




COVID-19 research progress: Bibliometrics and visualization analysis

Maryam Okhovati¹, Homa Arshadi*¹ 

Received: 20 Aug 2020

Published: 9 Feb 2021

Abstract

Background: Coronavirus primarily targets the human respiratory system, COVID-19 (Coronavirus disease 2019) triggered in China in the late 2019. In March 2020, WHO announced the COVID-19 pandemic. This study aims to analyze and visualize the scientific structure of the COVID-19 publications using co-citation and co-authorship.

Methods: This is a scientometric study. Web of Science Core Collection (WoSCC) was searched for all documents regarding COVID-19, MERS-Cov, and SARS-Cov from the beginning to 2020. An Excel spreadsheet was applied to gather and analyze the data and the CiteSpace was used to visualize and analyze the data.

Results: A total of 5159 records were retrieved in WoSCC. The structure of the network indicated that the network mean silhouette was low (0.1444), implying that the network clusters' identity is not identifiable with high confidence. The network modularity was 0.7309. The cluster analysis of the co-citation network on documents from 2003 to 2020 provided 188 clusters. The largest cluster entitled, "the Middle East respiratory syndrome coronavirus" had 255 nodes. The coauthorship network illustrated that the most prolific countries, USA, China, and Saudi Arabia, have focused on a specific field and have formed separate clusters.

Conclusion: The present study identified the important topics of research in the field of COVID-19 based on co-citation networks as well as the analysis of clusters of countries' collaborations. Despite the similarities in the production behavior in prolific countries, their thematic focus varies so that a country like China plays a role in "Quantitative Detection" cluster, while USA is the leading country in the "Biological Evaluation" cluster.

Keywords: Covid-19, Scientometrics, Visualization analysis, Co-citation, Collaboration network, Coauthorship, CiteSpace

Conflicts of Interest: None declared

Funding: None

*This work has been published under CC BY-NC-SA 1.0 license.

Copyright© Iran University of Medical Sciences

Cite this article as: Okhovati M, Arshadi H. COVID-19 research progress: Bibliometrics and visualization analysis. *Med J Islam Repub Iran.* 2021 (9 Feb);35:20. <https://doi.org/10.47176/mjiri.35.20>

Introduction

The coronavirus is one of the major pathogens that primarily targets the human respiratory system (1). In the past 2 decades, the outbreaks of the coronavirus infection have included 3 widespread epidemics: the severe acute respiratory syndrome (SARS) in China in 2002, the Middle East respiratory syndrome (MERS) in Saudi Arabia in 2012 (2), and COVID-19 triggered in Wuhan, China in the late 2019 (1, 3). The coronavirus was first introduced in

MeSH database in 1994 and COVID-19 was added to MeSH (medical subject headings) in February 2020. Searching "COVID-19" in MeSH database provides 37 hits, all of which are added to MeSH in 2020 (4). COVID-19 is highly infectious, and its main clinical symptoms include fever, dry cough, fatigue, myalgia, and dyspnea, headache, sore throat, chills, hemoptysis, diarrhea, lymphopenia, and consolidation of the lung (1, 5, 6).

Corresponding author: Homa Arshadi, h.arshadi@kmu.ac.ir

¹ Medical Library and Information Sciences Department, School of Management and Information Science, Kerman University of Medical Sciences, Kerman, Iran

↑What is "already known" in this topic:

Since the COVID-19 virus is unknown, there is little information on the structure of the studies in this field of science.

→What this article adds:

The results showed that despite the pandemic, the countries' research orientation was different. There were weak links between the thematic clusters in the map; for example, the United States focused on biological evaluation, while China focused on quantitative diagnosis.

It (Coronavirus disease 2019) seems to spread mainly through respiratory droplets and close contacts. It is a significant threat to global health (3). COVID-19 is a global public health concern and WHO declared public health emergency (1), and in March 11, 2020, the WHO announced the COVID-19 as a pandemic (7). Because of its pandemic spread, a large corpus of papers has been published (8). A huge body of literature on COVID-19 is free to accelerate the research. Therefore, the evaluation of the quantitative and qualitative value of the body of the literature is important to have the pattern of the literature and identify the gaps and use the results.

Bibliometrics is a method suitable for the assessment of the trends of research. It has widely been used to trace the scientific outputs in different fields, such as microbiome-gut-brain axis (MGBA) field (9), osteoarthritis (10), aflatoxin (9), oncolytic virus (11), and chronic hepatitis B treatment-related research (12), the zika (13), respiratory syncytial virus (14), the West Nile virus (15), and methanol poisoning (16). Chu et al did a bibliometrics study to identify the 100 most cited papers in head and neck cancer (17). The global research trends in the West Nile virus (15) and methanol poisoning (16) were studied using bibliometrics methods. A bibliometric assessment of the scientific output was published by Patiño-Barbosa et al (18). Zyoud analyzed the Middle East respiratory syndrome coronavirus output by bibliometric methods (19).

Within a few months after the outbreak of COVID-19, some studies traced the publications on COVID-19. In the study by Dehghanbanadaki et al (2020), Scopus was searched to retrieve the documents on COVID-19 between December 1, 2019, and April 1, 2020. A total of 923 documents were retrieved, of which 418 were original articles. The average citation was 2.76 per document, with an h-index of 23. China published the largest number of papers (348), followed by the United States (160), and the United Kingdom (80). According to this study, the Lancet and BMJ Clinical Research Ed published the most documents and University of Hong Kong and Huazhong University of Science and Technology published the most documents (20).

Bhattacharya and Singh (2020) applied a scientometric and altmetric research to review the COVID-19 output. The data were gathered from the Dimensions Database. Ten most popular research papers on COVID-19 were identified by altmetrics sites. Papers with the highest citations, most popular papers (Altmetric Score) and key collaborations were considered to identify the 10 most active countries. Medical and health sciences, public health and health services, and clinical sciences were the key areas on COVID-19 based on the papers. The words pandemics and China had the most occurrences. The influence of COVID-19 on the society was investigated using Google Trends (8).

Tao et al (2020) analyzed the research trends on the coronaviruses using bibliometric analysis to identify new prevention strategies. A total of 9760 publications between 2000 and 2020 and indexed in Web of Science Core Collection were investigated. The Journal of Virology has been the most popular journal according to the number of

papers published. The United States followed by China had the top positions worldwide based on the number of publications. Analyzing the keywords, they identified 5 coronavirus research hotspot clusters, and the top 25 keywords with the strongest citation bursts were identified (21).

