Covid-19 Knowledge Deconstruction and Retrieval: An Intelligent Bibliometric Solution

Mengjia Wu[1], Yi Zhang[1], Mark Markley[2], Caitlin Cassidy[2], Nils Newman[2], Alan Porter[2,3]

[1]Australian Artificial Intelligence Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia mengjia.wu@uts.edu.au, yi.zhang@uts.edu.au

[2]Search Technology, Inc., United States markmarkley@searchtech.com, caitlin.cassidy@searchtech.com, newman@searchtech.com, aporter@searchtech.com

[3]Science, Technology & Innovation Policy, Georgia Institute of Technology, United States aporter@searchtech.com

Corresponding Email: mengjia.wu@uts.edu.au

ORCID: 0000-0003-3956-7808 (Mengjia Wu); 0000-0002-7731-0301 (Yi Zhang); 0000-0001-9227-5156 (Mark Markley); 0000-0002-5506-8032 (Caitlin Cassidy); 0000-0003-2717-1267 (Nils Newman); 0000-0002-4520-6518 (Alan Porter).

Abstract

Covid-19 has been an unprecedented challenge that disruptively reshaped societies and brought a massive amount of novel knowledge to the scientific community. However, as this knowledge flood has surged, researchers have been disadvantaged by not having access to a platform that can quickly synthesize rapidly emerging information and link the expertise it contains to established knowledge foundations. Aiming to fill this gap, in this paper we propose a research framework that can assist scientists in identifying, retrieving, and understanding Covid-19 knowledge from the ocean of scholarly articles. Incorporating Principal Component Decomposition (PDC), a knowledge model based on text analytics, and hierarchical topic tree analysis, the proposed framework profiles the research landscape, retrieves topic-specific knowledge and visualizes knowledge structures. Addressing 127,971 Covid-19 research papers from PubMed, our PCD topic analysis identifies 35 research hotspots, along with their correlations and trends. The hierarchical topic tree analysis further segments the knowledge landscape of the whole dataset into clinical and public health branches at a macro level. To supplement this analysis, we also built a knowledge model from research papers on vaccinations and fetched 92,286 pre-Covid publications as the established knowledge foundation for reference. The hierarchical topic tree analysis results on the retrieved papers show multiple relevant biomedical disciplines and four future research topics: monoclonal antibody treatments, vaccinations in diabetic patients, vaccine immunity effectiveness and durability, and vaccination-related allergic sensitization.

Keywords

Covid-19; topic analysis; knowledge retrieval; intelligent bibliometrics

1. Introduction

Since the Covid-19 outbreak in 2020, scientists around the globe have published more than 200,000 research papers on the nature of this virus and ways to help mitigate its negative impacts. While beneficial, the sheer volume of information published has caused an information crisis (Chahrour et al., 2020; Xie et al., 2020). Apart from the problem of misinformation and rumors, the overwhelming influx of research papers has resulted in severe information overload, challenging scientists, healthcare professionals, and the general public to 1) keep up with the rapid accumulation of new knowledge; 2) accurately and comprehensively obtain knowledge on specific topics; and 3) understand the new knowledge emerging (Hossain, 2020; Wise et al., 2020; Yu et al., 2021). Although several open literature datasets and search tools are available online (Trewartha et al., 2020; Zhang et al., 2020a), a comprehensive framework incorporating effective analytical tools is still missing to help scientists meet these challenges. What is needed is a solution that can help researchers answer the following three questions:

- Question 1 (Q1): What are the key research topics in the emerging Covid-19 knowledge system?
- Question 2 (Q2): How can we retrieve established knowledge for specific Covid-19 research topics?
- Question 3 (Q3): How do we understand and utilize the retrieved knowledge?

Previous efforts to answer these questions mostly consist of Covid-19 topic analysis (Colavizza et al., 2021; Pourhatami et al., 2021; Tran et al., 2020; Zhang et al., 2021a), literature-based discovery studies (Wise et al., 2020; Wu et al., 2021b; Yu et al., 2021), and literature search tools (Chen et al., 2021; Trewartha et al., 2020). A common approach in current topic studies is to apply co-word clustering (Pourhatami et al., 2021) or topic modeling (Colavizza et al., 2021; Tran et al., 2020) to the Covid literature. Such studies have helped to track newly emerging knowledge but have often overlooked the relationships between new evidence and previously established coronavirus knowledge. For example, what are the similarities and differences between the diagnosing criteria, treatments, and social responses for SARS and SARS-CoV-2? In such cases, established knowledge can be a significant means of discovering and synthesizing new knowledge (Haghani & Bliemer, 2020; Haghani & Varamini, 2021; Hu et al., 2021; Petrosillo et al., 2020). In addition, current literature-based discovery studies are conducted on macro levels and do not focus on specific knowledge domains to discover targeted knowledge for people pursuing different interests (Wise et al., 2020; Yu et al., 2021). For example, the interests of virologists and pathologists lie in tracing the spike protein mutations of SARS-CoV-2 (Pairo-Castineira et al., 2021; Starr et al., 2021), while clinical doctors are eager to follow the latest progress in diagnosis and treatment (Felsenstein et al., 2020; Merrill et al., 2020). For this reason, combining topic analysis and literature-based discovery approaches is a promising way to fill these two gaps. Further, few of the available Covid-19 knowledge search tools provide visualizations or other efficient ways to assist users in understanding the retrieved results (Trewartha et al., 2020; Zhang et al., 2020a). A concise and appropriate visualization could save a huge amount of time in finding the right papers to follow or in narrowing down their search scope. Aiming to fill these research gaps, we developed a research framework that provides a systematic solution to answering the three cited research questions.

Q1 is answered via a strategy that involves two topic extraction methods, Principal Component Decomposition (PCD) (Watts & Porter, 1999; Watts et al., 1999) and hierarchical topic tree (HTT) analysis (Wu & Zhang, 2021). This approach identifies research topics from research papers and yields a bird's eye view of Covid-19's knowledge system. Compared to other topic extraction approaches like K-means text clustering (Wartena & Brussee, 2008) or topic modeling (Blei et al., 2003; Yau et al., 2014), PCD can generate robust document clustering results without introducing any randomization processes. HTT, on the other hand, profiles the research topics in a hierarchical structure to highlight the differences and inner connections between topics. The two topic profiling approaches complement each other in generating both macro-level knowledge overviews and meso-level knowledge hierarchies. With the Covid-19 topics identified, we further developed a document retrieval approach based on a knowledge model that supports topic-specific document retrieval. The approach parses the entire PubMed database and links each identified topic with semantically similar pre-Covid literature in PubMed. In this way, new knowledge is linked to foundational knowledge. Q2 is answered by combining the topic analysis with the results of the knowledge model. Targeting Q3, the focus is on hierarchy, a specific dimension of knowledge composition, where the hierarchical structures of a topic's knowledge body are profiled and visualized. This helps researchers to efficiently understand the knowledge structures in the retrieved papers, further supporting knowledge discovery. All in all, this study blends multiple data-driven bibliometric approaches to reveal insights into Covid-19 knowledge deconstruction, effective retrieval, and understanding. It is in line with the direction what we called "intelligent bibliometrics" (Zhang et al., 2020b), targeting problems in science, technology, and innovation (ST&I) studies and highlighting the development of computational models incorporating artificial intelligence and data science techniques with bibliometric indicators. Despite a specific focus on Covid-19 knowledge in this paper, the proposed framework is adaptable for knowledge deconstruction and retrieval in broad domains and scenarios.

To conduct our case study, we collected the abstracts of 127,971 research papers published in 2020 and 2021 from the PubMed database. Feeding those papers into the PCD analysis, we generated 35 PCD topics and revealed how the emphasis on different topics changed over different periods. In the beginning, the focus was on the epidemiological and clinical characteristics of the virus. However, over time the emphasis changed to the impacts of Covid-19 on different societies. The HTT results divided the explored knowledge into a clinical branch and a public health branch. The clinical branch focuses on discovering Covid-19-associated clinical factors and treatments. The public health branch addresses six particular public health concerns. Additionally, we constructed a knowledge model based on the most popular PCD topic of *vaccination* and ran a global search on PubMed for records published prior to 2020 to retrieve the knowledge foundations of this topic, resulting in 92,286 retrieved papers. Lastly, we used HTT to visualize the knowledge structures of the retrieved results. The HTT results highlighted multiple vaccination-related disciplines, including immunology, molecular biology, virology, etc. From the branches of those disciplines, we identified four future promising research directions: monoclonal antibody treatments, vaccination in diabetic patients, vaccination effectiveness in SARS-CoV-2 antigenic drift, and vaccination-related allergic sensitization. We empirically evaluated the results by matching evidence identified from the literature and identified research evidence in the latest studies. This empirical case not only demonstrates the reliability of our method, but also derives insights to support potential COVID-related R&D and strategic management for funding agencies, individual researchers, and institutions.

2. Literature review

2.1. Covid-19 topic analysis

Topic analysis has been applied in substantial bibliometric studies to facilitate knowledge profiling and retrieval (Begelman et al., 2006; Kajikawa et al., 2022; Mejia et al., 2021). Scholars cluster semantically similar text (e.g., a collection of documents or similar terms) as topics and develop topic analysis approaches with different foci, including topic identification (Small et al., 2014), tracking (Zhang et al., 2017), and visualization (Huang et al., 2014). Along with the rapid accumulation of Covid-19 studies, bibliometricians have started analyzing relevant literature to follow the latest research progress. However, Covid-19 stands out as a unique task because the unprecedented amount of emerging knowledge it brings is not only closely related to the established knowledge foundation but also rapidly reshaping a new knowledge structure. Hence, identifying the links between 'new' and 'old' knowledge becomes a significant task in Covid-19 knowledge profiling and retrieval. Early-stage bibliometric analysis presents descriptive analyses of country-level research productivity (Chahrour et al., 2020), supporting sources (Nasab & Rahim, 2020), collaborating dynamics (Cai et al., 2021; Fry et al., 2020), and citing patterns (Hossain, 2020; Kousha & Thelwall, 2020). Apart from these efforts to measure research activity, uncovering new knowledge from the rapidly accumulating literature, i.e., literature-based discovery, is becoming a more important task as such insights can support research and clinical and policy decisions (Hristovski et al., 2005; Swanson, 1986; Wu et al., 2021c). Following the literature-based discovery stream, Pourhatami et al. (2021) use co-word analysis to identify past coronavirus-related topics, pointing out promising research gaps in antibody-virus interactions, emerging infectious diseases, and coronavirus detection methods. Yu et al. (2021) apply entity metrics on an entity network extracted from the literature, highlighting ACE-2 and C-reactive protein as significant biomarkers and chemicals in diagnosing and treating Covid-19. Similar findings were also reported by Wu et al. (2021b), through network analysis on biomedical entities extracted from Covid-19 literature, with more significant biomarkers, drugs, and complications identified. Ebadi et al. (2021) applied machine learning approaches to different Covid publication sources and compared the highlights and differences in research topics. These literature-based discovery studies provide substantial evidence of explicit and implicit knowledge associations from the extant research and insights that inspire deeper explorations in the future.

2.2. Covid-19 knowledge retrieval

As mentioned, the Covid-19 pandemic has induced a serious information crisis (Xie et al., 2020). The influx of research papers has resulted in serious information overload and is stopping users from efficiently retrieving specific knowledge out of the sea of information. To alleviate this situation, multiple research institutions, communities, and companies have released several curated Covid-19 literature datasets and search tools to help scholars and the general public find relevant knowledge efficiently. For example, the global literature on coronavirus disease[1] is a multiple-language literature collection curated by the World Health Organization (WHO). Users can search publications from multiple sources worldwide based on annotated health science descriptors. LitCovid (Chen et al., 2021) is a PubMed-derived dataset that allows users to retrieve Covid-19 publications relevant to eight broad, high-level topics. This dataset was classified by a BioBERT-based deep learning model, and the eight topics include general information, mechanism, transmission, diagnosis, treatment, prevention, case report, and epidemic forecasting. However, despite having topic retrieving features, the WHO dataset and LitCovid only focus on post-2020 publications related to Covid-19 or SARS-CoV-2. Therefore, any relationships between the new Covid-19 papers and previously established knowledge on human coronaviruses are not considered (Haghani & Bliemer, 2020; Haghani & Varamini, 2021). That said, there are a few studies that comprehensively compare the current Covid-19 with previous coronaviruses, demonstrating a significant approach to understanding the clinical, epidemiological, and pathological features of Covid-19 (Hu et al., 2021).

