



Analyzing knowledge entities about COVID-19 using entitymetrics

Qi Yu^{1,2} · Qi Wang^{1,3} · Yafei Zhang¹ · Chongyan Chen⁴ · Hyeyoung Ryu⁵ · Namu Park⁶ · Jae-Eun Baek⁷ · Keyuan Li⁸ · Yifei Wu⁹ · Daifeng Li¹⁰ · Jian Xu¹⁰ · Meijun Liu^{11,12} · Jeremy J. Yang¹⁴ · Chenwei Zhang¹³ · Chao Lu¹⁵ · Peng Zhang⁹ · Xin Li¹⁶ · Baitong Chen¹⁷ · Islam Akef Ebeid⁴ · Julia Fensel¹⁸ · Chao Min¹⁹ · Yujia Zhai^{16,20} · Min Song²¹ · Ying Ding^{4,22} · Yi Bu²³ 

Received: 24 November 2020 / Accepted: 26 February 2021 / Published online: 12 March 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract

COVID-19 cases have surpassed the 109+ million markers, with deaths tallying up to 2.4 million. Tens of thousands of papers regarding COVID-19 have been published along with countless bibliometric analyses done on COVID-19 literature. Despite this, none of the analyses have focused on domain entities occurring in scientific publications. However, analysis of these bio-entities and the relations among them, a strategy called entity metrics, could offer more insights into knowledge usage and diffusion in specific cases. Thus, this paper presents an entitymetric analysis on COVID-19 literature. We construct an entity–entity co-occurrence network and employ network indicators to analyze the extracted entities. We find that ACE-2 and C-reactive protein are two very important genes and that lopinavir and ritonavir are two very important chemicals, regardless of the results from either ranking.

Keywords COVID-19 · Knowledge graph · Entity · Entitymetrics · Scientific publications · Bibliometrics

Introduction

As of February 19, 2021, the number of confirmed COVID-19 cases has increased to 109.6 million globally; more than 2.42 million people have died across the globe (World Health Organization, 2020). We are trapped in this weird situation about COVID-19, given it is so new and we have to learn it via experiencing it. Science plays a major role during the COVID-19. Scientists have been grasping at straws in an attempt to find effective treatments against the virus. Their collective efforts have generated thousands of academic

✉ Min Song
min.song@yonsei.ac.kr

✉ Yi Bu
buyi@pku.edu.cn

Extended author information available on the last page of the article

papers on peer-reviewed venues and pre-print platforms. Existing studies have implemented some preliminary bibliometric analyses on COVID-19-related literature, such as publication-country distribution (Chahrour et al., 2020), citation analysis (Hossain, 2020), and keyword co-occurrence networks (Rafiei Nasab & Rahim, 2020). These bibliometric analyses may quantitatively monitor research performance in this domain and reveal some potential implications on knowledge diffusion of COVID-19 literature.

Any scientific development or breakthrough can easily trigger the public and political attention. Thus, understanding scientific activities beyond simple ranking of authors or institutions is highly demanded by the public and policy-maker. Meanwhile, domain experts have also expressed the need to know which drugs or genes are heavily discussed in the fast-growing COVID-19 literature, and no scientist can read them all in a timely manner. They also expressed the need to know which team is working on which gene or drug so that they can seek collaboration. Therefore, it is quite important to measure the heavily discussed bio-entities such as drugs (e.g., Remdesivir), diseases (e.g., SARS), and proteins (e.g., ACE-2) at the entity level, as well as most “collaborative” bio-entities (more scientists are working on these bio-entities). Discovery of such entities can contribute uniquely to the society, and analysis of these bio-entities, as well as the relations among them, could offer more insights into knowledge usage and diffusion in specific cases. These knowledge-level analyses can reveal intricate inquiries, far beyond those that paper-level analysis can offer. With the advancement of established controlled vocabularies and algorithms in biomedical science, it is feasible to extract these bio-entities from scientific publications for in-depth analyses. This shows a typical way of implementing entitymetrics, a strategy of using entities “in the measurement of impact, knowledge usage, and knowledge transfer to facilitate knowledge discovery” (Ding et al., 2013, p. 2). Entitymetrics applies bibliometric approaches to knowledge entities and aims to contribute to knowledge discovery.

Therefore, this paper implements entitymetrics on scientific publications in the domain of coronavirus, and studies the patterns of coronavirus-related bio-entities, as well as their potential relations from literature. We pay special attention to entities’ scientific impact, such as which entities are most frequently studied by researchers; to this end, we examine several established indices, including popularity, promising, and collaboration indices (Li et al., 2020).

Literature review

Bibliometric analysis on COVID-19 literature

Starting from the outbreak of this pandemic, bibliometricians have started to analyze COVID-19 literature. Chahrour et al. (2020) utilized the PubMed database and the World Health Organization (WHO) database related to COVID-19 and selected 564 publications between December 2019 and March 18, 2020. They conducted a country-level analysis and reported that China produced the greater proportion of publications among all countries and that Singapore and Mauritius ranked the first in terms of the number of publications per million persons along with the number of publications per billion-dollar GDP. Using Scopus, Nasab and Rahim (2020) adopted VOSViewer as a tool to visualize bibliometric entities related to COVID-19 literature. They also analyzed financial supporters of these publications. Another bibliometric analysis based upon the Web of Science dataset focused more on citations, concluding that the number of citations per document is 2.47 (Hossain,

2020). Kousha and Thelwall (2020) examined database coverage, citations, and social media related to COVID-19. They found “a high degree of convergence between articles shared in the social web and citation counts, at least in the short term” (p.1). Fry et al. (2020) investigated an interesting question on whether COVID-19 accelerates or reverses international collaboration, observing that the U.S. and China continue to be at the center of the global co-authorship network as before. They also found that COVID-19 narrows scientists’ collaborations and favors elite structures. Yet, this study did not establish any baseline model and their conclusions need further additional empirical studies. Similar bibliometric analyses include those of Dehghanbanadaki et al. (2020), Farooq et al. (2021), Hamidah et al. (2020), and Zhai et al. (2020), etc.

Another branch of bibliometric analysis paid special attention to gender difference in scientists. By exploring working-paper publication in economics, Amano-Patiño et al. (2020) reported that female economists publish fewer working papers than male colleagues and argued that “the effects of lockdowns on the division of labor at home have been particularly detrimental to the research activity of women” (p. 1). Using questionnaires, Myers et al. (2020) concluded that scientists who have young children have decreased time for spending on research. Similarly, Vincent-Lamarre et al. (2020) used a large-scale bibliographic dataset and observed women submitted fewer papers to pre-print platforms such as *medRxiv*, *bioRxiv*, and *arXiv*.