Zhou and Chen (2020) considered some parameters, including time, regions, impact factors, and citation times, to analyze the global trends of the coronavirus from 2000 to 2020. Searching the Web of Science Core Collection, 9043 documents were retrieved on the coronavirus from 123 countries. The papers were published in 1202 journals and the Journal of Virology was the dominant journal according to the publication volume. USA (3101) followed by China (2230) published the largest number of papers and Germany (584) ranked the third. The 10 most cited articles on coronavirus research were reported in this paper and the keywords in the papers were visualized (3).

Another study by Joshua and Sivaprakasam examined the coronavirus publications published from 1968 to March 2020. The Web of Science Core Collection was searched. A total of 6424 documents were retrieved on the coronavirus. USA (N = 2345), China (N = 1067), and Germany (N = 480) published most papers, respectively. The University of Hong Kong was the most productive institute. Infectious diseases, microbiology, and virology were the top research areas, and the Journal of Virology was the most productive (22).

Bonilla-Aldana et al (2020) conducted a bibliometric analysis to examine SARS-CoV, MERS-CoV, and the 2019-novelCoV publications indexed in WOS, PubMed and Scopus from January 1951 to January 2020. Searching PubMed led to retrieving 14 455 articles, while searching WOS retrieved 11 775 papers. Among the papers indexed in these 2 databases, USA, China, and Germany had the most contribution; nonetheless, the Scopus search identified 18 158 articles, USA, China, and the United Kingdom were the leading countries in this regard. In all 3 databases, articles on SARS-CoV were more than papers on MERS-CoV (23).

Jia et al (2020) searched coronaviruses publications in the Web of Science Core Collection between 2003 and 2020 and found 8433 documents. The most prolific countries were USA (N = 2791 documents), China (N = 2231), and Germany (N = 564), respectively. The University of Hong Kong was the leading institute in this field (399). The top 10 cited authors and cited references contributed to the publications on the coronavirus were also reported. The coauthorship map indicated that the scholars could be roughly divided into 5 major academic groups. The top 5 keywords in terms of citation counts for coronavirus research were “SARS-CoV (1781 records)”, “infection (1528 records)”, “acute respiratory syndrome (1213 records)”, “identification (1141 records)” and “respiratory syndrome coronavirus (920 records)”. The top 5 keywords with high centrality were as follows: “infection (0.14)”, “identification (0.13)”, “infectious bronchitis virus (0.11)”, “respiratory syndrome coronavirus (0.10)”, “antibody (0.09)”. The 5 keyword clusters consisted of 133 nodes with 529 links, generating 5 keyword clusters. The

silhouette value of the clusters #0 to 4 was from 0.669 to 0.76, showing good homogeneity (24).

Zyoud and Al-Jabi (2020) studied the document type, country, collaboration patterns, affiliation, journal name, and citation patterns of papers on COVID-19 extracted from Scopus (25). ElHawary et al performed a systematic search using keywords related to COVID-19 in Web of Science, Scopus, and Google Scholar and studied the characteristics of the top 50 cited. These papers were cited 63 849 times until May 29, 2020. Type of study, journal, number of citations, number of authors, country of publication, and study content were other characteristics studied (26). Another paper analyzing COVID-19 research activity concluded that due to the pandemic and the high rate of infectivity, there is a global need for more relevant research that can help to be more informed about the clinical and pathogenic characteristics of this disease (27).

Bibliometric studies on different fields of science have attempted to analyze the scientific outputs to give a view of the patterns of the publications. Although the outbreak of COVID-19 occurred in March, 2020, some bibliographic studies have been done to analyze and visualize an international view of what has been done and what are the probable gaps. This study is different from previous studies in terms of the networks presented. It focuses on analyzing and visualizing the co-citation and discovering the characteristics of the clusters on the collaboration of the countries, and does not report the descriptive data, such as the leading authors, countries, and journals, which are presented in previous studies. The co-citation clusters were analyzed to introduce the leading documents based on centrality, degree, and burstiness, besides that, the main clusters of collaboration of the countries were also

separately analyzed. In other words, this study answers the following question: What is the pattern of the publications in COVID-19? Since visualizing the status of the research trend will show the gaps worthy of further exploration, it can help to inform policymakers to employ better plans. Thus, we hope the present study's findings, which show the current status of COVID-19, could improve policy-making.

Methods

This was a scientometrics study using co-citation and coauthorship analysis. The data (the full record and cited references) were collected from the Web of Science Core Collection (WoSCC). The search strategy was done with no time limit but no paper COVID-19 was found before 2003; thus, the time span covered was 2003 to 2020. The keyword "COVID-19" and all its synonyms were searched in the field "TOPIC" (title, abstract, and keywords). Different forms of the word and synonyms were extracted from MeSH and recent papers on COVID-19 and finally the terms were checked with a virology expert. An Excel spreadsheet was applied to analyze the data and CiteSpace version 5.6.R3 (Drexel University) was used to visualize and analyze the network indices. The search strategy is illustrated in Figure 1.

The definitions of each indicator calculated in this study, based on CiteSpace software, are reported in Table 1.

Results

Of the 5159 records retrieved in WOS, co-citation and collaboration of the countries were visualized and analyzed using 1895 nodes in CiteSpace. Table 2 indicates the 1859 selective groups that contributed in analysis pro-