There are also available datasets and search tools with a broader literature coverage. CORD-19 (Wang et al., 2020c) is one of the most popular Covid-19 literature datasets, which assembles publications and preprints on Covid-19 and relevant historical coronaviruses like SARS and MERS. Several search tools have also been developed based on CORD-19. For instance, Covid-Scholar (Trewartha et al., 2020) is a document search engine that integrates text mining and natural language processing techniques, including keyword extraction and document ranking. Additionally, it can visualize the pretrained embedding space of keywords and present the global semantic similarity web of this domain. The Neural Covidex (Zhang et al., 2020a) is another search system that uses the T5-based language model (Raffel et al., 2019) finetuned on biomedical text to apply unsupervised reranking on retrieved documents. It supports natural language questions, such as search queries, which makes it more like a question-answering system. There are also several industry-backed search tools, such as the AWS CORD-19 search engine[2] from Amazon and the Azure CORD-19[3] search from Microsoft. However, a common drawback of those search tools is that the search results are presented in a long article list format, which is extremely inefficient in conveying a global understanding of the knowledge delivered by all the articles.

3. Data and methods

The research framework is illustrated in Figure 1. Three research trajectories are designed to answer the three proposed questions. The first trajectory is to use PCD and HTT analysis to profile the research landscape of Covid-19 studies. Following this, a topic-specific knowledge model and semantic-based search compose the second trajectory, providing a document retrieval approach for identified research topics. Lastly, HTT analysis is exploited to reveal knowledge hierarchies of the search results, wrapping the last trajectory as a solution for understanding and using the retrieved knowledge.

[1] https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/
[2] https://aws.amazon.com/marketplace/pp/prodview-ybwpxcqlznbas
[3] https://docs.microsoft.com/en-us/azure/open-datasets/dataset-covid-19-open-research

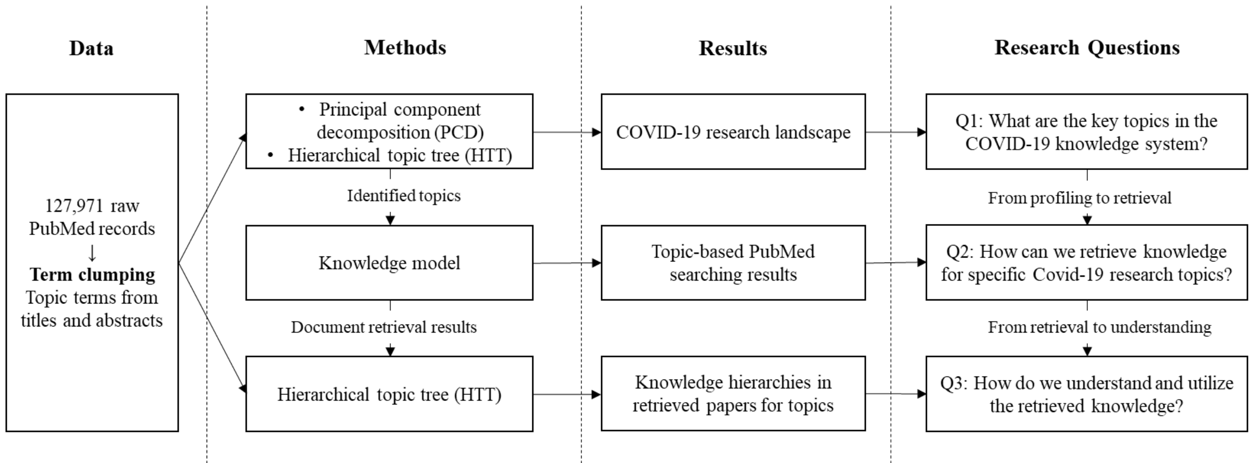

Figure 1. Research framework of the Covid-19 knowledge deconstruction and retrieval

3.1. Data collection and preprocessing

To collect Covid-19 bibliographic data, we compared multiple data sources in a pilot study (Porter et al., 2020) and ultimately decided to use PubMed, the globally largest and most comprehensive open-source biomedical database. Compared to other datasets that may have broader coverage on preprints and WHO materials (e.g., CORD-19), PubMed offers mostly peer-reviewed articles and affiliated curated metadata for our project analysis (e.g., MeSH Descriptors and Qualifiers). Lags in assignment of MeSH categories reduces their utility in the wildly accelerating COVID-19 research literature. The search process returned 127,971 relevant research papers from 1 Jan 2020 – 1 Jan 2022 as of early March 2022[4]. We further applied the natural language processing function of VantagePoint[5] to extract topic terms from titles and abstracts. The list of extracted terms was cleaned to remove stop words, consolidate similar terms, and eliminate all terms appearing only once (Zhang et al., 2014). The term clumping process and stepwise results are given in Table 1.

Table 1. Stepwise results of term clumping

Step	Detail	# Terms
1	Extract terms from titles and abstracts using VantagePoint NLP function	1,603,542
2	Remove terms starting/ending with non-alphabetic characters Remove common terms in scientific articles, e.g., "research framework" Remove meaningless terms, e.g., pronouns, prepositions, and conjunctions Consolidate synonyms based on expert knowledge, e.g., "Covid-19" and "Covid" Consolidate terms with the same stem, e.g., "severe patient" and "severe patients"	1,367,374
3	Filter terms with frequency above 10	33,281

3.2. Principal component decomposition (PCD)

PCD is essentially a robust and reproducible variant of principal components analysis (PCA) that groups scientific documents according to their textual features (Watts & Porter, 1999; Watts et al., 1999). Compared to the original

[4] We continue to update our Covid-19 PubMed research profile at: https://searchtechnology.github.io/VPDashboard/VantagePoint/Dashboard/1
[5] VantagePoint is a software platform for bibliometrics-based text analytics and knowledge management, owned by Search Technology Inc. More details can be found at the website: www.thevantagepoint.com.

PCA, PCD automatically decides the number of factors (derived PCA groupings) by minimizing the entropy and maximizing the cohesiveness of the derived factor groups. In our case, we exploited processed terms extracted from the titles and abstracts as document feature vectors. We then ran PCD on the document-term matrix to decide the factors automatically. The retained factors were deemed to be PCD topics.

3.3. Knowledge model-based document retrieval

The aim of knowledge model-based document retrieval is to retrieve scientific documents that have high semantic similarities with a given collection of textual data. Initially, we constructed a knowledge model containing a subset of relevant papers and their corresponding topic terms with the top and bottom 50 terms with the highest and lowest average term frequency-inverse document frequency (TF-IDF) constituting the feature vector of this set. We then used the knowledge model to search for related documents across all records in the PubMed database prior to 2020. At the time of retrieval, this amounted to over 30 million research papers. Lastly, the search results were ranked according to semantic similarity and the annual citation count. This formed the final output. The stepwise implementation is listed as follows:

Step 1: Select a specific PCD topic T and denote the set of grouped records as D_T, the entire corpus is denoted as D.

Step 2: Extract stems of all scientific terms in D_T and calculate the TF-IDF value for each stem using the following formula (Salton & Buckley, 1988):

$$\tau(t) = \frac{f_{t,D_T}}{\sum_{t' \in D_T} f_{t',D_T}} \log\left(\frac{|D|}{|t \in D|}\right)$$

where t denotes the stem of a scientific term and $\tau(t)$ is the TF-IDF value of t. f_{t,D_T} is the stem frequency of t in D_T, $|D|$ represents the total number of documents in D and $|t \in D|$ denotes the total number of documents in D that contains stem t.

Step 3: Construct the feature space V_T of topic T using terms with top and bottom 50 average TF-IDF values, for which we call V_T the knowledge model of T.

$$\overrightarrow{V_T} = [\tau(t_{\downarrow 1}), \tau(t_{\downarrow 2}), \dots, \tau(t_{\downarrow 50}), \tau(t_{\uparrow 1}), \tau(t_{\uparrow 2}), \dots \tau(t_{\uparrow 50})]$$

where $t_{\downarrow n}$ denotes the nth stem in TF-IDF descending order and $t_{\uparrow n}$ denotes the nth stem in TF-IDF ascending order.

Step 4: Construct and align the feature space for each document in the entire PubMed database by Steps 2 and 3, the aligned feature space of document d is denoted as $\overrightarrow{V_d}$. When calculating the IDF values for PubMed historical papers, we still adopted the original corpus as the total document set considering easier feature alignment and better algorithm scalability.

Step 5: Calculate the cosine similarity (Salton & McGill, 1986) of $\overrightarrow{V_T}$ and every $\overrightarrow{V_d}$; the records with similarity above a threshold ε are returned as a first pass, denoted as D_{TR}.

$$Cosine\left(\overrightarrow{V_T}, \overrightarrow{V_d}\right) = \frac{\overrightarrow{V_T} \cdot \overrightarrow{V_d}}{\left|\overrightarrow{V_T}\right|\left|\overrightarrow{V_d}\right|}$$

Step 6: Remove documents in D_{TR} that contain any of the terms in the PCD factor label of T.

Step 7: Rank the remaining publications in D_{TR} by the harmonic mean of cosine similarity and number of citations per year since publication, scaled between 0 and 1.

3.4. Hierarchical topic tree (HTT)

Hierarchical topic tree analysis (Wu & Zhang, 2021) is a network-based method that identifies research topics from scientific documents in a hierarchical way. From a co-term network as input, this method identifies term nodes with 1) notably high density; and 2) relatively far distance to other high-density nodes as community centers. Then each non-central node is assigned to its proximate community center to compose node communities (research topics). The subgraphs of the partitioned communities will serve as the next round of input for this process until no can community centers be found in the input graph. The community results of each iteration constitute a topic layer of the topic tree, with the community center denoting leaf labels. The finalized output is a hierarchically partitioned co-term network that represents the intellectual structure of a knowledge system (Wu et al., 2021a; Zhang et al., 2021b). The stepwise processes of this method are given below:

Step 1: Construct the co-term network from the documents and calculate the shortest distances of pairwise nodes.

$$G = (V, E)$$

where G is the co-term network, V is the set of term nodes and E is the set of co-occurrence edges.

$$w_{E_{ij\,(i \neq j)}} = \begin{cases} CF(V_i, V_j) & if\ V_i\ and\ V_j\ co-occur\ in\ at\ least\ one\ document \\ 0 & otherwise \end{cases}$$

where $w_{E_{ij\,(i \neq j)}}$ is the edge weight of the edge connecting nodes V_i and V_j, $CF(V_i, V_j)$ represents the co-occurring weight (number of documents that V_i and V_j co-occur in) of nodes V_i and V_j.

Step 2: Calculate the neighborhood density for each node and generate the shortest distance of every node to its closest node with a higher neighbor density. Considering the scalability of our algorithm on this network, we used neighborhood density as a proxy for the density measures of each node.

$$\rho_{V_i} = \exp\left(-\frac{1}{\Gamma(V_i)} \sum_{V_j \in \Gamma(V_i)} \frac{1}{{w_{E_{ij}}}^2}\right)$$

where ρ_{V_i} denotes the local density of node V_i, and $\Gamma(V_i)$ is the neighbor node set of V_i.

$$\delta_{V_i} = \begin{cases} \max\limits_{V_j \in \Gamma(V_i)} \left(\frac{1}{w_{V_i V_j}}\right) & if\ \rho_{V_i} = \max\limits_{V_j \in \Gamma(V_i)} (\rho_{V_j}) \\ \min\limits_{V_j \in V_{\rho_{V_j} > \rho_{V_i}}} \left(\frac{1}{w_{V_i V_j}}\right) & otherwise \end{cases}$$

where δ_{V_i} is the shortest distance from V_i to its closest neighbor node with larger local density.

Step 3: Locate the set of community centers that meet the density peak criterion (the formula below) and denote it as V_c; Then initialize them as community centers and allocate the rest of the nodes to the nearest node in V_c.

$$\rho_{v \in V_c} > \varepsilon \max_{V_i \in \Gamma(v)} \rho_{V_i}$$

where ε is the density threshold that decides the significance of a topic. This was set to 0.95 by default.

Step 4: Iterate Step 3 with the partitioned communities until no centroid node can be found in any sub-community. From the second iteration, an additional criterion is added to guarantee the identified centroids for the sub-communities are spread sparsely enough from each other:

$$\delta_{V_c} > \frac{1}{w_{V_r V_c}}$$

where V_r denotes the node centroid of its parent community.