Systematic review and meta-analysis on COVID-19 literature

Starting from the outbreak of COVID-19, scientists have started to implement systematic review and/or meta-analysis on this topic, just from different focuses. Yang et al. (2020) conducted a meta-analysis based on seven related studies with 1500+ patients included and summarized some prevalent clinical symptoms such as fever, cough, and fatigue with their statistical characteristics (e.g., confidence interval). Rodriguez-Morales et al. (2020) selected 19 from 660 articles in Medline/PubMed, Scopus, along with Web of Science, and analyzed the clinical, laboratory, and imaging features of COVID-19 patients. According to both quantitative and qualitative analyses, they highlighted that this type of coronavirus brings “a huge burden to healthcare facilities, especially in patients with comorbidities” (p. 2). Vardavas and Nikitara (2020) sought potential evidence from literature on the relations between COVID-19 and smoking by using a systematic review on PubMed and Science-Direct databases. Only five studies have been adopted in their systematic review, and many details were analyzed, such as setting, the number of patients, study design and time horizon, outcomes, smoking characteristics, etc. They observed that “smoking is most likely associated with the negative progression and adverse outcomes of COVID-19” (p. 22).

Besides these, systematic review and/or meta-analysis has been implemented in many COVID-19-related literatures, though with various objectives, such as the efficacy and safety of chloroquine as a treatment for COVID-19 (Cortegiani et al., 2020, p. 19), imaging profile of COVID-19 patients (Ng et al., 2020; Salehi et al., 2020), children and adults symptom comparisons (Ludvigsson, 2020), and school closure and management practices (Viner et al., 2020).

Entitymetrics

The idea of utilizing entities to understand the impact of knowledge has been proposed and executed by numerous previous studies. Bell et al. (2011) established a bio-entity network

containing protein–protein interactions, protein/gene regulations, protein–small molecule interactions, protein–gene ontology relationships, protein–pathway relationships, and pathway–disease relationships. Their built network was found to show plausible hypotheses for biological experiments. Ren et al. (2011) particularly focused on protein–protein networks, with which the “hub-proteins” (p. 734) could be easily identified in the system. Gysi et al. (2020) analyzed the molecular perturbations induced by the SARS-CoV-2 virus in order to identify potential drug repurposing candidates. Specifically, they adopted several network-level metrics, such as network proximity and diffusion state distance, and designed a graph neural network for COVID-19 treatment recommendations. Through these analyses, they identified 11 previously proposed drug-repurposing candidates for COVID-19 and additional 21 drugs that are under clinical trials when the authors wrote the paper.

As a novel way of characterizing the impact of knowledge units, entitymetrics has been applied to highlight the importance of entities within scientific literature (Ding et al., 2013). Entitymetrics is further used for knowledge discovery, such as drug repurposing quantifications (Li et al., 2020), comparison with other bibliometric networks (Lee et al., 2017), ego-centered bio-entity analyses (Song, 2016) and author profile analyses (Park et al., 2017), as well as implicit entity relation identifications (Song et al., 2013).

Data and methods

In March 2020, the Allen Institute of AI, together with other leading research groups, released a COVID-19 Open Research dataset, covering COVID-19-related scientific publication bibliographic metadata (COVID-19 Open Research Dataset (CORD-19), 2020). This dataset enabled us to establish our entitymetrics on COVID-19 literature. Based on the version released at the beginning of May 2020, there are about 57 thousand publications in 1951–2020 that come from different sources, including: (1) PubMed’s PMC open access corpus, (2) COVID-19 research articles from a corpus maintained by the WHO, and (3) *bioRxiv* and *medRxiv* pre-prints. For (1) and (3), the below query term is used:

COVID-19” OR Coronavirus OR “Corona virus” OR “2019-nCoV” OR “SARS-CoV” OR “MERS-CoV” OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome

For these publications, we extract bio-entities from titles and abstracts using PubTator, a web-based text mining tool for pre-annotating bio-entities (Wei et al., 2013). With this toolkit, we obtain the recognized bio-entities as well as their types (e.g., gene, chemical, species, mutation, etc.) in each scientific publication. This yields 39,914 distinct bio-entities with 10,265 chemicals and 7444 genes. Moreover, we follow Xu et al. (2020) to disambiguate authors’ names in this dataset; this enables us to implement future author-level analyses on COVID-19. In the following analyses, we are particularly interested in publications, along with their entities, between January and April 2020, as this is the early period of COVID-19 outbreaking.

To characterize top bio-entities, we use the following indicators: (1) popularity index (Li et al., 2020), (2) collaboration index (Li et al., 2020), (3) network indicators to describe the features of this entity co-occurrence network (i.e., PageRank, different types of centrality, and Average Degree. First of all, we want to address the needs of domain scientists and the public wanting to know more fine-grained level of scientific knowledge and activities related to COVID-19. Popularity index is the one to tell the

readers what the most discussed bio entities in this huge set of COVID-19 literature are. The collaboration index is the one to tell what the bio-entities relate to COVID-19 with more researchers working on are. Both of these can provide information which traditional bibliometric analysis could not, therefore adding value to the scientific communities. There are two more indicators of entitymetrics that we did not use, which are promising index and prestige index (Li et al., 2020). The promising index is about the bio-entities with the fast growth rate of publications discussing them. Because COVID-19 is a new domain, and most of the papers are published in 2020, and many of them are preprints from bioRxiv or medRxiv which do not have the publication year. So it is unlikely for us to be able to calculate the growth rate unless we narrow the publication year to publication month which requires manual effort to collect such data. Prestige index talks about the bio-entities mentioned in the journals with high impact journals. Roughly the same reason, given around 50% of the COVID-19 literature are preprints, we could not obtain the journal impact factors, therefore we did not use these two indicators. For the network indicators, we just adopt some commonly used indicators, including PageRank, different types of centralities, and average degree.