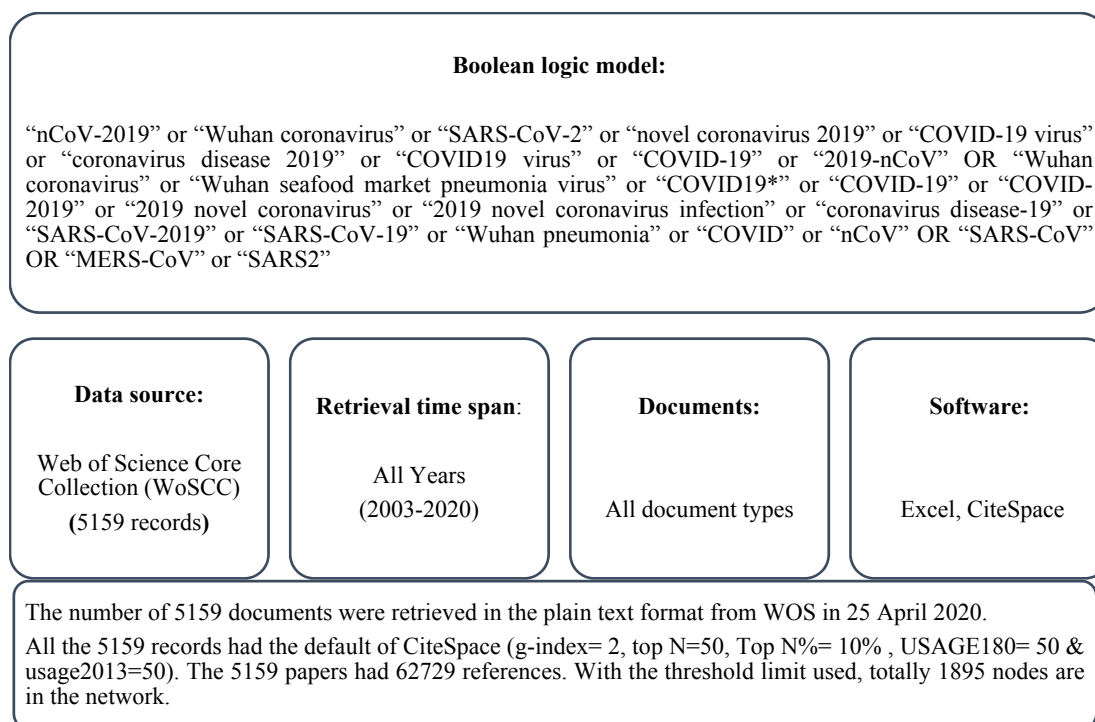


Fig. 1. Diagram of Search Strategy, Inclusion Criteria, and Search Results

Table 1. The definition of the indices

Indices	Definition
Silhouette	Silhouettes graphically interpret the clusters. The silhouettes are combined into a single plot to display the entire clustering. The average silhouette width is used to select an ‘appropriate’ number of clusters and provides an evaluation of clustering validity. This metric is useful in estimating the uncertainty involved in identifying the nature of a cluster, showing which objects lie well within their cluster, and which ones are merely somewhere in between clusters. The silhouette value of a cluster ranges from -1 to 1, the larger value implies higher consistency (inter-cluster connectivity). The higher the silhouette score means the homogeneity is higher (28).
Modularity	The value of modularity ranges from 0 to 1. A high modularity indicates a well-structured while a low modularity implies that the network cannot be reduced to clusters with clear boundaries (28).
Centrality	“Betweenness is useful as an index of the potential of a point for control of communication. it is based upon the frequency with which a point falls between pairs of other points on the shortest or geodesic paths connecting them”(29); a node with high betweenness centrality would be a pivotal point, or a turning point (30).
Burstiness	The burst detection in CiteSpace is based on Kleinberg’s algorithm, which measures the sharp increase (31). Burst as an indicator of a most active area of research is trying to show if a publication evidently has attracted an extraordinary degree of attention from its scientific community. Burst attempts to detect whether and when the citation count of a particular reference has surged. It determines if a given frequency function had significant fluctuations during a short time (32). Burst-detection can identify emergent terms and new research fronts even (33)

Table 2. The structural characteristics of the co-citation map

Records	Mean Silhouette	Modularity	Largest Component	Density	Links	Nodes
5159	0.1444	0.7309	1704 (89%)	0.0063	11284	1895

cess. These nodes have been linked 11284 times in co-citation network. The Silhouettes score is relatively low showing that the elements in each cluster are not consistent and there is no homogeneity. Furthermore, the existence of many small clusters is mainly the cause of this low score. The modularity of the network is relatively high (0.7309), which shows that the network has logically been divided into coupled clusters. The largest component having 1704 documents covers 89% of the documents.

Reference Co-citation Analysis

One of the maps, which can be presented by CiteSpace, is co-citation of the documents. The co-citation network is visualized by selecting cited references, and the co-citation map of the papers related to COVID-19 is shown in Figure 2. This network consists of 188 clusters, the biggest cluster is related to “the Middle East respiratory syndrome coronavirus” and the smallest cluster having 5 nodes is labeled “Inhibitors Design.” The cluster label implies that the cluster has been cited by the documents in

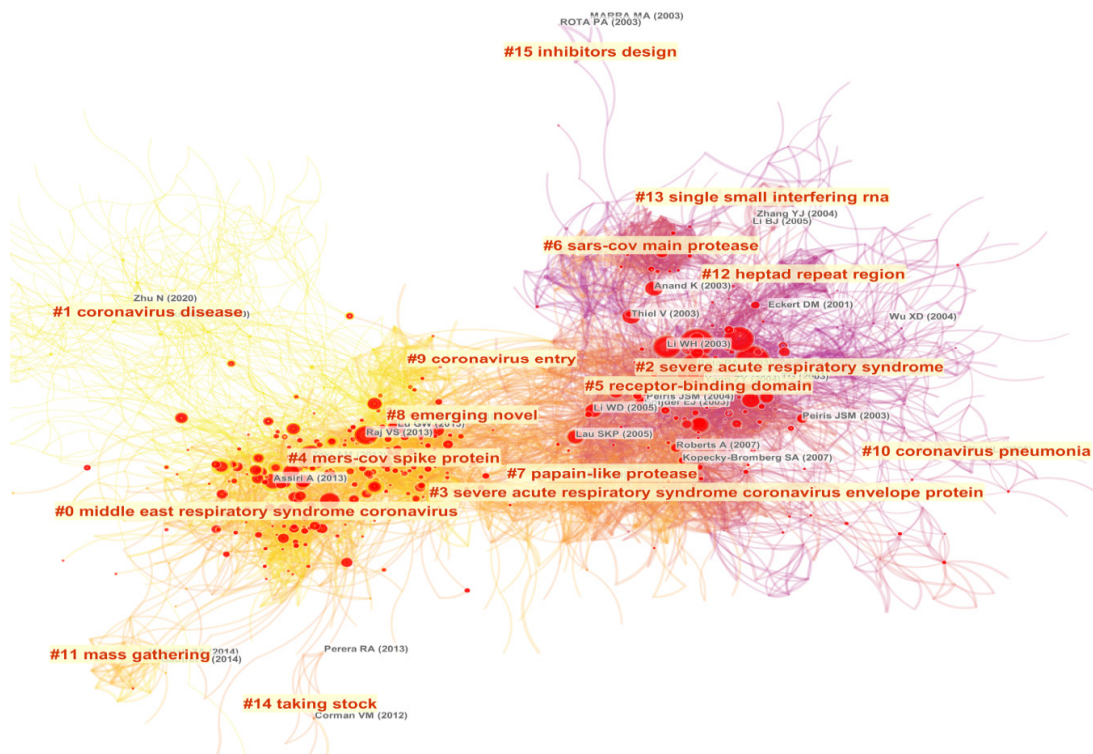


Fig. 2. Co-citation map of references from publications on coronavirus research (Log-Likelihood Ratio (LLR))

the same subject, for example, cluster 0 has been cited by papers on MERS coronavirus.

The number of nodes in each network, the silhouette mean, age, and the name of clusters are reported in Table 3. The highest silhouette mean belonged to the cluster 14 “taking stock” (0.998) having only 8 nodes. The age of the clusters is the mean of the papers’ year of publication. The biggest cluster is “the Middle East respiratory syndrome coronavirus.” This cluster consists of 255 nodes and its age is 2014. The silhouette mean is 0.765.