Ultimately, by applying Steps 1-4 to the co-term network, a set of hierarchical communities emerges. Communities (subgraphs) partitioned on different levels constitute different layers of topics. The label of the centroid node is used as the label for a community in the HTT visualizations.

4. Results

4.1. Data overview

Trends in Covid-19 publications can help us glimpse the response patterns of the scientific community given catastrophic events. To capture such trends, we first applied a descriptive bibliographic analysis to profile the external characteristics of Covid-19 studies regarding the monthly growth, institution ranking, and geographical distribution.

Figure 2 illustrates the basic monthly numbers of Covid-19 research papers. Early in 2020, these numbers were rapidly increasing, but by 2021, they had become relatively steady. The burst of Covid-19 publications can easily be attributed to the disruptiveness and uncertainty that Covid-19 has brought to previously established knowledge systems (Zhang et al., 2021a), which attracts research attention from massive new researchers (Wagner et al., 2022). However, the slowing increase might be due to multiple reasons: Is it due to research capacity limitations (e.g., journals, review speed, funding, etc.)? Or does it indicate that newly discovered knowledge is converging to a new stage? Will there be a decay period following? These possibilities only trigger more research questions to be answered and examined in future studies.

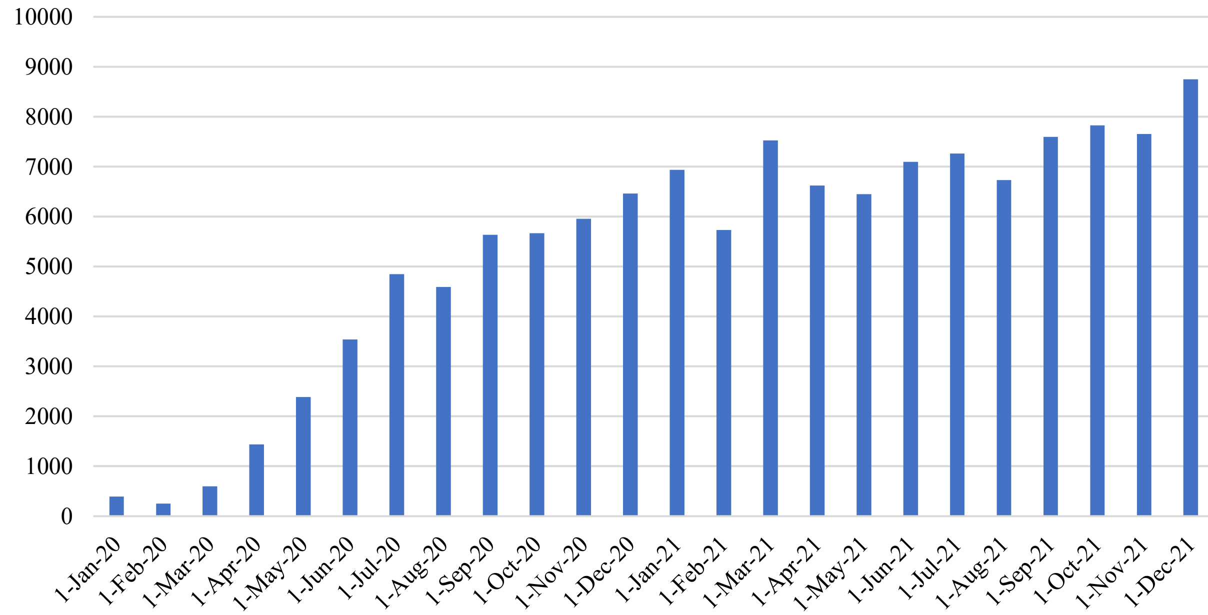

Figure 2. Monthly increasing trend of Covid-19 research papers

Figures 3 and 4 respectively profile the global distribution and ranking changes of Covid-19 research papers among worldwide countries/regions. Figure 5 lists the top 20 productive institutions. In terms of the absolute number of papers published at the national level in Figure 3, the United States and China unsurprisingly hold leading positions, followed by Italy, India, Germany, Canada, etc. From a retrospective view, the ranking changes in Figure 4 intuitively indicate the association between productivity and local epidemic severity (Wagner et al., 2022). For example, China took the first place in the initial few months because it was the first victim of Covid-19 and had first-hand access to massive numbers of clinical cases. However, the US soon overtook China and has held the first position since the middle of 2020. This may be because the US has solid research strength, but it could also be the result of how severely the Covid-19 pandemic hit the US (Burki, 2020). Italy maintained third place for a long time from March 2020 as it became the European Covid-19 epicenter, suffering high numbers of cases and mortality rates (Remuzzi & Remuzzi, 2020). Following a sharp decrease in March 2020, which could be a result of the 21-day nationwide lockdown at that time, India has remained high in the ranking list. The pandemic hit India severely, and multiple SARS-CoV-2 variants have emerged there (Bernal et al., 2021).

Diving into the institution level, we found that, compared to the earlier China-led trends in Covid-19 research (Fry et al., 2020), the momentum for US institutions to lead in this domain has continued to grow (Wu et al., 2021b). This indicates that even though China has published a substantial volume of papers, individual Chinese universities and research institutions have not demonstrated equal strength in competition with their global counterparts, particularly those from the States.

