/10.1101/2020.03.19.20038752.
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, doi: https://doi.org/10.1016/j.jcot.2020.04.030.
37 McKinney, W. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc. 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, doi: https://doi.org/10.1016/j.omega.2019.08.005.
39 De Felice, F.; Polimeni, A. Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis. *In Vivo* **2020**, *34*, 1613-1617, doi: https://doi.org/10.21873/invivo.11951.

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*, doi: https://doi.org/10.3145/epi.2020.ene.03.

41 VanderPlas, J. *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc. 2016.

42 Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523-538, doi: https://doi.org/10.1007/s11192-009-0146-3.

43 Levandowsky, M.; Winter, D. Distance between sets. *Nature* **1971**, 234, 34-35, doi: https://doi.org/10.1038/234034a0.

44 Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), June 28 - July 3; Austin, Texas; van der Walt, S., Millman, J., Eds.; 62-96.

45 Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90-95. doi: https://doi.org/10.1109/MCSE.2007.55.

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*, doi: https://doi.org/10.1371/journal.pone.0162721.

47 Loper, E.; Bird, S. NLTK: the natural language toolkit. arXiv:cs/0205028.

48 Miller, G.A. WordNet: a lexical database for English. Commun. ACM **1995**, *38*, 39-41, doi: https://doi.org/10.1145/219717.219748.

49 Perkins, J. *Python text processing with NLTK 2.0 cookbook*. Packt Publishing Ltd. 2010.

50 Lilleberg, J.; Zhu, Y.; Zhang, Y. Support vector machines and word2vec for text classification with semantic features. *In 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)* (136-140). IEEE, doi: https://doi.org/10.1109/ICCI-CC.2015.7259377.

51 Windmeijer, F.A. Goodness-of-fit measures in binary choice models. *Econom. Rev.* **1995**, *14*, 101-116, doi: https://doi.org/10.1080/07474939508800306.

52 Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning, 2006, June (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, doi: https://doi.org/10.25008/jkiski.v5i1.356.

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**, doi: https://doi.org/10.20944/preprints202003.0443.v1.

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, doi: https://doi.org/10.3390/ijerph17114095.

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**, doi: https://doi.org/10.20944/preprints202006.0161.v1.

57 Helmy, Y.A.; Fawzy, M.; Elaswad, 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, doi: https://doi.org/10.3390/jcm9041225.