As shown in Table 3, the silhouette score is relatively high in all major 15 clusters, which shows the quality of clusters.

Table 3 shows the characteristics of the major clusters in the network. The labels of the clusters were LLR (log-likelihood ratio), which according to the CiteSpace handbook, was suggested as the best method to obtain clustering results (32).

The top 10 papers for degree, burst, and frequency are

presented in Table 4. The most papers belonged to cluster 2 according to frequency, and the paper by Zaki, A.M. et al (2012) from the zero cluster had the highest frequency in this network.

The papers with the highest burstiness score belonged to cluster No. 2, which focuses on “severe acute respiratory syndrome.” The number of papers in this cluster increased rapidly, which was expected because of the outbreak of SARS. Cluster 5 with 159 papers had the highest degree value, and it is labeled “Receptor-Binding Domain” in the map.

According to Table 5, there are 140 nodes and 861 links in the countries’ collaboration network, which means that 140 countries have cooperated through 861 links. The mean silhouette score is low in the network, implying that the overall structure of the network is not of high quality and the nature of the clusters is not clear. The network modularity also indicates that the link between the clusters is weak. The network density implies that the cooperation

Table 3. The structural characteristics of the clusters

Cluster ID	Size	Silhouette	Mean (Y)	Cluster Name
0	255	0.765	2014	middle east respiratory syndrome coronavirus
1	243	0.855	2019	Coronavirus disease
2	233	0.666	2003	severe acute respiratory syndrome
3	169	0.811	2005	severe acute respiratory syndrome coronavirus envelope protein
4	165	0.832	2014	mers-cov spike protein
5	159	0.74	2005	receptor-binding domain
6	124	0.916	2002	sars-cov main protease
7	117	0.897	2008	papain-like protease
8	91	0.845	2011	emerging novel
9	64	0.894	2010	coronavirus entry
10	24	0.971	2003	coronavirus pneumonia
11	19	0.99	2014	mass gathering
12	19	0.98	1999	heptad repeat region
13	9	0.997	2004	single small interfering rna
14	8	0.998	2013	taking stock
15	5	0.996	1999	inhibitors design

Table 4. The top 10 documents based on 3 indices (frequency, burstiness, and degree)

	Frequency	Authors	Burst	Authors	Degree	Authors
1	920	Zaki AM, 2012 #0	215.87	Rota PA, 2003 #2	64	Bisht H, 2004 #5
2	763	Rota PA, 2003 #2	207.79	Ksiazek TG, 2003 #2	64	Weingartl H, 2004 #0
3	735	Ksiazek TG, 2003 #2	187.93	Drosten C, 2003 #2	60	Adney DR, 2014 #1
4	666	Drosten C, 2003 #2	187.36	Marra MA, 2003 #2	60	Bukreyev A, 2004 #5
5	664	Marra MA, 2003 #2	172.84	Zaki AM, 2012 #0	59	Raj VS, 2013 #4
6	557	Peiris JSM, 2003 #2	156.71	Peiris JSM, 2003 #2	57	Yang ZY, 2004 #5
7	418	Li WH, 2003 #5	120.87	Li WH, 2003 #5	57	Jiang LW, 2014 #4
8	403	Assiri A, 2013 #0	83.69	Assiri A, 2013 #0	56	He YX, 2005 #5
9	381	Raj VS, 2013 #4	78.63	Li WD, 2005 #8	53	Yang ZY, 2005 #5
10	338	Huang CL, 2020 #1	77.93	Guan Y, 2003 #2	53	Volz A, 2015 #4

Table 5. The structural characteristics of the countries’ network

Records	Mean Silhouette	Modularity	Largest Component	Density	Links	Nodes
5137	0.3125	0.26	127 (90%)	0.0885	861	140

among the countries is loose. In the largest cluster, 90% of the countries have cooperated through direct or indirect links.

The density of 0.0885 in this study indicated that the network of countries in articles on the coronavirus is quite loose.

The country cluster includes 17 subclusters and the biggest cluster has 38 nodes. The collaborations of countries and regions were analyzed using the CiteSpace from 2003-2020. The top 10 contributing countries are presented in Table 4 in terms of frequency, burst, centrality, and degree. The high centrality of the network indicates the importance of the network nodes. The centrality value is a number between 0 and 1, and the closer to 1, the more important in the network. According to Figure 3, the highest centrality belongs to the United States (0.29), followed by China (0.2), and Germany (0.19).

The purple lines surrounding the nodes of the USA, China, Germany, France, Saudi Arabia, Spain, and Eng-

land demonstrate the high centrality and the high thickness of the rings for USA and China, showing the value of these countries (nodes) in the network (32).

A strong burstiness of a node demonstrates a potentially interesting work, which means it has attracted significant attention in a short period of time (32). The burstiness has been shown in tree rings in red in Figure 4. Saudi Arabia has the highest burstiness value (85.27), followed by China (45.94), and Taiwan (45.94).

In a total of 140 countries, only 19 had the burstiness score and most of the countries belonged to the cluster zero. The highest score of burstiness in the zero cluster belonged to Taiwan.

Collaboration Network between Countries Based on the Degree Score

The degree shows the co-authorship score of a researcher with other researchers in the network. This is the simplest and the most common value to evaluate the coau-