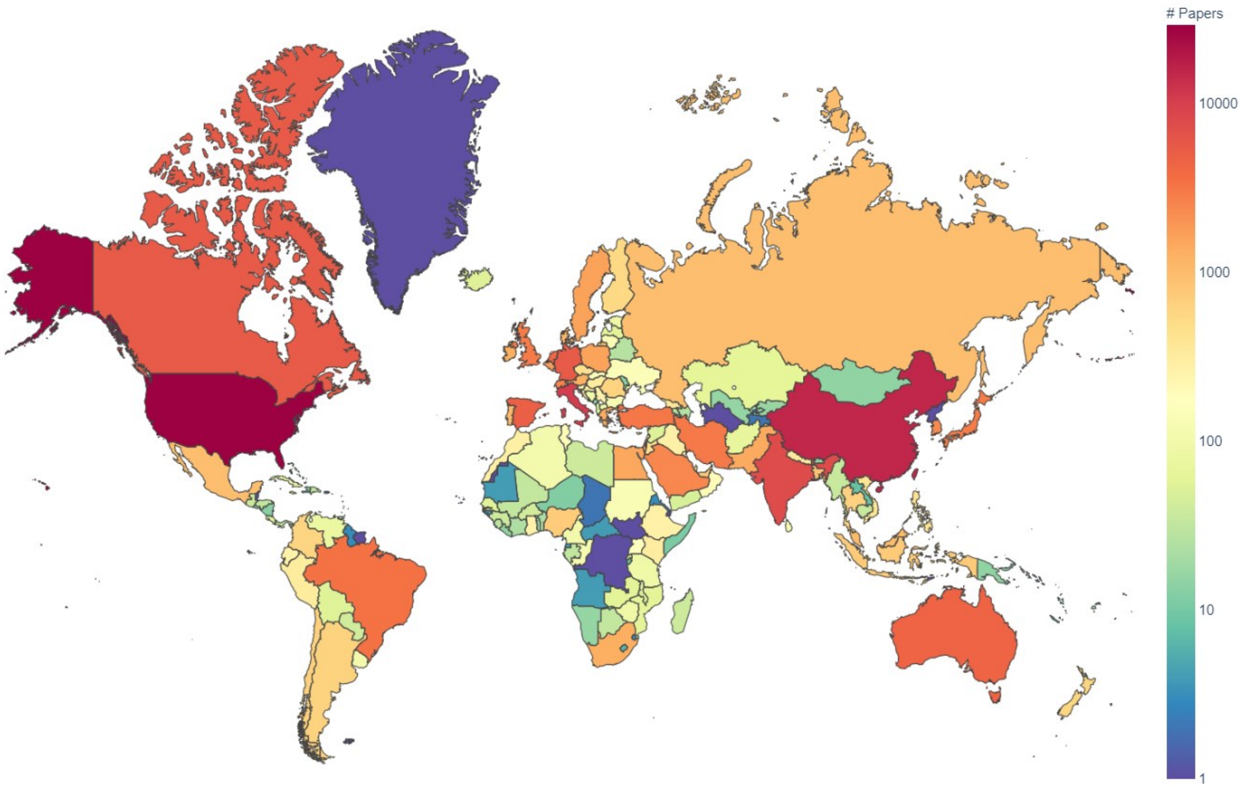

Figure 3. The geographical distribution of Covid-19 papers

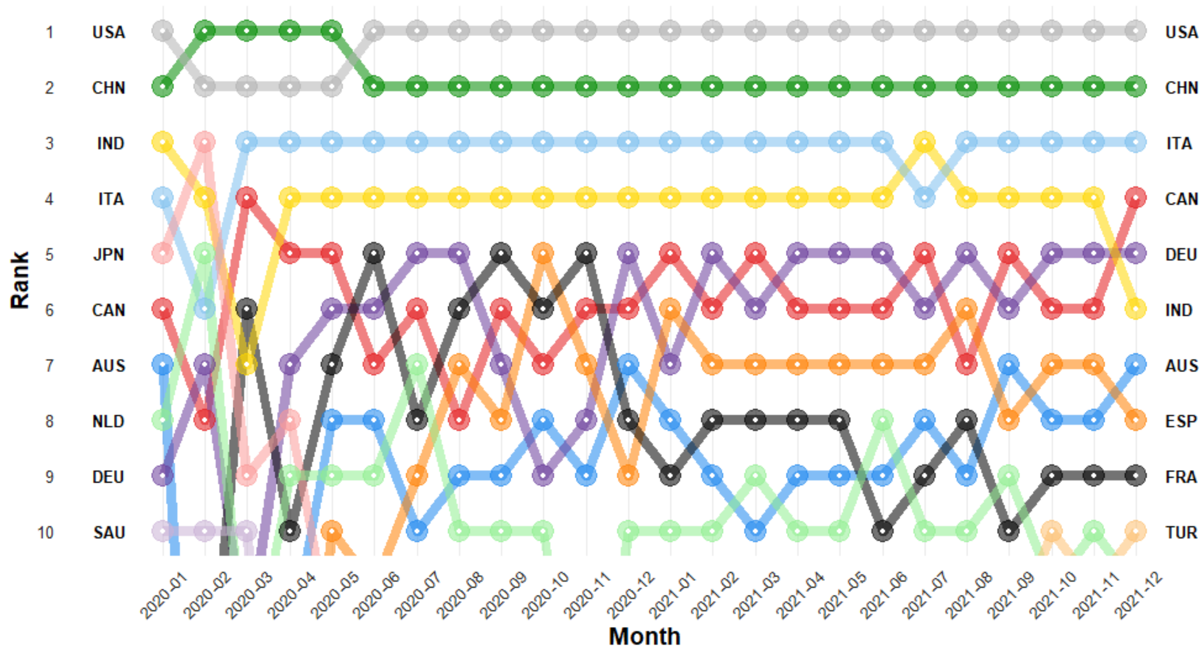

Figure 4. The ranking changes of countries

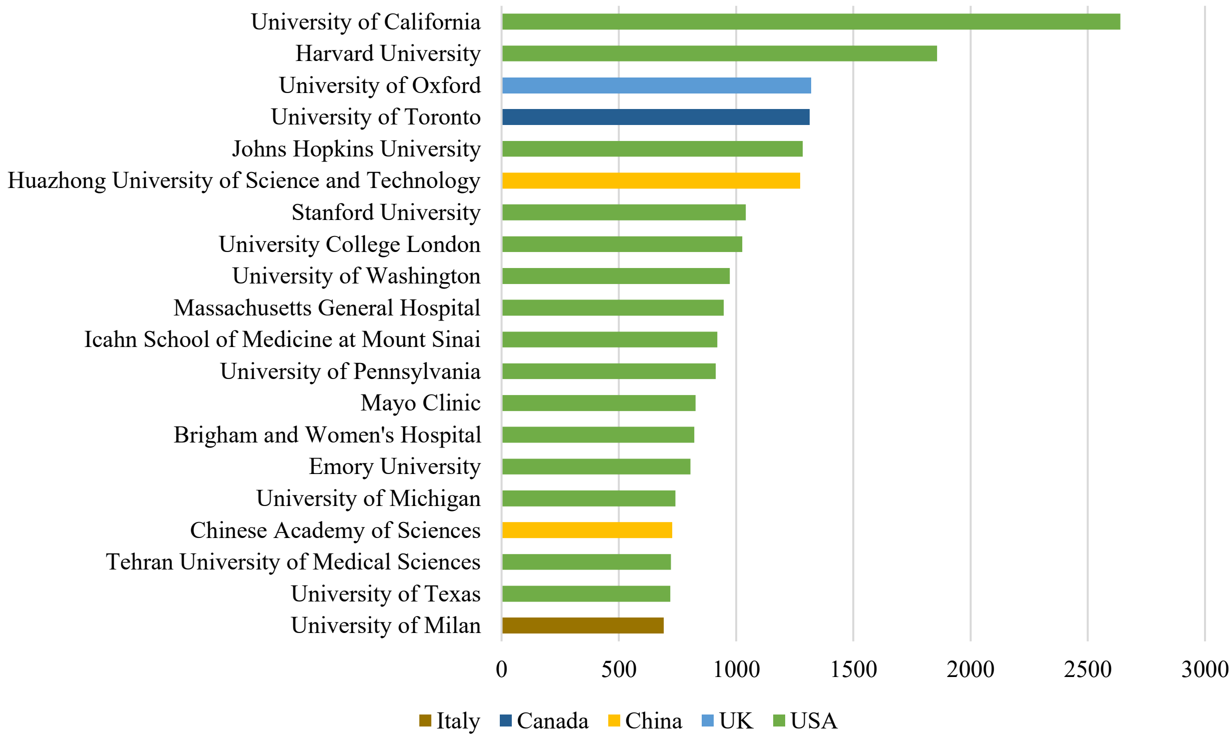

Figure 5. Top 20 prolific research institutions

4.2. Research landscape of Covid-19

Feeding the extracted topic terms into the PCD analysis, we distilled 35 research topics[6]. Further, we plotted a topic correlation map in Figure 6, with each bubble representing a PCD research topic and the size denoting its associated paper count, and the links denoting a cosine correlation above 0.5 (Salton & McGill, 1986). The correlation map of the 35 topics highlights a core topic cluster in the middle, representing a set of clinical manifestations and hospitalization factors of Covid-19. The other scattered topics cover a broad range, including public health, education, economics, etc. More details are provided on those topics in the following analysis.

[6] The terms contained in each topic are available at https://github.com/IntelligentBibliometrics/Covid_knowledge/blob/main/Topic_terms.xlsx

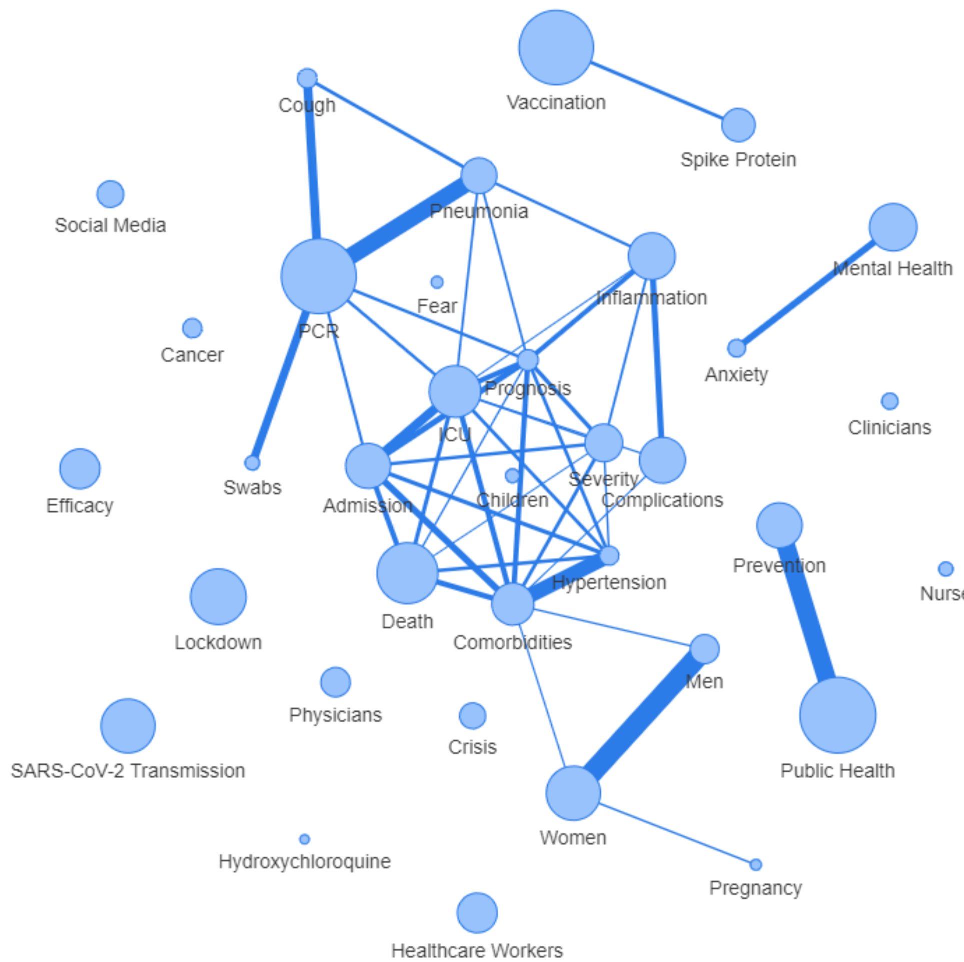

Figure 6. The distribution and cross-correlation of PCD topics

The monthly ranking changes of the top ten topics are given in Figure 7, indicating different stages of Covid-19 research. Among these topics, the rankings of *PCR* and *Public Health* maintain the top, while other topics show significant fluctuating trends.

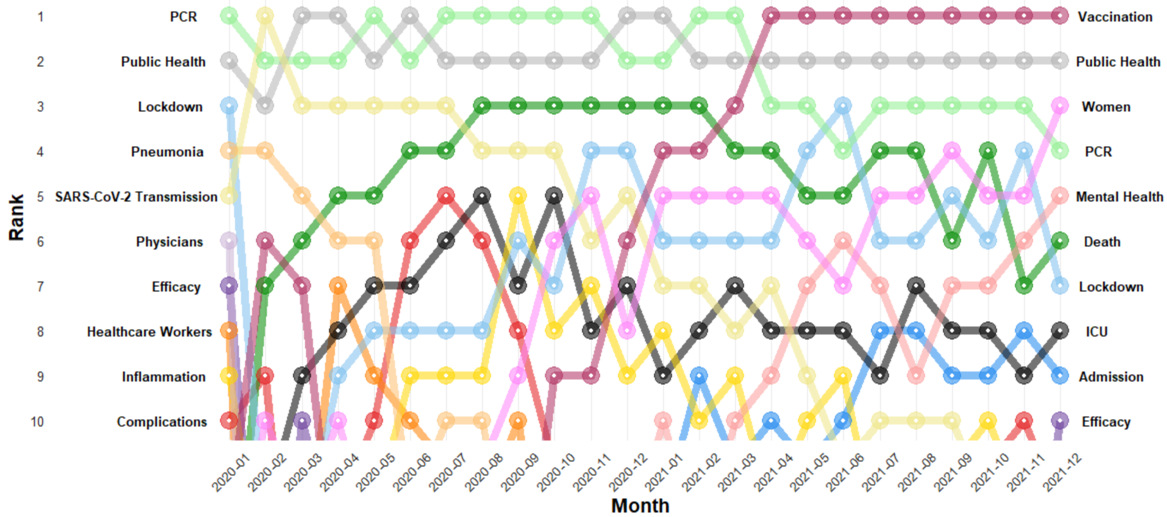

Figure 7. Monthly increasing trend of PCD topics

At the beginning of the Covid-19 breakout in Wuhan, the PCD topics *Pneumonia* and *SARS-CoV-2 Transmission* attracted massive attention, as first-hand clinical and epidemiological investigations were urgently needed to improve Covid treatments and control its transmission (Huang et al., 2020; Li et al., 2020b; Lu et al., 2020; Wang et al., 2020b). In such investigations and following clinical trials, the gender difference is an essential analyzing factor as indicated by the continuing ranking rise of PCD topics *Women* and *Men*, additional attention was put on the female group due to studies on the vulnerabilities of pregnant women or women at lactating ages (Chen et al., 2020). As Covid-19 turned from regional transmissions into a global pandemic, scientists started to look into the social impacts of Covid-19 as illustrated by the rise of topics *Lockdown* (Ruktanonchai et al., 2020; Shepherd et al., 2021) and *Mental Health*. The former broadly covers the social impacts of lockdown measures on healthcare services (Shepherd et al., 2021), economy (Bonaccorsi et al., 2020), education (Engzell et al., 2021), and environment (Venter et al., 2020), etc.; The latter topic discusses mental health issues among the general public (Brülhart et al., 2021; Shi et al., 2020) and healthcare workers (Lai et al., 2020). As the Covid pandemic progressed, rankings of the topics *Death* and *ICU* decreased relatively steadily.

Notably, the change in *Vaccination*-related papers illustrates two waves of vaccine studies. The first wave appeared at the beginning of the Covid-19 breakout and peaked in February 2020. These early-stage papers mainly focus on reviewing past coronavirus vaccines, calling for rapid vaccine development procedures, and proposing possible vaccine development approaches (Ahmed et al., 2020; Ahn et al., 2020; Pang et al., 2020; Prompetchara et al., 2020). With the advent of multiple available vaccines, the next wave emerged in the third quarter of 2020 and continued to rise. In addition to the massive numbers of basic medicine and clinical trial studies around the safety and efficacy of those vaccines (Polack et al., 2020; Thomas et al., 2021; Xia et al., 2020), the rollout of vaccines also triggers researchers' concerns about the social implications, including the vaccine hesitancy phenomena (Biswas et al., 2021; Dror et al., 2020), vaccine allocation strategies (Duch et al., 2021) and vaccination incentives (Campos-Mercade et al., 2021; Dai et al., 2021). As vaccination offers one of the most effective measures in preventing Covid-19, we will demonstrate how we used our knowledge model to retrieve historical knowledge of vaccination studies in the next section.

The PCD results yield a flat view of the Covid-19 research landscape. To dive further into the hierarchy of Covid-19 knowledge and obtain more details about research topics, we ran the HTT algorithm and constructed a co-term network using terms with a frequency of above 10. The characteristics of our input network are given in Table 2. The generated

HTT map is shown in Figure 8, with the tree trimmed to only show nontrivial branches[7]. The node size indicates the prevalence of the topic, and the edge thickness denotes the co-occurring strength of the two connected topics.