Details of our adopted indicators are listed as follows:

- The *popularity index* (PI) of a certain bio-entity is defined as the number of publications that mention this entity during a certain period divided by the number of publications in the research field in the same period. Thus, we can see that the popularity index of a certain bio-entity reflects the proportions of scientific publications that discuss this entity.
- As for the *collaboration index* (CI) of a certain bio-entity, it is stipulated as the number of distinct authors whose publications mention the entity divided by the total number of distinct authors in the same period of time.
- *PageRank* The PageRank algorithm is a method for node ranking in a network and also for evaluating the node's importance (Brin & Page, 1998). In PageRank, network nodes that are linked to many other nodes indicate that they are quite important, and their PageRank values will be thus correspondingly great. Meanwhile, if a node with a high PageRank is linked to another node, the PageRank of that node will increase accordingly.
- *Closeness centrality, betweenness, and eigenvalue centrality* Centrality is a commonly used network concept that is used to indicate a "central" node. According to the way of defining "center", there are many types of centralities, such as degree centrality, closeness centrality, harmonic closeness centrality, and betweenness. Closeness centrality is to calculate the total distance of a node to and all other nodes, and the smaller the total value is, the shorter the path of this node to and all other nodes, and the closer the node is to all other nodes. A node with a high degree of closeness to the center indicates that the node is closest to any other node and is spatially reflected in the center position. Betweenness calculates the number of shortest paths through a node. The greater the number of shortest paths through a node, the greater its value of centrality. Another type of centrality is the eigenvalue centrality. The basic idea is that points connected with points with high centrality are regarded as more important. Eigenvalue Centrality is calculated by using the adjacency matrix of a graph.
- *Average degree* This indicator calculates the degree of each node and counts the number of nodes with the same degree. From this, we can observe which nodes have the highest value of degree and examine whether there are more points connecting them.

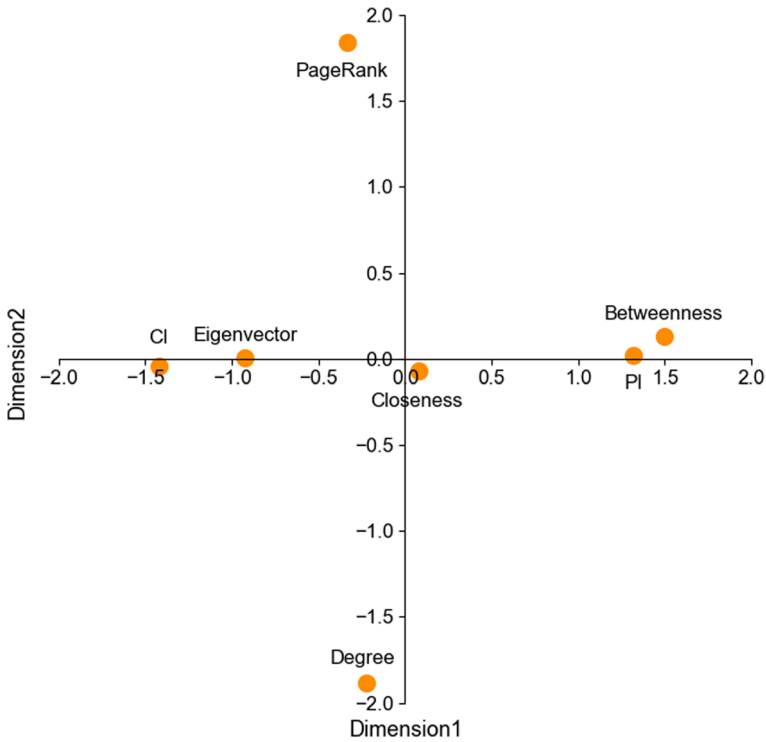


Fig. 1 Two-dimensional map of seven indicators by use multidimensional scaling algorithm on pairwise cosine similarity between indicators

Results

Overview

As mentioned, each entity has a rank value for a specific indicator. Therefore, we are eager to know the relationship among these different ranks. To this end, we employ the multidimensional scaling analysis (MDS), a statistical method that has been widely used in social sciences, to classify samples or variables with many dimensions according to their similarity (near) or dis-similarity (far, i.e. by calculating their distance). This is essentially a dimension reduction strategy. In an MDS map, space and distance are used to reflect the relationship between each point and judge the distribution of each point in the network and the density of the network, etc. That is, you can find out what groups are distributed throughout the network.¹ As shown in Fig. 1, two main groups were formed based on their similarities except PageRank and Average Degree: Group 1 includes CI and Eigenvector centrality; Groups 2 includes PI and Betweenness. In order to illustrate the sequence

¹ We have obtained a table where each row represents a bio-entity and each column represents an indicator. We then calculated the correlation matrix (i.e., a similarity matrix) among indicators based on the table. This matrix is used as the input of MDS.

Table 1 Top-ranked genes in terms of Betweenness in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	AST	ACE-2	ACE-2	AKt2
2	LCN-2	Spike	Aspartate aminotransferase	LPS
3	Granzyme B	CD4	CRP	IL-10
4	ALB	CD8	CD4	CTX
5	cTnI	mTOR	AST	IL-17
6	ACE-2	TNF	LPS	IL-8
7	Gremlin-1	LPS	TNF	IL-1beta
8	PARPs	IL-1beta	Albumin	IL-12
9	CD4	cTnI	cTnI	IGFBP-2
10	TNF	IL-10	IL-10	IGFBP-3

analysis results of seven measurements for the data of January to April 2020, we select the representative indicators in each group, namely betweenness and the collaboration index (PI) from the right of Fig. 1, and the popularity index (CI) from the left of Fig. 1 for further analysis.

Betweenness

Table 1 presents top-ranked genes when we calculate betweenness index. We can see that ACE-2 (Angiotensin-converting enzyme 2) ranks rises in the first three periods, ACE-2 lowers blood pressure by catalyzing the hydrolysis of angiotensin II, which is a vasoconstrictor peptide, into angiotensin, a vasodilator. Since the coronavirus SARS-CoV-2 was found to be particularly deleterious to patients with cardiovascular disease per existing studies (e.g., South et al., 2020), the mechanism for SARS-CoV-2 infection is the requisite binding of the virus to the membrane-bound form of ACE-2 and internalization of the complex by the host cell.

Some recent studies show potential relation between ACE-2 and ibuprofen, a commonly used drug to treat fever and mild to severe pain. A letter published in *The Lancet Respiratory Medicine*, for instance, stated that increased expression of ACE-2 could facilitate an infection with COVID-19. The letter states that thiazolidinediones and ibuprofen increase ACE-2; however, this appears to be based on animal studies (Fang et al., 2020). A statement attributed to WHO spokesperson Lindmeier recommending paracetamol and avoiding ibuprofen as a self-medication was widely circulated in the media; however, such a position could not be found on the WHO website or other official sources. As of March 18, 2020, WHO has not recommended against the use of ibuprofen per its Twitter.² The COVID-19 evidence table updated daily³ also shows that “there have been unsubstantiated reports of younger, healthy patients who took ibuprofen and suffered severe outcomes with COVID-19.” (p. 40). Yet, there lacks official and scientific case reports. Although no compelling evidence shows associations between ibuprofen and negative outcomes in patients

² <https://twitter.com/WHO/status/1240409217997189128>

³ <https://www.ashp.org/-/media/assets/pharmacy-practice/resource-centers/Coronavirus/docs/ASHP-COVID-19-Evidence-Table.ashx?la=en&hash=B414CC64FD64E1AE8CA47AD753BA744EDF4FFB8C>.