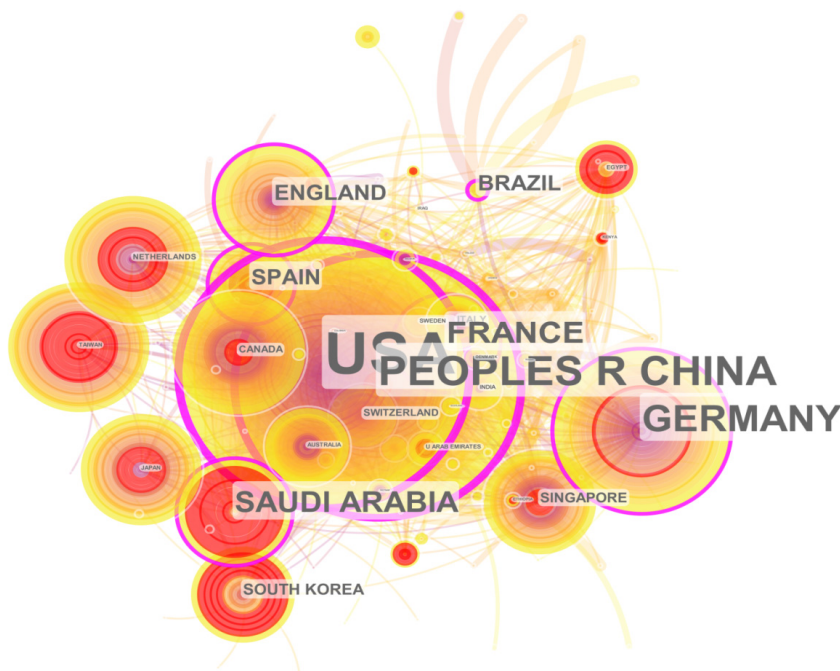


Fig. 3. Network map of countries collaborations based on centrality

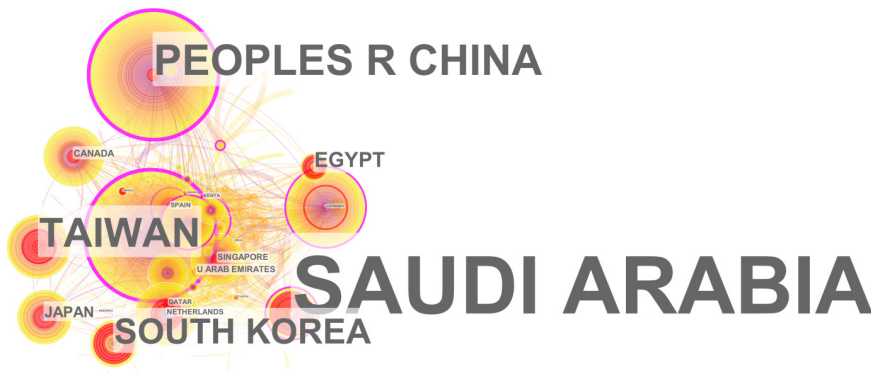


Fig. 4. Network map of countries' collaborations based on burstiness

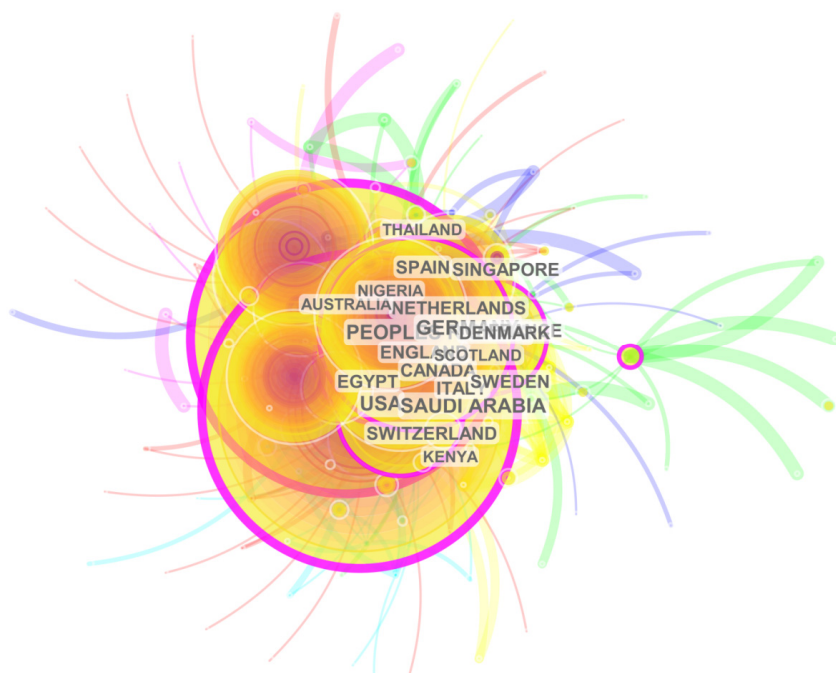


Fig. 5. Network map of countries' collaborations based on degree

Table 6. Top 10 contributing countries by frequency, burstiness, and centrality

Frequency	Countries	Burst	Countries	Centrality	Countries	Degree	Countries
1578	USA	85.27	Saudi Arabia	0.29	USA	63	USA
1574	Peoples R China	45.94	Peoples R China	0.2	Peoples R China	63	Germany
377	Saudi Arabia	43.44	Taiwan	0.19	Germany	59	Peoples R China
335	Germany	37.83	South Korea	0.16	France	57	Saudi Arabia
307	England	22.31	Egypt	0.14	Saudi Arabia	51	England
263	South Korea	16.89	Japan	0.11	Spain	47	France
245	Canada	10.86	Canada	0.11	England	40	Singapore
231	Taiwan	9.58	U Arab Emirates	0.11	Brazil	39	Spain
210	Netherlands	9.11	Singapore	0.08	South Korea	39	Netherlands
179	Singapore	8.44	Netherlands	0.07	Singapore	38	Sweden

thorship. It is obvious that the actors with high coauthorship are the most active in the scientific community and have an important role in information communication (34). According to Lu and Hsu, the degree score indicates the ability of the network nodes to collaborate, and a node having a high degree value indicates the high degree of collaboration (35). The degree score of countries is shown in Figure 5.

Finally, the top 10 countries according to these 3 indices (centrality, burstiness, and degree) are summarized in Table 6.

Table 6 shows that USA is in the first rank for frequency, centrality, and degree indices, with a value of (1578), (0.29) and (63), respectively. Nevertheless, the first place for burstiness belongs to Saudi Arabia (85.27).

Table 7 shows the network clusters of countries' collaborations. Six clusters are sorted according to the number and the size of nodes. The largest node includes 38 countries collaborating in the field of COVID-19. As the table indicates, the silhouette score is relatively high for all the 6 clusters. It implies that different countries focusing on 1 or some more countries have formed separate clusters and their nature can be determined by clusters, while the overall collaboration network having a low sil-

houette score lacks a coherent structure. The highest silhouette score belonged to cluster 4 and the oldest cluster in the collaboration network belonged to cluster 5, which is characterized by quantitative detection.

The biggest cluster (Cluster 0) mainly focused on biological evaluation, the clusters on next ranks concentrated on travel cancellation (Cluster 1), evidence-based recommendation (Cluster 2), the Middle East respiratory syndrome coronavirus (Cluster 3), public health (Cluster 4), and the final cluster (Cluster 5) focused on quantitative detection (Fig. 6). The clusters labeled with each subject were cited by articles on the same subjects. For instance, the cluster labeled 0 indicated that this cluster was cited by articles on biological evaluation, and cluster 1 was cited by articles on travel cancellation.