Table 2. The characteristics of Covid-19 term co-occurrence network

	Number	Weight			
		Max	Min	Avg.	Std.
Node	33,281	14,944	10	45.448	237.106
Edge	7,504,641	3,618	1	1.568	6.246
Average degree			450.980		

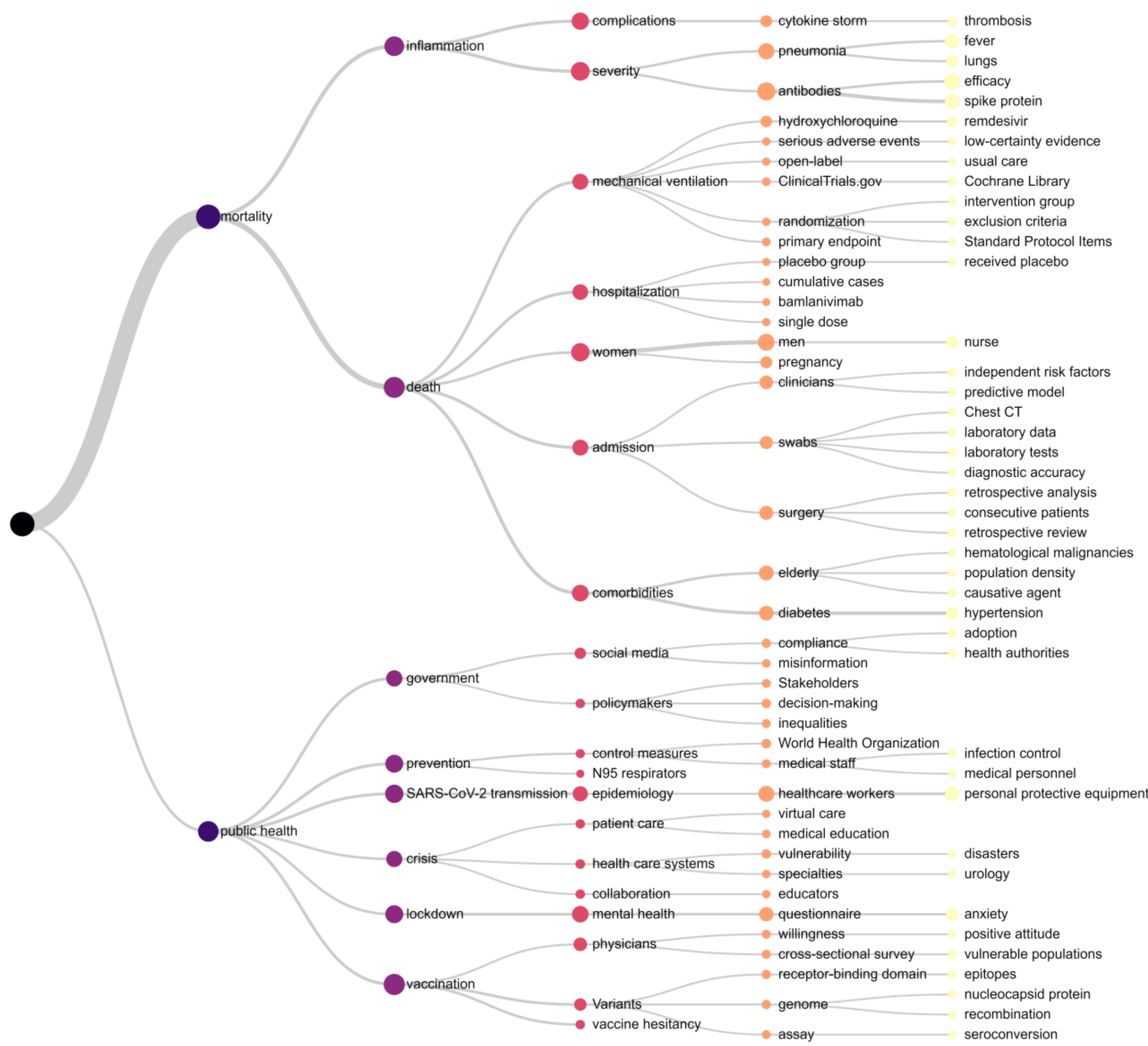

Figure 8. The hierarchical knowledge landscape of Covid-19 literature

[7] The entire HTT can be found at https://github.com/IntelligentBibliometrics/Covid_knowledge/blob/main/HTT_overall.svg

The hierarchical topic tree shows more detail on every individual topic. The HTT map covers most PCD topics and arranges them hierarchically according to their topological importance in the co-occurrence network. This empirical evidence, discovered through PCD and HTT, coincides with knowledge manually identified from the literature, which might in some sense endorse the performance of the method. *Mortality* and *Public health* are two HTT topics that hold the top positions in this hierarchy and represent the two major research branches: clinical and public health studies.

The clinical branch spans efforts to uncover the clinical associated factors of Covid-19 and find effective therapies. As illustrated in Figure 8, such explored clinical factors include gender – *women*, *men* (Jin et al., 2020), complications – *inflammation*, *cytokine storm* (Jose & Manuel, 2020), *thrombosis* (Levi et al., 2020), age – *elderly* (Liu et al., 2020a), and comorbidities – *diabetes* (Muniyappa & Gubbi, 2020), *hypertension* (Fang et al., 2020). The treatments studied in clinical case reports and clinical trials consist of *mechanical ventilation*, *hydroxychloroquine* (Gautret et al., 2020), *remdesivir* (Beigel et al., 2020), and *bamlanivimab* (Gottlieb et al., 2021), etc. In all, this branch contains various clinical case reports and clinical trial studies devoted to revealing the associated factors of Covid-19 severity/mortality/prognosis and finding effective treatments.

For the public health branch, six subtopics are highlighted as follows. 1) *Government*: This branch discusses the role of government in fighting Covid-19. One of its subordinate branches points to *policymakers*, and, within this, handling *inequalities* in different groups of people has become a notable concern in the policymaking process (Chu et al., 2020; Garcia et al., 2021). The other subordinate branch of *social media* indicates the role of social media as a double-edged sword for governments when it comes to information dissemination and evaluation (Islam et al., 2020b; Li et al., 2020a; Tsao et al., 2021), given the presence of misinformation. 2) *Prevention*: This set of HTT topics reflects some of the major explorations of Covid -19 prevention being: face mask production and use issues (Brooks et al., 2021; Long et al., 2020; Wu et al., 2020); identifying effective control measures (Nussbaumer-Streit et al., 2020; Wang et al., 2020d); and how to protect frontline healthcare workers (Ding et al., 2020; Islam et al., 2020a). 3) *SARS-CoV-2 transmission*: This set of topics explores the epidemiological characteristics of Covid-19, among which the transmission between healthcare workers (Bergwerk et al., 2021; Sikkema et al., 2020) and the use of personal protective equipment (Mick & Murphy, 2020) have attracted substantial research attention. 4) *Crisis*: This topic set discusses the implications of Covid-19 on healthcare systems (Liu et al., 2020b; Spinelli & Pellino, 2020) and medical education (Hall et al., 2020). 5) *Lockdown*: As one of the strictest restrictions, lockdown measures were frequently explored for their associations with mental health issues in the general public and medical staff (Wang et al., 2020a; Williams et al., 2020), 6) *Vaccination*: Apart from one branch highlight the basic biomedical studies for vaccine development (Polack et al., 2020; Xia et al., 2020), the other two branches respectively address attention to vaccination in healthcare workers (Bergwerk et al., 2021) and the vaccination hesitancy issue (Dai et al., 2021; Machingaidze & Wiysonge, 2021).

4.3. Knowledge model search results: The case of 'Vaccination'

This section demonstrates the utility of our knowledge model approach in retrieving historical knowledge from the entire PubMed database, using the most prominent PCD research topic, *Vaccination,* as an example. We extracted 15,967 papers related to this topic and calculated the TF-IDF values of all the extracted terms of those papers. Then, a knowledge model was built up with its top 50 and bottom 50 term stems. The details of this knowledge model are given in Table 3.

Table 3. The knowledge model of topic *Vaccination*

Top				Bottom			
Stem	**TF-IDF (avg.)**	**Stem**	**TF-IDF (avg.)**	**Stem**	**TF-IDF (avg.)**	**Stem**	**TF-IDF (avg.)**
vaccin	0.0711	bnt162b2	0.0054	synchronis	4.13E-07	e31del	4.82E-07
antibodi	0.0289	case	0.0054	africain	4.26E-07	f888l	4.82E-07
immun	0.0138	protect	0.0053	d253g	4.38E-07	formerlygr	4.82E-07
neutral	0.0121	individu	0.0053	q954h	4.38E-07	h69del	4.82E-07
dose	0.0101	diseas	0.0052	s373p	4.38E-07	havevaccin	4.82E-07
variant	0.0097	popul	0.0051	s375f	4.38E-07	k848	4.82E-07
infect	0.0093	coronaviru	0.005	y505h	4.38E-07	l212i	4.82E-07
respons	0.0091	antigen	0.0049	voor	4.44E-07	n211del	4.82E-07
cell	0.009	efficaci	0.0049	andb	4.54E-07	n417	4.82E-07
mrna	0.0089	titer	0.0048	d796y	4.54E-07	n969k	4.82E-07
spike	0.0087	receiv	0.0047	f157l	4.54E-07	namelyepsilon	4.82E-07
protein	0.0082	report	0.0047	l981f	4.54E-07	q1071h	4.82E-07
test	0.0076	posit	0.0047	r190	4.54E-07	q19e	4.82E-07
patient	0.0075	serolog	0.0047	severityof	4.54E-07	r32del	4.82E-07
anti	0.0074	epitop	0.0047	t1027i	4.54E-07	s33del	4.82E-07
develop	0.0069	human	0.0046	thebeta	4.54E-07	s929i	4.82E-07
hesit	0.0065	syndrom	0.0046	v70del	4.54E-07	spathogen	4.82E-07
assai	0.0064	clinic	0.0045	variantwa	4.54E-07	tegallyet	4.82E-07
viru	0.0062	respiratori	0.0045	1092k	4.82E-07	theb	4.82E-07
bind	0.0059	trial	0.0044	156del	4.82E-07	thebind	4.82E-07
effect	0.0057	influenza	0.0044	157del	4.82E-07	thedelta	4.82E-07
viral	0.0057	mutat	0.0044	2020in	4.82E-07	theepsilon	4.82E-07
sever	0.0056	receptor	0.0043	351also	4.82E-07	thefus	4.82E-07
detect	0.0055	base	0.0043	7variantwa	4.82E-07	theheptad	4.82E-07
specif	0.0055	group	0.0043	a63t	4.82E-07	thep	4.82E-07

Further, we ran the knowledge model-based document retrieval over the entire PubMed database and retrieved 92,286 historical records out of the Covid dataset[8]. We removed records containing the /vaccin/ stem and empirically set the

[8] The full list of retrieved paper is at https://github.com/IntelligentBibliometrics/Covid_knowledge/blob/main/retrieved_papers.xlsx

cosine similarity retrieving threshold to 0.25. The next section demonstrates how to deconstruct the retrieved results and mine the knowledge structures.

4.4. Knowledge structure visualization

We further mapped the 92,286 records to Open Academic Graph (OAG)[9] and retrieved 89,951 records with the field of study (FoS) information. The FoS is essentially constituted by Wikipedia entities being assigned to scholarly papers via a Naïve Bayes-based tagging process (Shen et al., 2018). OAG originates from Microsoft academic graph (MAG) and currently covers more than 240 million publications. Compared to scientific terms extracted from titles and abstracts in Section 4.2, the FoS system adopts Wikipedia entity entries as the topics of each paper, which we consider more suitable for representing established knowledge foundations. To efficiently understand and visualize the knowledge in the search results, we constructed the FoS co-occurrence network of the 89,951 records and ran our HTT algorithm over it. The detail of the constructed network is given in Table 4.

Table 4. The characteristics of the FoS network

	Number	Weight			
		Max.	Min.	Avg.	Std.
Node	27,596	39,542	1	3.459	44.336
Edge	922,252	18,737	1	28.135	427.105
Average degree			66.840		

We trimmed this HTT to retain the main body of knowledge. This is presented in Figure 9[10], yielding a hierarchical overview of the search results. Immunology is the root topic of this HTT, indicating that vaccination studies are mostly constructed based on immunology knowledge. The presented topics are primarily highlighted as discipline-level topics (green font) and entity-level topics (red font). By comparing and contrasting the historical records (regarded as the knowledge foundations) with the latest research evidence, we drew the following insights on four significant topics: ***Monoclonal antibodies, Antigenic drift, Diabetes, Allergic sensitization.*** In the next section, we will detail the evidence to validate our findings empirically.