Table 2 Top-ranked chemicals in terms of betweenness in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	Lactate	Cysteine	Pangolins	Serine
2	Ritonavir	PhJ	Bilirubin	NO ₂
3	Ribavirin	Nitrogen	Lactate	Urea
4	Alanine	Corticosteroids	Aspartate	ADP
5	Iopinavir	Lactate	Lopinavir	Nitrogen
6	Ser	Pangolins	Nitrogen	Lactate
7	Abidol	ADP	Corticosteroids	CO ₂
8	Sofosbuvir	Aspartate	Dexamethasone	Ser
9	PPE	Ritonavir	Hydrogen	PPE
10	5-OH	Ebselen	Hydroxychloroquine	–

with COVID-19, some experts have recommended preferentially using acetaminophen for treatment of fever (e.g., Alhazzani et al., 2020).

We can observe from Table 1 that CD4 and cTnI occurs in the first three periods. In molecular biology, CD4 is a glycoprotein found on the surface of immune cells such as T helper cells, monocytes, macrophages, and dendritic cells, etc. Existing literature showed that SARS-CoV-2 mainly attacks human CD4 T lymphocytes (Giamarellos-Bourboulis et al., 2020), and that the HIV virus of AIDS also attacks CD4 T cells (Li et al., 2019). This hints, though theoretically, potential drug repurposing motivations of using HIV drugs on COVID-19. The high rank of CD4 in “2020/1 to 2020–3” period demonstrates its popularity in February and March 2020.

We also observe cTnI in Table 1. Basically Troponin (cTn) is divided into T and I subtypes. In the early stage of myocardial injury, cTnT and cTnI free in the cytoplasm will be rapidly released into the blood. After 4–6 h, the increased troponin can be maintained in the blood for 5–10 days. Troponin is currently the serum “gold standard” for the diagnosis of AMI, especially for patients with minor heart damage who cannot be diagnosed by ECG changes and without typical clinical symptoms, it is the best auxiliary diagnostic indicator.

We can also observe from Table 1 that IL-10 occurs in the first three periods; particularly, in April, it ranks third. IL-10, an abbreviation standing for Interleukin10, is an interleukin that acts as both a pro-inflammatory cytokine and an anti-inflammatory myokine. IL-10 inhibitors may ameliorate severe damage to lung tissue caused by cytokine release in patients with serious COVID-19 infections. From the table, one can see that other interleukin genes, e.g., IL-8 and IL-12, also occur in the lists.

Table 2 presents most popular chemicals in 2020 COVID-19 literature when we calculate betweenness index, where we observe that Lopinavir and Ritonavir keep ranking high in January, February, and March. Lopinavir is an HIV-related drug. Although it does not cure HIV or AIDS directly, its combinations with other drugs slow down the disease progress and prolong life. Because currently there are no available pharmacological treatments for COVID-19, scientists are trying their best to re-purpose currently available drugs for immediate use on COVID-19 (see more detailed discussion later); among these, Lopinavir is one candidate. Lopinavir is formulated in combination with another protease inhibitor, Ritonavir, branded as Kaletra or Aluvia. Interestingly, Ritonavir is also top-ranked in Like SARS-CoV and MERS-CoV, the SARS-CoV-2 virus is a single-stranded RNA beta-coronavirus. These viruses enter host cells and replicate, producing strands that contain multiple copies of the viral genetic material (e.g., RNA). The strands of genetic material

Table 3 Top-ranked genes in terms of the popularity index (PI) in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	ACE-2	ACE-2	ACE-2	Spike
2	C-reactive protein	Spike	C-reactive protein	TMPRSS2
3	CRP	C-reactive protein	Spike	CRP
4	AST	CD8	CRP	RdRp
5	CD4	CD4	CD8	C-reactive protein
6	Aspartate aminotransferase	CRP	CD4	IL-1
7	Spike	IL-10	AST	CD4
8	IL-1	IL-8	TMPRSS2	CD8
9	IFN-gamma	mTOR	IL-10	SAA
10	Albumin	TNF	IL-8	IL-8

Table 4 Top-ranked chemicals in terms of PI in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	Lopinavir	Lopinavir	Ritonavir	Hydroxychloroquine
2	Ritonavir	Ritonavir	Lactate	PPE
3	Abidol	Nitrogen	Lopinavir	Lopinavir
4	Lactate	Lactate	Hydroxychloroquine	Ritonavir
5	Alanine	Urea	Urea	Lactate
6	Bilirubin	ADP	Bilirubin	Serine
7	Ser	Serine	PPE	NO2
8	Sofosbuvir	Pangolins	Ribavirin	CO2
9	Hydrogen	Cysteine	Nitrogen	Corticosteroids
10	Urea	Alanine	Hydrogen	Pangolins

accumulate at the periphery of the cell, ready to be cleaved, packaged, and prepared for release from the host cell. The enzyme 3-chymotrypsin-like protease (3CL^{Pro}) plays a crucial role in processing the viral RNA (Zhang et al., 2020, p. 2). As Lopinavir/Ritonavir is a protease inhibitor, it may inhibit the action of 3CL^{Pro}, thereby disrupting the process of viral replication and release from host cells (Zhang et al., 2020, p. 2). Recent evidence suggests that lopinavir has antiviral activity against SARS-CoV-2 in vitro (Choy et al., 2020). However, coronavirus proteases, including 3CL^{Pro}, do not contain a C2-symmetric pocket, which is the target of HIV protease inhibitors, leading some to question the potential potency of HIV protease inhibitors in treating these viruses (Sheahan et al., 2020). Darunavir, another HIV protease inhibitor, is reportedly not active against SARS-CoV-2 in an unpublished in vitro study,⁴ and a recent study using in vitro and mouse models found stronger evidence for anti MERS-CoV activity for the antiviral Remdesivir compared to Lopinavir/Ritonavir (Sheahan et al., 2020).

⁴ <https://www.njn.com/lack-of-evidence-to-support-darunavir-based-hiv-treatments-for-coronavirus>.