Table 7. The Structural Characteristics of Countries' Clusters

Cluster ID	Size	Silhouette	mean(Year)
0	38	0.696	2009
1	23	0.802	2012
2	21	0.531	2010
3	19	0.662	2014
4	16	0.835	2014
5	10	0.786	2008

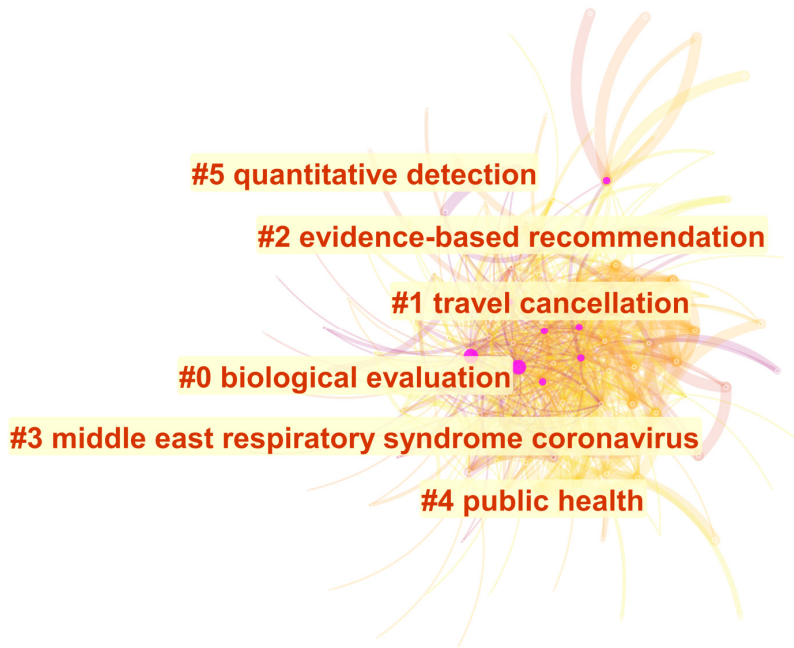


Fig. 6. The cluster view of the countries' collaboration network (Log-Likelihood Ratio (LLR))

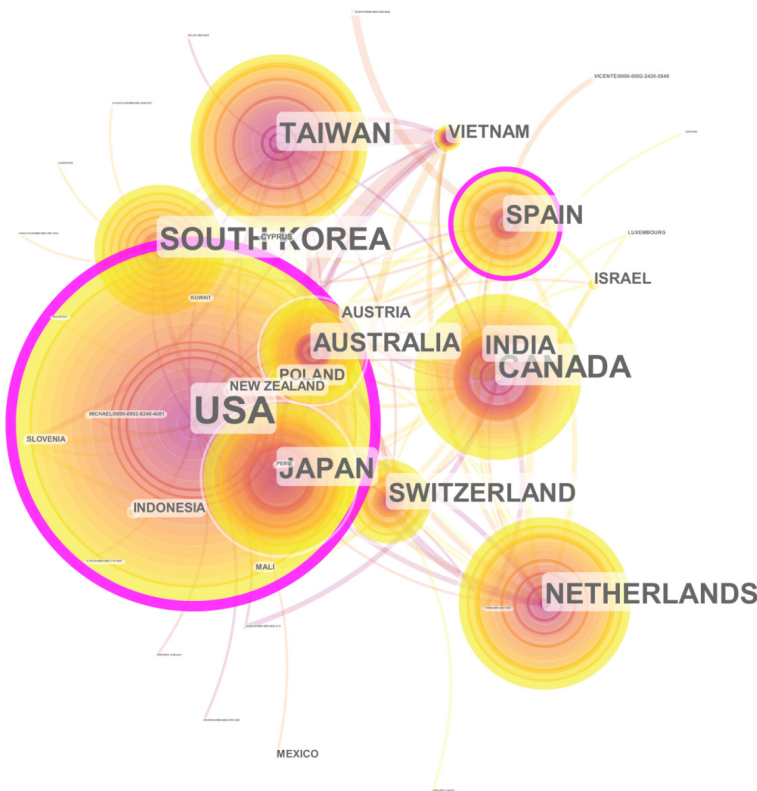


Fig. 7. Cluster 0 (biological evaluation)

Cluster 0: Biological Evaluation

The largest cluster describing the countries collaborations focuses on biological evaluation and include 38 countries (Fig. 7). The age of the cluster is 2009 and due to the silhouette score, it has a relatively high-quality structure. USA preceded other countries in frequency,

degree, and centrality indices. The purple color around each node implies its high centrality and it shows centrality value 0.29 for the USA and 0.11 for Spain. However, Taiwan and South Korea had a high score in burstiness; in fact, the biggest sudden change has happened in Taiwan.

The most prolific countries in the largest cluster of

countries' collaboration are presented in Table 8, which have been cited by papers in the field of biological evaluation. USA ranked first for all the 3 indices, while Taiwan has experienced an abrupt change in burst, with the score of 43.44.

The rest of 5 main clusters of countries are presented in Figure 8.

Characteristics of Clusters 1 to 5

Cluster 1: Travel Cancellation

This cluster addresses travel cancellation and there are 23 countries in this cluster. England, Singapore, and Italy have the most publications in this cluster. England has the highest frequency (37), centrality (0.11), and degree value

(51) while the highest burstiness score belongs to Singapore (9.11).

Cluster 2: Evidence-based Recommendation

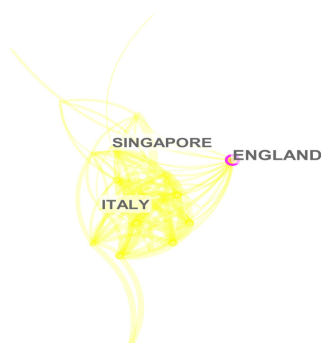
This cluster has a lower silhouette score than cluster 5 and includes 21 nodes. Germany has the highest score in all 4 indices: frequency (335), centrality (0.19), degree (63), and burstiness (4.08). In fact, Germany, collaborating with other countries such as Belgium and Brazil, has focused on evidence-based recommendation.

Cluster: 3 Middle East Respiratory Syndrome Coronavirus

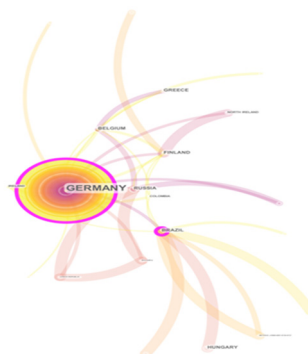
In the third cluster, 19 countries have formed a cluster

Table 8. Top countries contributing to the largest cluster (biological evaluation)

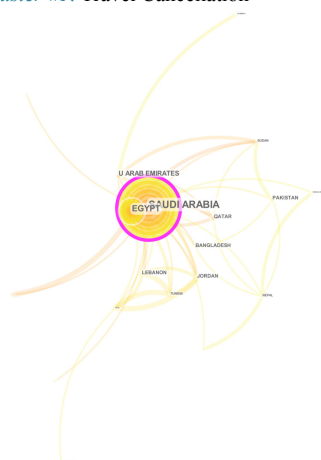
Countries	Frequency	Burst	Degree	Centrality
USA	1578		63	0.29
South Korea	263	37.83	20	0.08
Canada	245	10.86	37	0.06
Taiwan	231	43.44	20	0.04
Netherlands	210	8.44	39	0.04
Japan	172	16.89	22	0.03
Australia	146		27	0.04
Switzerland	119		34	0.05
Spain	108	7.85	39	0.11
India	94		22	0.03
Vietnam	28		22	0.01