[9] https://www.aminer.cn/oag-2-1
[10] The entire HTT can be found at https://github.com/IntelligentBibliometrics/Covid_knowledge/blob/main/Vacc_all.svg

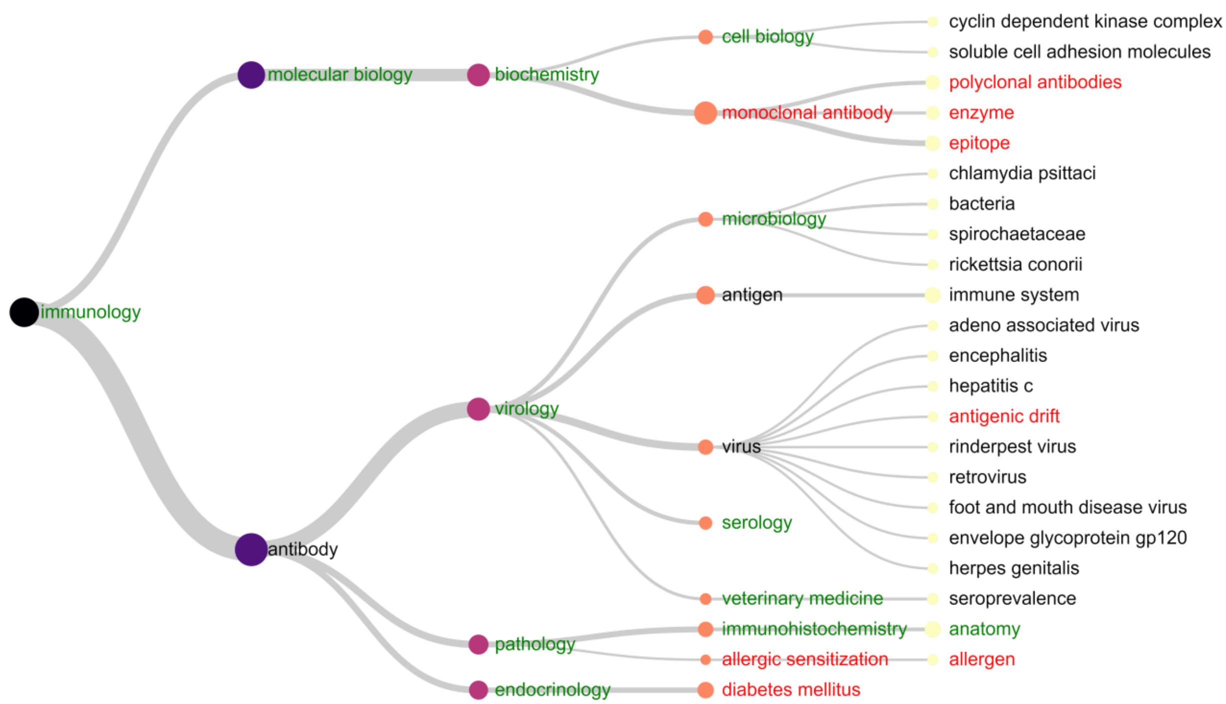

Figure 9. The hierarchical knowledge landscape of retrieved results

4.5. Empirical validation

Given that our three methods have been separately validated in pilot studies (Cassidy, 2020; Watts et al., 1999; Wu & Zhang, 2021) and Covid-19 research is still in the infant age, we validated our empirical results with literature-based evidence and dived into the historical papers and the newest Covid-19 research articles related to the four topics. The knowledge connections we identified from the papers are presented as follows:

- ***Monoclonal antibodies***: This topic is positioned in the branch of *molecular biology – biochemistry*. Diving into this topic, we can trace many historical studies on developing monoclonal antibodies as a treatment for existing human and animal coronaviruses, including severe acute respiratory syndrome (SARS) coronavirus (Traggiai et al., 2004; Zhu et al., 2007) and bovine coronaviruses (Deregt & Babiuk, 1987; Mockett et al., 1984). Such studies can provide instructive research clues for developing novel monoclonal antibody treatments for Covid-19. With the approval of multiple monoclonal antibody treatments for Covid-19, more efforts will predictably be put into finding efficient methods of extracting and producing such monoclonal antibodies (Taylor et al., 2021).
- ***Antigenic drift***: This topic exists in the *virus* branch, describing a natural phenomenon of antigen genetic mutations that also happens in the SARS-CoV-2 virus (Yuan et al., 2021). Medical experts can trace historical studies of influenza viruses (Pica et al., 2012; Yu et al., 2008) and other possibly related viruses (Coulson et al., 1985) in search results to infer and analyze the impacts of antigenic drift on vaccination implementations. The effectiveness and immune durability of current vaccines for various SARS-CoV-2 variants (including currently concerning Omicron) may need deeper exploration (Cameroni et al., 2022; Koyama et al., 2020).
- ***Diabetes***: Located in the endocrinology branch, this topic consists of historical papers clarifying the autoimmune-mediated beta-cell damage mechanisms (Van Belle et al., 2011), significant autoantigens (Wenzlau et al., 2007), and different subtypes of type 1 diabetes (Imagawa et al., 2000; Stenstrom et al., 2005). Recent studies report that two types of diabetes are associated with higher odds of Covid-19 hospital deaths (Barron et al., 2020; Holman et al., 2020), and SARS-CoV-2 infection possibly induces negative effects on beta-cells (Apicella et al., 2020; Bornstein et al., 2020; Lim et al., 2021; Marchand et al., 2020). Consequently, vaccination in diabetic patients has become a trending topic among vaccination studies. On

the one hand, many researchers have called for prioritizing vaccination in diabetic patients as they are more vulnerable to Covid-19 (Pal et al., 2021; Powers et al., 2021). On the other hand, associating the knowledge from our search results with Covid vaccinations (especially for Type 1 diabetes) is worth deeper exploration because the current evidence is still limited (Boddu et al., 2020; Marchand et al., 2020).

- ***Allergic sensitization***: Historical studies related to this topic comprehensively discuss the reactivity of immunoglobin E in allergic reactions (Aalberse et al., 2001; Eibensteiner et al., 2000; Jenmalm et al., 2001), which can provide instructive insights for Covid-19 vaccination allergy studies (Cabanillas et al., 2020; Kounis et al., 2021).

5. Discussion

Covid-19 has brought about a global public health pandemic and an overwhelming flood of research knowledge. Aiming to efficiently discover and use the knowledge contained in the massive body of Covid-19 scientific studies, we devised a research framework that: 1) profiles the Covid-19 knowledge landscape and research topics at both the flat and hierarchical levels; 2) retrieves the foundations of knowledge related to specific topics; and 3) visualizes the retrieved knowledge to support knowledge understanding and discovery. We anticipate that this research methodology and our key findings will support a) scientific researchers to quickly absorb emerging new knowledge and identify their future study directions, and b) research policymakers to make informed decisions about research funding allocations.

5.1. Key findings

Q1: What are the key topics of the emerging Covid-19 knowledge system?

We exploited PCD and HTT analysis to profile the Covid-19 research landscape. The PCD results highlight 35 research hotspots and multiple research emphases over different periods. The changing trends in PCD topic rankings indicate that early Covid-19 studies focus on uncovering the clinical and epidemiology characteristics of Covid-19, while the subsequent studies throw more light on the societal impacts of the pandemic. Intriguingly, the change in PCD topic *vaccination* papers reflects two waves of vaccination studies – the first appearing at the start of the Covid outbreak and the second after the rollout of multiple available vaccines. The HTT results consistently reveal clinical and public health studies as two major branches of research in this domain. Complementarily, the HTT results generate more detailed insights on 1) the clinically investigated factors associated with Covid-19 mortality/severity and effective treatments; and 2) six segments of public health concern: *government*, *prevention*, *SARS-CoV-2 transmission*, *crisis*, *lockdown*, and *vaccination*.

Q2: How can we retrieve established knowledge for specific Covid-19 research topics?

We developed a knowledge model based on text analytics to discover the foundations of the topic-specific knowledge and demonstrated the utility of this approach using the topic of *vaccination* in Section 4.3. Using this knowledge model, we conducted a global search of the entire PubMed database and retrieved 92,286 papers that hold high semantic similarities with records on this topic at the document level.

Q3: How do we understand and utilize the retrieved knowledge?

We ran our HTT algorithm over the search results from the knowledge model to reveal the hierarchies of topics. At the top levels of the HTT, we identified multiple significant medical disciplines, including *immunology*, *molecular biology*, *virology*, and so on. In addition to these disciplines, we also uncovered four directions worthy of more attention in future vaccination-related studies. These are 1) monoclonal antibody treatments, 2) vaccination priority and immune responses in diabetic patients, 3) the effectiveness and durability of vaccines on various SARS-CoV-2 mutations, and 4) vaccination allergies.

5.2. Technical implications

There are three methodological contributions of this paper worth highlighting. Initially, incorporating PCD topic analysis and knowledge model searches provides an effective topic-based mode of knowledge retrieval. This approach first clusters research documents into research topics and then searches the entire PubMed dataset for the foundational knowledge on the target topic, generating a more narrowed, focused, and accurate search scope in knowledge retrieval. Additionally, our HTT results offer a way for researchers to visualize and understand thousands of papers efficiently. By highlighting the topologically significant terms in the co-occurrence network, the HTT can help researchers quickly clarify complex knowledge structures and identify intriguing topics. Last, but not least, our research framework provides a paradigm for research profiling and knowledge retrieval. This methodology is adaptable to various cases and can be transferred with little cost.

From the practical perspective, this paper profiles the knowledge landscape of Covid-19 studies both in flat (PCD) and hierarchical (HTT) manners, yielding hotspots for researchers to follow. Furthermore, the insights offered in Section 4.4 identify four intriguing vaccination-related research directions. Such insights can: 1) inspire medical researchers to conduct future studies with enriched knowledge foundations; and 2) assist scientific policymakers in making informed decisions about research funding allocations.

5.3. Limitations and future directions

This study does come with limitations. From a methodological perspective, even though the three approaches exploited in this study are developed in our pilot studies (Cassidy, 2020; Watts et al., 1999; Wu & Zhang, 2021) with separate validations, there is a limitation stemming from our knowledge model and HTT approaches, in that both methods require parameter configurations. We empirically selected parameters in this case to achieve better results but developing an automatic data-driven parameter fine-tuning process is the direction we are heading. From the theoretical perspective, we profiled the knowledge landscape of Covid-19 and identified the knowledge foundation of the *vaccination* topic. Compared to literature-based evidence, it might be more interesting to validate the results with clinical trials and in-depth expert consultations.

Acknowledgements

Mengjia Wu and Yi Zhang are supported by the Australian Research Council under Discovery Early Career Researcher Award DE190100994. Mark Markley, Caitlin Cassidy, and Alan Porter received support through NSF Award # 2029673 – "RAPID: Exploring Causes and Cures for COVID-19 through Improved Access to Biomedical Research," and NSF Award #1759960 – "Indicators of Technological Emergence" to Search Technology, Inc. The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the U.S. National Science Foundation.

Pilot studies of part of this work have been submitted to the 12th Global TechMining Conference 2021 and the workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE 2022) at the 2022 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2022).

Declarations

Conflicts of interests

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Availability of data, material, and code

All data, materials, and code of this study are available on request to mengjia.wu@uts.edu.au

Author contributions

Designed research: Mengjia Wu, Yi Zhang, Mark Markley, Caitlin Cassidy, Nils Newman, Alan Porter; Performed research: Mengjia Wu, Mark Markley, Caitlin Cassidy; Contributed new reagents or analytic tools: Mengjia Wu, Nils Newman, Mark Markley, Caitlin Cassidy, Alan Porter; Analyzed data: Mengjia Wu, Mark Markley, Caitlin Cassidy; Wrote the paper: Mengjia Wu, Yi Zhang, Caitlin Cassidy, Alan Porter.

References

Aalberse, R. C., Akkerdaas, J., & Van Ree, R. (2001). Cross‐reactivity of IgE antibodies to allergens. *Allergy, 56*(6), 478-490.

Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses, 12*(3), 254.

Ahn, D.-G., Shin, H.-J., Kim, M.-H., Lee, S., Kim, H.-S., Myoung, J., et al. (2020). Current Status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). *Journal of Microbiology and Biotechnology, 30*(3), 313-324. doi:10.4014/jmb.2003.03011

Apicella, M., Campopiano, M. C., Mantuano, M., Mazoni, L., Coppelli, A., & Del Prato, S. (2020). COVID-19 in people with diabetes: Understanding the reasons for worse outcomes. *The Lancet Diabetes & Endocrinology, 8*(9), 782-792.

Barron, E., Bakhai, C., Kar, P., Weaver, A., Bradley, D., Ismail, H., et al. (2020). Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England: A whole-population study. *The Lancet Diabetes & Endocrinology, 8*(10), 813-822.

Begelman, G., Keller, P., & Smadja, F. (2006). *Automated tag clustering: Improving search and exploration in the tag space.* Paper presented at the collaborative web tagging workshop at WWW2006, Edinburgh, Scotland.

Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19. *New England Journal of Medicine, 383*(19), 1813-1826.

Bergwerk, M., Gonen, T., Lustig, Y., Amit, S., Lipsitch, M., Cohen, C., et al. (2021). Covid-19 breakthrough infections in vaccinated health care workers. *New England Journal of Medicine, 385*(16), 1474-1484.

Bernal, J. L., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., et al. (2021). Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *New England Journal of Medicine*.

Biswas, N., Mustapha, T., Khubchandani, J., & Price, J. H. (2021). The nature and extent of COVID-19 vaccination hesitancy in healthcare workers. *Journal of Community Health, 46*(6), 1244-1251.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022.