Popularity index (PI)

Tables 3 and 4 present top-ranked genes and chemicals, respectively, when we calculate the popularity index (PI). From the tables, we see that ACE-2 still ranks first in the first and the third columns, indicating that people are increasingly focusing on this gene from January to February, and from March to April 2020. We also observe that the genes in Tables 3 have some overlaps with those in Table 1.

We can also observe from Table 3 that C-reactive protein and Spike are also top-ranked in these periods. C-reactive protein (CRP) is a protein made by the liver. CRP levels in the blood increase when there is a condition causing inflammation somewhere in the body. A CRP test measures the amount of CRP in the blood to detect inflammation due to acute conditions or to monitor the severity of disease in chronic conditions. Existing studies have presented CRP levels of patients in the early stage of COVID-19 (Wang, 2020, p. 19).

We observe that Lopinavir and Ritonavir keep ranking high in January, February, March and April. Meanwhile, we find that Hydroxychloroquine ranks fourth in March 2020 but jumps to the first one in April 2020 albeit ranking top 10 in neither January nor February. Hydroxychloroquine is a medication used to prevent and treat malaria in areas where malaria remains sensitive to chloroquine. Other uses include treatment of rheumatoid arthritis, lupus, and porphyria cutanea tarda. Common side effects include vomiting, headache, changes in vision, and muscle weakness. Severe side effects may include allergic reactions, vision problems, and heart problems. Although all risk cannot be excluded, it remains a treatment for rheumatic disease during pregnancy. The U.S. President, Trump, has actively promoted the possibility that the antimalarial drugs chloroquine and hydroxychloroquine might prove to be miracle cures for COVID-19. In April, the Bill & Melinda Gates Foundation, the Wellcome Trust, and Mastercard announced the launch of the COVID-19 Therapeutics Accelerator, which is a ~\$125 million fund to address the coronavirus pandemic. The Accelerator has started by looking at how existing drugs can help treat COVID-19, a typical drug repositioning process. Hydroxychloroquine is an important one among these funded drugs. However, one recent publication on the *New English Journal of Medicine* studied 1.3+ thousand COVID-19 patients in the New York City, and found that there is no statistical significance in terms of the impact on the risk of the most severe outcomes from COVID-19 (Geleris et al., 2020, p. 19). As an observational study, this research employed the propensity score strategy to compare the patients in the study and the control groups. The paper, finally accepted for publication on May 7, 2020, observed that “patients who had received hydroxychloroquine were more likely to have had a primary end-point event than were patients who did not (hazard ratio, 2.37; 95% CI, 1.84 to 3.02)” (p. 6). In that paper, the authors also highlighted that the clinical guidance in the affiliation of the authors had already been updated by removing this drug from the guidance file, indicating that hydroxychloroquine may not have effects on COVID-19 treatments.

Although recommended for clinical trial by IDSA 38, NIH COVID-19 Treatment Guidelines Panel states that clinical data are insufficient to recommend either for or against use of hydroxychloroquine for the treatment of COVID-19.⁵ China established several hydroxychloroquine-related pilot studies; nevertheless, the efficacy and safety of hydroxychloroquine for treatment or prevention of COVID-19 has not yet been established till

⁵ <https://www.covid19treatmentguidelines.nih.gov/>.

Table 5 Top-ranked genes in terms of the collaboration index (CI) in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	ACE-2	ACE-2	ACE-2	TMPRSS2
2	C-reactive protein	CD8	C-reactive protein	Spike
3	AST	IL-6	CRP	RdRp
4	CD4	C-reactive protein	Spike	IL-6
5	Calcium	IL-10	IL-6	CRP
6	Albumin	CD4	CD8	C-reactive protein
7	CRP	CRP	CD4	CD4
8	Aspartate aminotransferase	Spike	TMPRSS2	CTSL
9	IL-6	TNF	CD147	ACE
10	ALB	Albumin	IL-1	IL-1

Table 6 Top-ranked chemicals in terms of the collaboration index (CI) in 2020 COVID-19 literature

Rank	2020/1	2020/2	2020/3	2020/4
1	Lopinavir	Lactate	Lactate	Hydroxychloroquine
2	Ritonavir	Nitrogen	Lopinavir	PPE
3	Abidol	Urea	Ritonavir	Ritonavir
4	Bilirubin	PhJ	Hydroxychloroquine	Lopinavir
5	5-OH	Pangolins	Urea	Serine
6	Alanine	Ebselen	Dexamethasone	Lactate
7	Lactate	Aspartate	Bilirubin	CO2
8	Ribavirin	Lopinavir	Nitrogen	Bilirubin
9	Serine	Ritonavir	Aspartate	Aspartate
10	Aspartate	Cysteine	–	Corticosteroids

now.⁶ Till May 8, 2020, there are at least 10 clinical trials to evaluate hydroxychloroquine for treatment of COVID-19 are registered at clinicaltrials.gov, and 10+ clinical trials for prevention of COVID-19.⁷ On May 22, a *Lancet* paper (Mehra et al., 2020a) reported an increasing risk of using hydroxychloroquine/chloroquine with 96 K+ patients’ experiments and thus concluded that they could not confirm the benefit of hydroxychloroquine/chloroquine. However, this paper was finally retracted by the *Lancet* on June 3 (Mehra et al., 2020b). A following *Science* comment paper pointed out that “a mysterious company’s coronavirus papers in top medical journals may be unraveling”.⁸

⁶ <https://www.fda.gov/media/137250/download>.

⁷ <https://www.ashp.org/-/media/assets/pharmacy-practice/resource-centers/Coronavirus/docs/ASHP-COVID-19-Evidence-Table.ashx?la=en&hash=B414CC64FD64E1AE8CA47AD753BA744EDF4FFB8C>.

⁸ <https://www.sciencemag.org/news/2020/06/mysterious-company-s-coronavirus-papers-top-medical-journals-may-be-unraveling>.

Table 7 Top-ranked genes in terms of the betweenness, PI, and CI

	Betweenness	PI	CI
Same genes	ACE-2, CD4, CD8, CPR, AST, IL-10, IL-1, spike TNF, aspartate aminotransferase, albumin		
Different genes	cTnI, LCN-2, granzyme B, gremlin-1, PARPs, LPS, IL-1beta, AST, CTX, AKt2, IL-17, IL-12, IGFBP- 2, IGFBP-3 IL-8, mTOR	IL-1, RdRp, TMPRSS2 Calcium	IL-6, SAA, IFN-gamma, CD147, CTSL, ACE
	ALB		ALB

Collaboration index (CI)

Tables 5 and 6 show top-ranked genes and chemicals in terms of collaboration index in January, February, March, and April 2020. when we calculate the collaboration index (CI). We can see that most genes and chemicals in the two tables are quite similar to those in previous tables, indicating that almost all publications are collaborative in this domain.