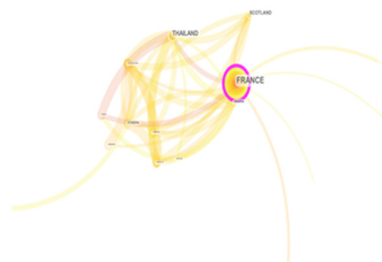
Cluster #1: Travel Cancellation



Cluster #2: Recommendation Evidence-based



Cluster#3: Middle East Respiratory Syndrome coronavirus



Cluster#4: Public Health



Cluster # 5: Quantitative detection

Fig. 8. Five other clusters of countries' collaborations

and the network structure is relatively good. It can be inferred that the outbreak of MERS may have encouraged Saudi Arabia to collaborate with neighborhood countries and making a cluster. Saudi Arabia has the highest score for all 4 indices with frequency (377), burst (85.27), degree (57), and centrality (0.14).

Cluster 4: Public Health

The first rank for all indices belonged to France, except burstiness. There are 16 countries in this cluster and the core is France and the focus is on public health. The centrality, degree, and frequency values for France are 0.16, 47, and 163, respectively.

Cluster 5: Quantitative Detection

The last cluster paid attention to quantitative detection of COVID-19. It is interesting that China with Portugal and Malaysia formed this cluster and the focus is on China. The centrality score in this cluster is 0.2, degree = 59, frequency = 1574 and burstiness = 45.94.

Discussion

COVID-19 triggered in China and its widespread prevalence in China and other parts of the world in a short time made the WHO to announce it as a pandemic. This study was conducted to analyze and visualize the co-citation and the countries' contribution in the scientific output besides the identification of the top authors and countries. The CiteSpace software was used to analyze and visualize the networks.

The co-citation network of the documents showed that various clusters were formed and each cluster focused on a topic. The link between the clusters was not strong. For example, the cluster 2 having 233 actors were cited by papers on SARS, while the papers with the topic "SARS Coronavirus envelope protein" are in a separate cluster (No. 3) with 169 nodes. The low quality of the network and the low silhouette value may be due to various clusters.

A total of 140 countries all over the world contributed to the scientific output on COVID-19. The network of countries' contribution included 17 clusters and both metrics-silhouette and modularity- were low, 0.3125 and 0.26, respectively, which shows that there was no homogeneity and the link between the clusters was not strong. The nodes within the clusters had not strong links. The weak links may be because of the unknown identity of the virus, which has a different behavior according to indigenous and genetic conditions of the countries and the focus was mostly on native documents rather than international output. Nevertheless, the centrality and degree metrics were high in USA rather than in other countries, showing that this country had the most contribution in the network because a powerful centrality has a great impact on the information flow (36). Burstiness shows the rate of change in a particular period of time (31). The high value of burst metric in Saudi Arabia may imply the outbreak of MERS and change in publication and citation behavior, which finally resulted in an increase in this metric. In Tao's study, USA, China and Saudi Arabia had the highest cen-

trality score, which was consistent with the result of our study. The centrality score of USA was 0.29 in our study and it was 0.24 in Tao's study (21).

According to Bhattacharya and Singh (2020), Dehghanbanadaki et al (2020) and Zhou and Chen (2020) studies, USA and China were in the first and the second ranks for the number of documents (3, 8, 20). Bhattacharya and Singh demonstrated that the keywords "pandemic" and "China" had the highest co-occurrence (8). In the study of Dehghanbanadaki et al, the highest co-occurrence belonged to the keyword "COVID-19" with 139 repeats (20). According to the results of the research of Zhou and Chen, the highest density belonged to "coronavirus" and "infection" keywords (3).

There was an interesting result in the clusters; various topics from "quantitative diagnosis" to "biological evolution" were identified in the coauthorship network of the countries; each cluster included a certain number of countries. For example, in the cluster 0, the largest cluster, with the topic biological evaluation, USA, South Korea, and Canada formed the cluster, while the leading countries of China, Saudi Arabia, and Germany were not in this cluster.

The study conducted by Jia et al (2020) covered the same time span as our study (2003-2020) and found that Germany and France with 0.16 score for centrality had the leading role in information flow. According to the findings of their study, USA, China, and Germany obtained the highest centrality degree. Jia et al presented 13 clusters and the largest cluster labeled "mers-cov" which is the same as in our study (24). The next similar cluster was cluster 1 labeled "spike protein" which matched the cluster 4 "MERS-Cov spike protein" in this study. The other similar cluster is cluster 3 in Jia et al's study, which matches cluster 2 in the present study (24). Zhai et al (2020) did a bibliographic research on coronavirus, covering the same time span as our study. They chose the periods 2003 to 2020 because the SARS outbreak occurred after 2003 and the leading countries were USA, China, and Germany. The most common keywords were "SARS", "Coronavirus", and "infection" (37). Zhu and Chen (2020) also reported that "coronavirus" and "infection" were the most frequent keywords in the documents studied (3).

The leading countries formed separate clusters and they focused on different topics. Because of the disease pandemics, it was expected that the countries collaborate have a more cohesive collaboration, which may be related to different policies in different countries. The leading countries in each cluster were USA, England, Germany, Saudi Arabia, France, and China working on biological evaluation, travel cancellation, evidence-based recommendation, public health, and quantitative detection. It is suggested that the relationship between the policies of different countries about the disease prevention, treatment, and their focus on topic, their publication behavior, and the rate of mortality be considered for next studies.

Conclusion

COVID-19 visualization indicated the countries' research collaboration and top papers and authors in differ-

ent COVID-19 subfields were also identified. The inconsistency in collaboration network among leading countries in COVID-19 publications and the countries focus on different research fields need in-depth studies considering their political conditions and health policy. It is suggested that future studies compare the prevalence and causes of mortality and research orientations to determine whether the research was tailored to the needs of those countries. Due to the rapid increase in the research on COVID-19, more maps can be presented to visualize the scientific production based on PubMed and Scopus.

Conflict of Interests

The authors declare that they have no competing interests.