Boddu, S. K., Aurangabadkar, G., & Kuchay, M. S. (2020). New onset diabetes, type 1 diabetes and COVID-19. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14*(6), 2211-2217.

Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences, 117*(27), 15530-15535.

Bornstein, S. R., Rubino, F., Khunti, K., Mingrone, G., Hopkins, D., Birkenfeld, A. L., et al. (2020). Practical recommendations for the management of diabetes in patients with COVID-19. *The Lancet Diabetes & Endocrinology, 8*(6), 546-550.

Brooks, J. T., Beezhold, D. H., Noti, J. D., Coyle, J. P., Derk, R. C., Blachere, F. M., et al. (2021). Maximizing fit for cloth and medical procedure masks to improve performance and reduce SARS-CoV-2 transmission and exposure, 2021. *Morbidity and Mortality Weekly Report, 70*(7), 254.

Brülhart, M., Klotzbücher, V., Lalive, R., & Reich, S. K. (2021). Mental health concerns during the COVID-19 pandemic as revealed by helpline calls. *Nature, 600*(7887), 121-126.

Burki, T. (2020). China's successful control of COVID-19. *The Lancet Infectious Diseases, 20*(11), 1240-1241.

Cabanillas, B., Akdis, C., & Novak, N. (2020). Allergic reactions to the first COVID-19 vaccine: A potential role of Polyethylene glycol. *Allergy, 76*(6), 1617-1618.

Cai, X., Fry, C. V., & Wagner, C. S. (2021). International collaboration during the COVID-19 crisis: Autumn 2020 developments. *Scientometrics, 126*(4), 3683-3692.

Cameroni, E., Bowen, J. E., Rosen, L. E., Saliba, C., Zepeda, S. K., Culap, K., et al. (2022). Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature, 602*(7898), 664-670.

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., & Wengström, E. (2021). Monetary incentives increase COVID-19 vaccinations. *Science, 374*(6569), 879-882.

Cassidy, C. (2020). Parameter tuning Naïve Bayes for automatic patent classification. *World Patent Information, 61*, 101968.

Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A. A., Salhab, H., Fares, M., et al. (2020). A bibliometric analysis of COVID-19 research activity: A call for increased output. *Cureus, 12*(3).

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., et al. (2020). Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: A retrospective review of medical records. *The Lancet, 395*(10226), 809-815.

Chen, Q., Allot, A., & Lu, Z. (2021). LitCovid: An open database of COVID-19 literature. *Nucleic Acids Research, 49*(D1), D1534-D1540.

Chu, I. Y.-H., Alam, P., Larson, H. J., & Lin, L. (2020). Social consequences of mass quarantine during epidemics: A systematic review with implications for the COVID-19 response. *Journal of Travel Medicine, 27*(7), taaa192. doi:https://doi.org/10.1093/jtm/taaa192

Colavizza, G., Costas, R., Traag, V. A., Van Eck, N. J., Van Leeuwen, T., & Waltman, L. (2021). A scientometric overview of CORD-19. *PloS One, 16*(1), e0244839.

Coulson, B. S., Fowler, K., Bishop, R., & Cotton, R. (1985). Neutralizing monoclonal antibodies to human rotavirus and indications of antigenic drift among strains from neonates. *Journal of Virology, 54*(1), 14-20.

Dai, H., Saccardo, S., Han, M. A., Roh, L., Raja, N., Vangala, S., et al. (2021). Behavioural nudges increase COVID-19 vaccinations. *Nature, 597*(7876), 404-409.

Deregt, D., & Babiuk, L. A. (1987). Monoclonal antibodies to bovine coronavirus: Characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology, 161*(2), 410-420.

Ding, J., Fu, H., Liu, Y., Gao, J., Li, Z., Zhao, X., et al. (2020). Prevention and control measures in radiology department for COVID-19. *European Radiology, 30*(7), 3603-3608.

Dror, A. A., Eisenbach, N., Taiber, S., Morozov, N. G., Mizrachi, M., Zigron, A., et al. (2020). Vaccine hesitancy: The next challenge in the fight against COVID-19. *European Journal of Epidemiology, 35*(8), 775-779.

Duch, R., Roope, L. S., Violato, M., Becerra, M. F., Robinson, T. S., Bonnefon, J.-F., et al. (2021). Citizens from 13 countries share similar preferences for COVID-19 vaccine allocation priorities. *Proceedings of the National Academy of Sciences, 118*(38).

Ebadi, A., Xi, P., Tremblay, S., Spencer, B., Pall, R., & Wong, A. (2021). Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing. *Scientometrics, 126*(1), 725-739.

Eibensteiner, P., Spitzauer, S., Steinberger, P., Kraft, D., & Valenta, R. (2000). Immunoglobulin E antibodies of atopic individuals exhibit a broad usage of VH-gene families. *Immunology, 101*(1), 112-119.

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences, 118*(17).

Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *The Lancet Respiratory Medicine, 8*(4), e21.

Felsenstein, S., Herbert, J. A., McNamara, P. S., & Hedrich, C. M. (2020). COVID-19: Immunology and treatment options. *Clinical Immunology, 215*, 108448.

Fry, C. V., Cai, X., Zhang, Y., & Wagner, C. S. (2020). Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research. *PloS One, 15*(7), e0236307.

Garcia, M. A., Homan, P. A., García, C., & Brown, T. H. (2021). The color of COVID-19: Structural racism and the disproportionate impact of the pandemic on older Black and Latinx adults. *The Journals of Gerontology: Series B, 76*(3), e75-e80.

Gautret, P., Lagier, J.-C., Parola, P., Meddeb, L., Mailhe, M., Doudier, B., et al. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International Journal of Antimicrobial Agents, 56*(1), 105949.

Gottlieb, R. L., Nirula, A., Chen, P., Boscia, J., Heller, B., Morris, J., et al. (2021). Effect of bamlanivimab as monotherapy or in combination with etesevimab on viral load in patients with mild to moderate COVID-19: a randomized clinical trial. *JAMA, 325*(7), 632-644.

Haghani, M., & Bliemer, M. C. (2020). Covid-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCoV literature. *Scientometrics, 125*(3), 2695-2726.

Haghani, M., & Varamini, P. (2021). Temporal evolution, most influential studies and sleeping beauties of the coronavirus literature. *Scientometrics, 126*(8), 7005-7050.

Hall, A. K., Nousiainen, M. T., Campisi, P., Dagnone, J. D., Frank, J. R., Kroeker, K. I., et al. (2020). Training disrupted: Practical tips for supporting competency-based medical education during the COVID-19 pandemic. *Medical Teacher, 42*(7), 756-761.

Holman, N., Knighton, P., Kar, P., O'Keefe, J., Curley, M., Weaver, A., et al. (2020). Risk factors for COVID-19-related mortality in people with type 1 and type 2 diabetes in England: A population-based cohort study. *The Lancet Diabetes & Endocrinology, 8*(10), 823-833.

Hossain, M. M. (2020). Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. *SSRN*. doi:10.2139/ssrn.3547824

Hristovski, D., Peterlin, B., Mitchell, J. A., & Humphrey, S. M. (2005). Using literature-based discovery to identify disease candidate genes. *International Journal of Medical Informatics, 74*(2-4), 289-298.

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology, 19*(3), 141-154.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497-506.

Huang, L., Zhang, Y., Guo, Y., Zhu, D., & Porter, A. L. (2014). Four dimensional Science and Technology planning: A new approach based on bibliometrics and technology roadmapping. *Technological Forecasting and Social Change, 81*, 39-48.

Imagawa, A., Hanafusa, T., Miyagawa, J.-i., & Matsuzawa, Y. (2000). A novel subtype of type 1 diabetes mellitus characterized by a rapid onset and an absence of diabetes-related antibodies. *New England Journal of Medicine, 342*(5), 301-307.

Islam, M. S., Rahman, K. M., Sun, Y., Qureshi, M. O., Abdi, I., Chughtai, A. A., et al. (2020a). Current knowledge of COVID-19 and infection prevention and control strategies in healthcare settings: A global analysis. *Infection Control & Hospital Epidemiology, 41*(10), 1196-1206.

Islam, M. S., Sarkar, T., Khan, S. H., Kamal, A.-H. M., Hasan, S. M., Kabir, A., et al. (2020b). COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene, 103*(4), 1621.

Jenmalm, M., Van Snick, J., Cormont, F., & Salman, B. (2001). Allergen‐induced Th1 and Th2 cytokine secretion in relation to specific allergen sensitization and atopic symptoms in children. *Clinical & Experimental Allergy, 31*(10), 1528-1535.

Jin, J.-M., Bai, P., He, W., Wu, F., Liu, X.-F., Han, D.-M., et al. (2020). Gender differences in patients with COVID-19: Focus on severity and mortality. *Frontiers in Public Health*, 152.

Jose, R. J., & Manuel, A. (2020). COVID-19 cytokine storm: The interplay between inflammation and coagulation. *The Lancet Respiratory Medicine, 8*(6), e46-e47.

Kajikawa, Y., Mejia, C., Wu, M., & Zhang, Y. (2022). Academic landscape of Technological Forecasting and Social Change through citation network and topic analyses. *Technological Forecasting and Social Change, 182*, 121877.

Kounis, N. G., Koniari, I., de Gregorio, C., Velissaris, D., Petalas, K., Brinia, A., et al. (2021). Allergic reactions to current available COVID-19 vaccinations: Pathophysiology, causality, and therapeutic considerations. *Vaccines, 9*(3), 221.

Kousha, K., & Thelwall, M. (2020). COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quantitative Science Studies, 1*(3), 1068-1091.

Koyama, T., Weeraratne, D., Snowdon, J. L., & Parida, L. (2020). Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens, 9*(5), 324.

Lai, J., Ma, S., Wang, Y., Cai, Z., Hu, J., Wei, N., et al. (2020). Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Network Open, 3*(3), e203976-e203976.

Levi, M., Thachil, J., Iba, T., & Levy, J. H. (2020). Coagulation abnormalities and thrombosis in patients with COVID-19. *The Lancet Haematology, 7*(6), e438-e440.

Li, H. O.-Y., Bailey, A., Huynh, D., & Chan, J. (2020a). YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Global Health, 5*(5), e002604.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020b). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

Lim, S., Bae, J. H., Kwon, H.-S., & Nauck, M. A. (2021). COVID-19 and diabetes mellitus: From pathophysiology to clinical management. *Nature Reviews Endocrinology, 17*(1), 11-30.

Liu, K., Chen, Y., Lin, R., & Han, K. (2020a). Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection, 80*(6), e14-e18.

Liu, Q., Luo, D., Haase, J. E., Guo, Q., Wang, X. Q., Liu, S., et al. (2020b). The experiences of health-care providers during the COVID-19 crisis in China: A qualitative study. *The Lancet Global Health, 8*(6), e790-e798.

Long, Y., Hu, T., Liu, L., Chen, R., Guo, Q., Yang, L., et al. (2020). Effectiveness of N95 respirators versus surgical masks against influenza: A systematic review and meta-analysis. *Journal of Evidence-Based Medicine, 13*(2), 93-101.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet, 395*(10224), 565-574.

Machingaidze, S., & Wiysonge, C. S. (2021). Understanding COVID-19 vaccine hesitancy. *Nature Medicine, 27*(8), 1338-1339.

Marchand, L., Pecquet, M., & Luyton, C. (2020). Type 1 diabetes onset triggered by COVID-19. *Acta Diabetologica, 57*(10), 1265-1266.

Mejia, C., Wu, M., Zhang, Y., & Kajikawa, Y. (2021). Exploring topics in bibliometric research through citation networks and semantic analysis. *Frontiers in research metrics and analytics, 6*.

Merrill, J. T., Erkan, D., Winakur, J., & James, J. A. (2020). Emerging evidence of a COVID-19 thrombotic syndrome has treatment implications. *Nature Reviews Rheumatology, 16*(10), 581-589.

Mick, P., & Murphy, R. (2020). Aerosol-generating otolaryngology procedures and the need for enhanced PPE during the COVID-19 pandemic: A literature review. *Journal of Otolaryngology-Head & Neck Surgery, 49*(1), 1-10.

Mockett, A. A., Cavanagh, D., & Brown, T. D. K. (1984). Monoclonal antibodies to the S1 spike and membrane proteins of avian infectious bronchitis coronavirus strain Massachusetts M41. *Journal of General Virology, 65*(12), 2281-2286.