Comparison

We have already listed top-ranked bio-entities based on betweenness, PI, and CI. Nonetheless, we found that there are many top-ranked bio-entities that are duplicated in these results. To this end, we need to compare the genes and chemicals described by the three measurements to illustrate the similarities and differences.

Table 7 shows the similarities and differences of genes in betweenness, PI, and CI ranking algorithms. It can be seen that ACE-2, CD4, CD8, CPR, and other indicators have appeared in the three indicators, and the ranking is relatively high. CTnI and lcn-2 were only found in the betweenness index. IL-6, SAA, and IFN gamma only appeared in CI, and Calcium only appeared in PI index. IL-8 and mTOR appeared in betweenness and PI, while IL-1, RdRp, and TMPRSS2 appeared in PI and CI.

Table 8 shows the similarities and differences of chemicals in betweenness, PI, and CI ranking algorithms. It can be seen that Lopinavir, ritonavir, lactate and other chemicals appear in all three rankings. Ser only appeared in the betweenness index. Sofosbuvir, NO2 and ADP appeared in betweenness and PI. Common chemicals of betweenness and CI include aspartate, 5-OH, PHJ, and ebselen.

A case of ACE-2

Per the results in the previous section, ACE-2 ranks quite top in many tables. ACE-2 counters the activity of the related angiotensin-converting enzyme (ACE) by reducing the

Table 8 Top-ranked chemicals in terms of the betweenness PI and CI in 2020 COVID-19 literature

	Betweenness	PI	CI
Same chemicals	Lopinavir, ritonavir, lactate, ribavirin, alanine, abidol, nitrogen, corticosteroids, bilirubin, CO ₂ , PPE, Hydroxychloroquine, urea, serine, pangolins, cysteine		
Different chemicals	sofosbuvir, NO ₂ , ADP, Ser, hydrogen Ser Aspartate, 5-OH, PhJ, Ebselen		Aspartate, 5-OH, PhJ, Ebselen

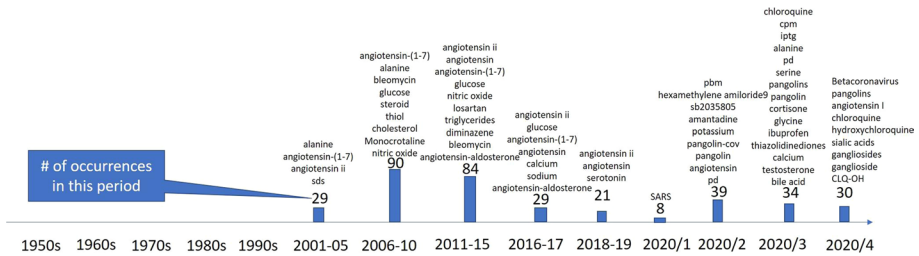


Fig. 2 Evolution of ACE-2 and co-occurred chemicals

amount of angiotensin-II and increasing angiotensin-(1–7), making it a promising drug target for treating cardiovascular diseases. ACE-2 is also quite an important protein for potential COVID-19 treatments by its popularity in COVID-19 literature.

Given the popularity of ACE-2 and its close relationship with COVID-19, we implement a case study on ACE-2 to examine their evolutions over years based on our dataset. We divide 1951–2020 into several time periods. Before 2000, we divided periods into a 10-year-long time window; between 2001 and 2019, we selected 2001–2005, 2006–2010, 2011–2015, 2016–2017, and 2018–2019 windows. As for 2020, we particularly adopt month-long time windows. Figure 2 shows the evolution of ACE-2 over years, in which the height of bins is proportional to its number of occurrences in literature in a given period—in 2006–2010, for example, ACE-2 occurs 16 times in our dataset. Names of entities above each bin present the co-occurred chemicals in that period.⁹ From Fig. 1, we can see that before 2020, ACE-2 was studied more frequently between 2011 and 2015 (43 times), probably because of the pandemic of Middle East respiratory syndrome (MERS, 2012–). After the breaking out of COVID-19 (January 2020), ACE-2 started to be mentioned, with 11 times in February and 14 in March. The fact that ACE-2 was not mentioned in January is probably attributable to the delay of scientific research and publishing. Figure 1 also shows that in 2006–2010, angiotensin, bleomycin, octapeptide ANG II, and saralasin are four co-occurred chemicals with ACE-2. Among these, angiotensin is a peptide hormone that causes vasoconstriction and an increase in blood pressure, which is quite related to ACE-2 because both of them are involved in the process of blood pressure regulation. In 2011–2015, an increasing number of chemicals can be found together with ACE-2 in literature. For example, Angiotensin (1–7) is an active heptapeptide of the renin–angiotensin system. It is shown that Angiotensin (1–7) is a vasodilator agent that plays important roles in cardiovascular. Figures 2 and 3 show the co-occurred entities of ACE-2 after removing low-frequency co-occurred entities, in which we can see that renin, ANG-II, and angiotensin II are all top co-occurred genes of ACE-2, which is similar to the results in Fig. 1.

Recognition that ACE-2 is the coreceptor for the coronavirus has prompted new therapeutic approaches to block the enzyme or reduce its expression to prevent the cellular entry and SARS-CoV-2 infection in tissues that express ACE-2 including lung, heart, kidney, brain, and gut. ACE-2, however, is a key enzymatic component of the renin–angiotensin–aldosterone system (RAAS); ACE-2 degrades ANG II, a peptide with multiple actions that promote CVD, and generates Ang-(1–7), which antagonizes the effects of ANG II.

⁹ Note that we remove some meaningless chemicals from the figure manually.



Fig. 3 Co-occurred genes (left) and chemicals (right) of ACE-2

Moreover, experimental evidence suggests that RAAS blockade by ACE inhibitors, ANG II type 1 receptor antagonists, and mineralocorticoid antagonists, as well as statins, enhance ACE-2 which, in part, contributes to the benefit of these regimens. In lieu of the fact that many older patients with hypertension or other CVDs are routinely treated with RAAS blockers and statins, new clinical concerns have developed regarding whether these patients are at greater risk for SARS-CoV-2 infection, whether RAAS and statin therapy should be discontinued, and the potential consequences of RAAS blockade to COVID-19-related pathologies such as acute and chronic respiratory disease. The current perspective critically examines the evidence for ACE-2 regulation by RAAS blockade and statins, the cardiovascular benefits of ACE-2, and whether ACE-2 blockade is a viable approach to attenuate COVID-19.