References

- Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun.* 2020;102433.
- Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *In. J Epidemiol.* 2020.
- Zhou Y, Chen L. Twenty-Year Span of Global Coronavirus Research Trends: A Bibliometric Analysis. *Int J Environ Res Public Health.* 2020;17(9):3082.
- MeSH. COVID-19 [Supplementary Concept] 2020 [cited 2020 10 november]. Available from: <https://www.ncbi.nlm.nih.gov/mesh/2050205>.
- Hu X, Chen J, Jiang X, Tao S, Zhen Z, Zhou C, et al. CT imaging of two cases of one family cluster 2019 novel coronavirus (2019-nCoV) pneumonia: inconsistency between clinical symptoms amelioration and imaging sign progression. *Quant Imaging Med Surg.* 2020;10(2):508.
- Zhong BL, Luo W, Li HM, Zhang QQ, Liu XG, Li WT, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. *Int J Biol Sci.* 2020;16(10):1745.
- Organization WH. WHO Timeline-COVID-19. Ebola virus disease 2020 [Available from: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>].
- Bhattacharya S, Singh S. Visible Insights of the Invisible Pandemic: A Scientometric, Altmetric and Topic Trend Analysis. *arXiv preprint arXiv:200410878.* 2020.
- Sa'ed HZ, Smale S, Waring WS, Sweileh WM, Al-Jabi SW. Global research trends in microbiome-gut-brain axis during 2009–2018: A bibliometric and visualized study. *BMC Gastroenterol.* 2019;19(1):158.
- Zhang Q, Hou Y, Xing D, Lin J. Tracing scientific outputs in the osteoarthritis research field in China based on publications in the Web of Science. *Osteoarthr Cartil.* 2019;1(1-2):100007.
- Zou Y, Luo Y, Zhang J, Xia N, Tan G, Huang C. Bibliometric analysis of oncolytic virus research, 2000 to 2018. *Medicine.* 2019;98(35).
- Li G, Lin J, Jiang C, Feng Q, Wen L. Trends in chronic hepatitis B treatment-related research from 1973 to 2018: a bibliometric and visual analysis. *J Int Med Res.* 2019;0300060519893234.
- Albuquerque PC, Castro MJC, Santos-Gandelman J, Oliveira AC, Peralta JM, Rodrigues ML. Bibliometric indicators of the Zika outbreak. *PLoS Negl Trop Dis.* 2017;11(1):e0005132.
- Brüggmann D, Köster C, Klingelhöfer D, Bauer J, Ohlendorf D, Bundschuh M, et al. Respiratory syncytial virus: a systematic scientometric analysis of the global publication output and the gender distribution of publishing authors. *BMJ Open.* 2017;7(7).
- Al-Jabi SW. Global research trends in West Nile virus from 1943 to 2016: a bibliometric analysis. *Glob Health.* 2017;13(1):55.
- Sa'ed HZ, Al-Jabi SW, Sweileh WM, Awang R, Waring WS. Bibliometric profile of the global scientific research on methanol poisoning (1902–2012). *J Occup Med Toxicol.* 2015;10(1):17.
- Chu T, Kwok H, Chan J, Tse F. The 100 most cited manuscripts in head and neck cancer: a bibliometric analysis. *J Laryngol Otol.* 2019;133(11):936-42.
- Patiño-Barbosa AM, Bedoya-Arias JE, Cardona-Ospina JA, Rodríguez-Morales AJ. Bibliometric assessment of the scientific production of literature regarding Mayaro. *J Infect Public Health.* 2016;9(4):532.
- Sa'ed HZ. Global research trends of Middle East respiratory syndrome coronavirus: a bibliometric analysis. *BMC Infect Dis.* 2016;16(1):255.
- Dehghanbanadaki H, Seif F, Vahidi Y, Razi F, Hashemi E, Khoshmirasfa M, et al. Bibliometric analysis of global scientific research on Coronavirus (COVID-19). *Med J Islam Repub Iran.* 2020;34(1):354-62.
- Tao Z, Zhou S, Yao R, Wen K, Da W, Meng Y, et al. COVID-19 will stimulate a new coronavirus research breakthrough: a 20-year bibliometric analysis. *Ann Transl Med.* 2020;8(8).
- Joshua V, Sivaprakasam S. Coronavirus: Bibliometric analysis of scientific publications from 1968 to 2020. *Med J Islam Repub Iran.* 2020;34(1):456-63.
- Bonilla-Aldana DK, Quintero-Rada K, Montoya-Posada JP, Ramírez-Ocampo S, Paniz-Mondolfi A, Rabaan AA, et al. SARS-CoV, MERS-CoV and now the 2019-novel CoV: Have we investigated enough about coronaviruses?—A bibliometric analysis. *Travel Med Infect Dis.* 2020;33:101566.
- Jia Q, Shi S, Yuan G, Shi J, Shi S, Hu Y. Analysis of knowledge bases and research hotspots of coronavirus from the perspective of mapping knowledge domain. *Medicine.* 2020;99(22):e20378.
- Zyoud SeH, Al-Jabi SW. 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(1):561.
- ElHawary H, Salimi A, Diab N, Smith L. Bibliometric Analysis of Early COVID-19 Research: The Top 50 Cited Papers. *Int J Infect Dis.* 2020;13:1178633720962935.
- Chahrour M, Assi S, Bejjani M, Nasrallah AA, Salhab H, Fares M, et al. A Bibliometric Analysis of COVID-19 Research Activity: A Call for Increased Output. *Cureus.* 2020;12(3):e7357-e.
- Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *J Assoc Inf Sci Technol.* 2010;61(7):1386-409.
- Abbasi A, Hossain L, Leydesdorff L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *J Informetr.* 2012;6(3):403-12.
- Wang S, Du Y, Deng Y. A new measure of identifying influential nodes: Efficiency centrality. *Commun Nonlinear Sci Numer Simul.* 2017;47:151-63.
- Sohrabi B, Vanani IR, Jalali SMJ, Abedin E. Evaluation of Research Trends in Knowledge Management: A Hybrid Analysis through Burst Detection and Text Clustering. *J Inf Knowl Manag.* 2019;18(04):1950043.
- Chen C. *CiteSpace: A practical guide for mapping scientific literature*: Nova Science Publishers, Incorporated; 2016.
- Zhou W, Chen J, Huang Y. Co-Citation Analysis and Burst Detection on Financial Bubbles with Scientometrics Approach. *Econ Res.* 2019;32(1):2310-28.
- Ye Q, Li T, Law R. A coauthorship network analysis of tourism and hospitality research collaboration. *J Hosp Tour Res.* 2013;37(1):51-76.
- Luo YL, Hsu CH, editors. *An empirical study of research collaboration using social network analysis*. 2009 International Conference on Computational Science and Engineering; 2009: IEEE.
- Borgatti SP. Centrality and network flow. *Soc Networks.* 2005;27(1):55-71.
- Zhai F, Zhai Y, Cong C, Song T, Xiang R, Feng T, et al. Research Progress of Coronavirus Based on Bibliometric Analysis. *Int J Environ Res Public Health.* 2020;17(11):3766.