Muniyappa, R., & Gubbi, S. (2020). COVID-19 pandemic, coronaviruses, and diabetes mellitus. *American Journal of Physiology-Endocrinology and Metabolism*. doi:10.1152/ajpendo.00124.2020

Nasab, F.-R., & Rahim, F. (2020). Bibliometric analysis of global scientific research on SARS-CoV-2 (COVID-19). doi:10.1101/2020.03.19.20038752

Nussbaumer-Streit, B., Mayr, V., Dobrescu, A. I., Chapman, A., Persad, E., Klerings, I., et al. (2020). Quarantine alone or in combination with other public health measures to control COVID-19: A rapid review. *Cochrane Database of Systematic Reviews*(9).

Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2021). Genetic mechanisms of critical illness in Covid-19. *Nature, 591*(7848), 92-98.

Pal, R., Bhadada, S. K., & Misra, A. (2021). COVID-19 vaccination in patients with diabetes mellitus: Current concepts, uncertainties and challenges. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 15*(2), 505-508.

Pang, J., Wang, M. X., Ang, I. Y. H., Tan, S. H. X., Lewis, R. F., Chen, J. I.-P., et al. (2020). Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): A systematic review. *Journal of Clinical Medicine, 9*(3), 623.

Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: Are they closely related? *Clinical Microbiology and Infection, 26*(6), 729-734.

Pica, N., Hai, R., Krammer, F., Wang, T. T., Maamary, J., Eggink, D., et al. (2012). Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proceedings of the National Academy of Sciences, 109*(7), 2573-2578.

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*. doi:10.1056/NEJMoa2034577

Porter, A. L., Zhang, Y., Huang, Y., & Wu, M. (2020). Tracking and Mining the COVID-19 Research Literature. *Frontiers in Research Metrics and Analytics, 5*, 12.

Pourhatami, A., Kaviyani-Charati, M., Kargar, B., Baziyad, H., Kargar, M., & Olmeda-Gómez, C. (2021). Mapping the intellectual structure of the coronavirus field (2000–2020): A co-word analysis. *Scientometrics, 126*(8), 6625-6657.

Powers, A. C., Aronoff, D. M., & Eckel, R. H. (2021). COVID-19 vaccine prioritisation for type 1 and type 2 diabetes. *The Lancet Diabetes & Endocrinology, 9*(3), 140-141.

Prompetchara, E., Ketloy, C., & Palaga, T. (2020). Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pacific Journal of Allergy and Immunology, 38*(1), 1-9.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*. doi:10.48550/arXiv.1910.10683

Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *The Lancet, 395*(10231), 1225-1228.

Ruktanonchai, N. W., Floyd, J., Lai, S., Ruktanonchai, C. W., Sadilek, A., Rente-Lourenco, P., et al. (2020). Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science, 369*(6510), 1465-1470.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management, 24*(5), 513-523.

Salton, G., & McGill, M. J. (1986). *Introduction to Modern Information Retrieval*: McGraw-Hill, Inc.

Shen, Z., Ma, H., & Wang, K. (2018). A web-scale system for scientific knowledge exploration. *arXiv preprint arXiv:1805.12216*.

Shepherd, J. P., Moore, S. C., Long, A., Kollar, L. M. M., & Sumner, S. A. (2021). Association between COVID-19 lockdown measures and emergency department visits for violence-related injuries in Cardiff, Wales. *JAMA, 325*(9), 885-887.

Shi, L., Lu, Z.-A., Que, J.-Y., Huang, X.-L., Liu, L., Ran, M.-S., et al. (2020). Prevalence of and risk factors associated with mental health symptoms among the general population in China during the coronavirus disease 2019 pandemic. *JAMA Network Open, 3*(7), e2014053-e2014053.

Sikkema, R. S., Pas, S. D., Nieuwenhuijse, D. F., O'Toole, Á., Verweij, J., van der Linden, A., et al. (2020). COVID-19 in health-care workers in three hospitals in the south of the Netherlands: A cross-sectional study. *The Lancet Infectious Diseases, 20*(11), 1273-1280.

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy, 43*(8), 1450-1467.

Spinelli, A., & Pellino, G. (2020). COVID-19 pandemic: Perspectives on an unfolding crisis. *Journal of British Surgery, 107*(7), 785-787.

Starr, T. N., Greaney, A. J., Addetia, A., Hannon, W. W., Choudhary, M. C., Dingens, A. S., et al. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science, 371*(6531), 850-854.

Stenstrom, G., Gottsater, A., Bakhtadze, E., Berger, B., & Sundkvist, G. (2005). Latent autoimmune diabetes in adults: Definition, prevalence, β-cell function, and treatment. *Diabetes, 54*(suppl_2), S68-S72.

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine, 30*(1), 7-18.

Taylor, P. C., Adams, A. C., Hufford, M. M., De La Torre, I., Winthrop, K., & Gottlieb, R. L. (2021). Neutralizing monoclonal antibodies for treatment of COVID-19. *Nature Reviews Immunology, 21*(6), 382-393.

Thomas, S. J., Moreira Jr, E. D., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., et al. (2021). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months. *New England Journal of Medicine, 385*(19), 1761-1773. doi:10.1056/NEJMoa2110345

Traggiai, E., Becker, S., Subbarao, K., Kolesnikova, L., Uematsu, Y., Gismondo, M. R., et al. (2004). An efficient method to make human monoclonal antibodies from memory B cells: Potent neutralization of SARS coronavirus. *Nature Medicine, 10*(8), 871-875.

Tran, B. X., Ha, G. H., Nguyen, L. H., Vu, G. T., Hoang, M. T., Le, H. T., et al. (2020). Studies of novel coronavirus disease 19 (COVID-19) pandemic: A global analysis of literature. *International Journal of Environmental Research and Public Health, 17*(11), 4095.

Trewartha, A., Dagdelen, J., Huo, H., Cruse, K., Wang, Z., He, T., et al. (2020). COVIDScholar: An automated COVID-19 research aggregation and analysis platform. *arXiv preprint arXiv:2012.03891*. doi:10.48550/arXiv.2012.03891

Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. A. (2021). What social media told us in the time of COVID-19: A scoping review. *The Lancet Digital Health, 3*(3), e175-e194.

Van Belle, T. L., Coppieters, K. T., & Von Herrath, M. G. (2011). Type 1 diabetes: Etiology, immunology, and therapeutic strategies. *Physiological Reviews, 91*(1), 79-118.

Venter, Z. S., Aunan, K., Chowdhury, S., & Lelieveld, J. (2020). COVID-19 lockdowns cause global air pollution declines. *Proceedings of the National Academy of Sciences, 117*(32), 18984-18990.

Wagner, C. S., Cai, X., Zhang, Y., & Fry, C. V. (2022). One-year in: COVID-19 research at the international level in CORD-19 data. *Plos one, 17*(5), e0261624.

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., McIntyre, R. S., et al. (2020a). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain, Behavior, and Immunity, 87*, 40-48.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020b). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA, 323*(11), 1061-1069.

Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., et al. (2020c). Cord-19: The covid-19 open research dataset. *ArXiv*. doi:10.48550/arXiv.2004.10706

Wang, Y., Wang, Y., Chen, Y., & Qin, Q. (2020d). Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID‐19) implicate special control measures. *Journal of Medical Virology, 92*(6), 568-576.

Wartena, C., & Brussee, R. (2008). *Topic detection by clustering keywords.* Paper presented at the 2008 19th International Workshop on Database and Expert Systems Applications.

Watts, R. J., & Porter, A. L. (1999). *Mining foreign language information resources.* Paper presented at the PICMET'99: Portland International Conference on Management of Engineering and Technology. Proceedings Vol-1: Book of Summaries (IEEE Cat. No. 99CH36310).

Watts, R. J., Porter, A. L., & Courseault, C. (1999). Functional analysis: Deriving systems knowledge from bibliographic information resources. *Information Knowledge Systems Management, 1*(1), 45-61.

Wenzlau, J. M., Juhl, K., Yu, L., Moua, O., Sarkar, S. A., Gottlieb, P., et al. (2007). The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proceedings of the National Academy of Sciences, 104*(43), 17040-17045.

Williams, S. N., Armitage, C. J., Tampe, T., & Dienes, K. (2020). Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: A UK-based focus group study. *BMJ Open, 10*(7), e039334.

Wise, C., Ioannidis, V. N., Calvo, M. R., Song, X., Price, G., Kulkarni, N., et al. (2020). COVID-19 knowledge graph: Accelerating information retrieval and discovery for scientific literature. *arXiv preprint arXiv:2007.12731*. doi:10.48550/arXiv.2007.12731

Wu, H.-l., Huang, J., Zhang, C. J., He, Z., & Ming, W.-K. (2020). Facemask shortage and the novel coronavirus disease (COVID-19) outbreak: Reflections on public health measures. *EClinicalMedicine, 21*, 100329.

Wu, M., Kozanoglu, D. C., Min, C., & Zhang, Y. (2021a). Unraveling the capabilities that enable digital transformation: A data-driven methodology and the case of artificial intelligence. *Advanced Engineering Informatics, 50*, 101368.

Wu, M., & Zhang, Y. (2021). *Hierarchical topic tree: A hybrid model comprising network analysis and density peak search*. Paper presented at the 18th International Conference on Scientometrics and Informetrics, Belgium.

Wu, M., Zhang, Y., Grosser, M., Tipper, S., Venter, D., Lin, H., et al. (2021b). Profiling COVID-19 genetic research: A data-driven study utilizing intelligent bibliometrics. *Frontiers in Research Metrics and Analytics, 6*, 30.

Wu, M., Zhang, Y., Zhang, G., & Lu, J. (2021c). Exploring the genetic basis of diseases through a heterogeneous bibliometric network: A methodology and case study. *Technological Forecasting and Social Change, 164*, 120513.

Xia, S., Duan, K., Zhang, Y., Zhao, D., Zhang, H., Xie, Z., et al. (2020). Effect of an inactivated vaccine against SARS-CoV-2 on safety and immunogenicity outcomes: Interim analysis of 2 randomized clinical trials. *JAMA, 324*(10), 951-960.

Xie, B., He, D., Mercer, T., Wang, Y., Wu, D., Fleischmann, K. R., et al. (2020). Global health crises are also information crises: A call to action. *Journal of the Association for Information Science and Technology, 71*(12), 1419-1423.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics, 100*(3), 767-786.

Yu, Q., Wang, Q., Zhang, Y., Chen, C., Ryu, H., Park, N., et al. (2021). Analyzing knowledge entities about COVID-19 using entitymetrics. *Scientometrics, 126*(5), 4491-4509.

Yu, X., Tsibane, T., McGraw, P. A., House, F. S., Keefer, C. J., Hicar, M. D., et al. (2008). Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature, 455*(7212), 532-536.

Yuan, M., Huang, D., Lee, C.-C. D., Wu, N. C., Jackson, A. M., Zhu, X., et al. (2021). Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science, 373*(6556), 818-823.

Zhang, E., Gupta, N., Nogueira, R., Cho, K., & Lin, J. (2020a). Rapidly deploying a neural search engine for the covid-19 open research dataset: Preliminary thoughts and lessons learned. *arXiv preprint arXiv:2004.05125*. doi:10.48550/arXiv.2004.05125

Zhang, Y., Cai, X., Fry, C. V., Wu, M., & Wagner, C. S. (2021a). Topic evolution, disruption and resilience in early COVID-19 research. *Scientometrics, 126*(5), 4225-4253.

Zhang, Y., Porter, A. L., Cunningham, S., Chiavetta, D., & Newman, N. (2020b). Parallel or Intersecting Lines? Intelligent Bibliometrics for Investigating the Involvement of Data Science in Policy Analysis. *IEEE Transactions on Engineering Management*, 1-13. doi:10.1109/TEM.2020.2974761

Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014). "Term clumping" for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change, 85*, 26-39.

Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021b). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems, 222*, 106994.

Zhang, Y., Zhang, G., Zhu, D., & Lu, J. (2017). Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics. *Journal of the Association for Information Science and Technology, 68*(8), 1925-1939.

Zhu, Z., Chakraborti, S., He, Y., Roberts, A., Sheahan, T., Xiao, X., et al. (2007). Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. *Proceedings of the National Academy of Sciences, 104*(29), 12123-12128.