Conclusions

This paper presents an entitymetric analysis on COVID-19 literatures. We construct an entity–entity co-occurrence network and employ network indicators to analyze the extracted entities. We find that ACE-2 and C-reactive protein are two very important genes and that lopinavir and ritonavir are two very important chemicals, regardless of the results from which rankings. However, there are several limitations of the current paper. For instance, we only extract the entity co-occurrence relations from COVID-19 papers. In the future, we plan to adopt more advanced algorithms to extract various relations among entities, especially their semantic relations. We believe that these will improve the quality and the performance of entitymetrics.

Acknowledgements The authors are grateful to Allen Institute of AI for publicly sharing the COVID-19 open research dataset and Drs. Zhiyong Lu and Chih-Hsuan Wei for their kind help on PubTator usage. The authors thank the anonymous reviewer to offer many suggestions to improve the quality of this paper. Qi Yu acknowledges financial supports from National Natural Science Foundation of China (Grant Number: 71573162) and the Shanxi Scholarship Council of China (Grant Number: HGKY2019057). This work was supported by Min Song acknowledges financial supports from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (Grant Number: 2020S1A5B1104865). Ying Ding

acknowledges financial support from the NSF RAPID (Grant Number: 2028717). Baitong Chen acknowledges financial support from the Ministry of Education of China Project of Humanities and Social Sciences (Grant Number: 18YJC870002).

Declaration

Conflict of interest The authors declare no competing interests.

References

- Alhazzani, W., Møller, M. H., Arabi, Y. M., Loeb, M., Gong, M. N., Fan, E., Oczkowski, S., Levy, M. M., Derde, L., & Dzierba, A. (2020). Surviving sepsis campaign: Guidelines on the management of critically ill adults with Coronavirus Disease 2019 (COVID-19). *Intensive Care Medicine*, 1–34.
- Amano-Patiño, N., Faraglia, E., Giannitsarou, C., & Hasna, Z. (2020). Who is doing new research in the time of COVID-19? Not the female economists. *VoxEU.Org*. <https://voxeu.org/article/who-doing-new-research-time-covid-19-not-female-economists>
- Bell, L., Chowdhary, R., Liu, J. S., Niu, X., & Zhang, J. (2011). Integrated Bio-Entity network: A system for biological knowledge discovery. *PLoS ONE*, 6(6), e21474. <https://doi.org/10.1371/journal.pone.0021474>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Chahrour, M., Assi, S., Bejjani, M., Nasrallah, A. A., Salhab, H., Fares, M. Y., & Khachfe, H. H. (2020). A Bibliometric Analysis of COVID-19 Research Activity: A Call for Increased Output. *Cureus*, 12(3).
- Choy, K.-T., Wong, A. Y.-L., Kaewpreedee, P., Sia, S.-F., Chen, D., Hui, K. P. Y., Chu, D. K. W., Chan, M. C. W., Cheung, P. P.-H., & Huang, X. (2020). Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Research*, 104786.
- Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A., & Einav, S. (2020). A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *Journal of Critical Care*.
- COVID-19 Open Research Dataset (CORD-19). (2020). *CORD-19 | Semantic Scholar*. <https://pages.semanticscholar.org/coronavirus-research>
- Dehghanbanadaki, H., Seif, F., Vahidi, Y., Razi, F., Hashemi, E., Khoshmirsafa, M., & Aazami, H. (2020). Bibliometric analysis of global scientific research on Coronavirus (COVID-19). *Medical Journal of the Islamic Republic of Iran*, 34, 51.
- Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013). Entitymetrics: Measuring the Impact of Entities. *PLoS ONE*, 8(8). <https://doi.org/10.1371/journal.pone.0071416>
- 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*.
- Farooq, R. K., Rehman, S. U., Ashiq, M., Siddique, N., & Ahmad, S. (2021). Bibliometric analysis of coronavirus disease (COVID-19) literature published in Web of Science 2019–2020. *Journal of Family and Community Medicine*, 28(1), 1.
- Fry, C. V., Cai, X., Zhang, Y., & Wagner, C. (2020). Consolidation in a crisis: Patterns of International Collaboration in COVID-19 Research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3595455>
- Geleris, J., Sun, Y., Platt, J., Zucker, J., Baldwin, M., Hripcsak, G., Labella, A., Manson, D., Kubin, C., Barr, R. G., Sobiles, M. E., & Schluger, N. W. (2020). Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. *New England Journal of Medicine*, 0(0), null. <https://doi.org/10.1056/NEJMoa2012410>
- Giamarellos-Bourboulis, E. J., Netea, M. G., Rovina, N., Akinosoglou, K., Antoniadou, A., Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M.-E., & Katsaounou, P. (2020). Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host & Microbe*.
- Gysi, D. M., Valle, Í. D., Zitnik, M., Ameli, A., Gan, X., Varol, O., Sanchez, H., Baron, R. M., Ghiasian, D., Loscalzo, J., & Barabási, A.-L. (2020). Network medicine framework for identifying drug repurposing opportunities for covid-19. *ArXiv Preprint*. arXiv:2004.07229v1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7280907/>.
- Hamidah, I., Sriyono, S., & Hudha, M. N. (2020). A Bibliometric analysis of Covid-19 research using VOSviewer. *Indonesian Journal of Science and Technology*, 5(2), 209–216.

- Hossain, M. M. (2020). Current Status of Global Research on Novel Coronavirus Disease (COVID-19): A Bibliometric Analysis and Knowledge Mapping. *Available at SSRN 3547824*.
- 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.
- Lee, K., Kim, S., Kim, E.H.-J., & Song, M. (2017). Comparative evaluation of bibliometric content networks by tomographic content analysis: An application to Parkinson's disease. *Journal of the Association for Information Science and Technology*, 68(5), 1295–1307. <https://doi.org/10.1002/asi.23752>
- Li, C.-X., Li, Y.-Y., He, L.-P., Kou, J., Bai, J.-S., Liu, J., Tian, B., Cao, L.-J., Wang, K.-H., & Kuang, Y.-Q. (2019). The predictive role of CD4+ cell count and CD4/CD8 ratio in immune reconstitution outcome among HIV/AIDS patients receiving antiretroviral therapy: An eight-year observation in China. *BMC Immunology*, 20(1), 31.
- Li, X., Rousseau, J. F., Ding, Y., Song, M., & Lu, W. (2020). Understanding drug repurposing from the perspective of biomedical entities and their evolution: A bibliographic research using aspirin. *JMIR Medical Informatics*.
- Ludvigsson, J. F. (2020). Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatrica*, 109(6), 1088–1095.
- Mehra, M. R., Desai, S. S., Ruschitzka, F., & Patel, A. N. (2020a). RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6)
- Mehra, M. R., Desai, S. S., Ruschitzka, F., & Patel, A. N. (2020b). Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis. *The Lancet*.
- Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour*, 1–4.
- Ng, M.-Y., Lee, E. Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M. M., Lo, C. S.-Y., Leung, B., & Khong, P.-L. (2020). Imaging profile of the COVID-19 infection: Radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1), e200034.
- Park, J., Kim, T., & Song, M. (2017). Entitymetrics Analysis of the Research Works of Dong-ju Yun using Textmining. *Journal of the Korean BIBLIA Society for library and Information Science*, 28(1), 191–207. <https://doi.org/10.14699/kbiblia.2017.28.1.191>
- Rafiei Nasab, F., & Rahim, F. (2020). Bibliometric analysis of Global Scientific Research on SARS-CoV-2 (COVID-19) [Preprint]. *Health Informatics*. <https://doi.org/10.1101/2020.03.19.20038752>
- Ren, L.-H., Shen, Y.-Z., Ding, Y.-S., & Chou, K.-C. (2011). Bio-entity network for analysis of protein–protein interaction networks. *Asian Journal of Control*, 13(5), 726–737. <https://doi.org/10.1002/asjc.395>
- Rodriguez-Morales, A. J., Cardona-Ospina, J. A., Gutiérrez-Ocampo, E., Villamizar-Peña, R., Holguin-Rivera, Y., Escalera-Antezana, J. P., Alvarado-Arnez, L. E., Bonilla-Aldana, D. K., Franco-Paredes, C., & Henao-Martinez, A. F. (2020). Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Medicine and Infectious Disease*, 101623.
- Salehi, S., Abedi, A., Balakrishnan, S., & Gholamrezaezhad, A. (2020). Coronavirus disease 2019 (COVID-19): A systematic review of imaging findings in 919 patients. *American Journal of Roentgenology*, 1–7.
- Sheahan, T. P., Sims, A. C., Leist, S. R., Schäfer, A., Won, J., Brown, A. J., Montgomery, S. A., Hogg, A., Babusis, D., & Clarke, M. O. (2020). Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. *Nature Communications*, 11(1), 1–14.
- Song, M. (2016). Analyzing Impact, Trend, and Diffusion of Knowledge associated with Neoplasms Research. *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, 1. <https://www.aclweb.org/anthology/W16-4701>
- Song, M., Han, N.-G., Kim, Y.-H., Ding, Y., & Chambers, T. (2013). Discovering implicit entity relation with the gene-citation-gene network. *PLoS ONE*, 8(12), e84639. <https://doi.org/10.1371/journal.pone.0084639>
- South, A. M., Diz, D. I., & Chappell, M. C. (2020). COVID-19, ACE2, and the cardiovascular consequences. *American Journal of Physiology. Heart and Circulatory Physiology*, 318(5), H1084–H1090. <https://doi.org/10.1152/ajpheart.00217.2020>
- Vardavas, C. I., & Nikitara, K. (2020). COVID-19 and smoking: A systematic review of the evidence. *Tobacco Induced Diseases*, 18.
- Vincent-Lamarre, P., Cassidy, S., & Vincent, L. (2020). The decline of women's research production during the coronavirus pandemic. *Nature Index*.
- Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *The Lancet Child & Adolescent Health*.

- Wang, L. (2020). C-reactive protein levels in the early stage of COVID-19. *Medecine Et Maladies Infectieuses*. <https://doi.org/10.1016/j.medmal.2020.03.007>
- Wei, C.-H., Kao, H.-Y., & Lu, Z. (2013). PubTator: A web-based text mining tool for assisting biocuration. *Nucleic Acids Research*, *41*(W1), W518–W522.
- World Health Organization. (2020). *World Health Organization*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- Xu, J., Kim, S., Song, M., Jeong, M., Kim, D., Kang, J., Rousseau, J. F., Li, X., Xu, W., Torvik, V. I., Bu, Y., Chen, C., Ebeid, I. A., Li, D., & Ding, Y. (2020). Building a PubMed knowledge graph. *Scientific Data*, *7*(1), 1–15.
- Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., & Zhou, Y. (2020). Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: A systematic review and meta-analysis. *International Journal of Infectious Diseases*.
- Zhai, F., Zhai, Y., Cong, C., Song, T., Xiang, R., Feng, T., Liang, Z., Zeng, Y., Yang J., & Liang, J. (2020). Research progress of coronavirus based on bibliometric analysis. *International Journal of Environmental Research and Public Health*, *17*(11), 3766.
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*, *368*(6489), 409–412.

Authors and Affiliations

Qi Yu^{1,2} · Qi Wang^{1,3} · Yafei Zhang¹ · Chongyan Chen⁴ · Hyeoung Ryu⁵ · Namu Park⁶ · Jae-Eun Baek⁷ · Keyuan Li⁸ · Yifei Wu⁹ · Daifeng Li¹⁰ · Jian Xu¹⁰ · Meijun Liu^{11,12} · Jeremy J. Yang¹⁴ · Chenwei Zhang¹³ · Chao Lu¹⁵ · Peng Zhang⁹ · Xin Li¹⁶ · Baitong Chen¹⁷ · Islam Akef Ebeid⁴ · Julia Fensel¹⁸ · Chao Min¹⁹ · Yujia Zhai^{16,20} · Min Song²¹ · Ying Ding^{4,22} · Yi Bu²³ 

¹ Institute of Medical Data Sciences, Shanxi Medical University, Taiyuan, China

² School of Management, Shanxi Medical University, Taiyuan, China

³ Key Laboratory of Cellular Physiology (Shanxi Medical University), Ministry of Education, Taiyuan, China

⁴ School of Information, University of Texas, Austin, TX, USA

⁵ The Information School, University of Washington, Seattle, WA, USA

⁶ Department of Digital Analytics, Yonsei University, Seoul, South Korea

⁷ Department of Library and Information Science, Dae-Gu University, Gyeongsan, South Korea

⁸ Department of Statistics, College of Arts and Sciences, Indiana University, Bloomington, IN, USA

⁹ Department of Computer Science and Technology, Tsinghua University, Beijing, China

¹⁰ School of Information Management, Sun Yat-Sen University, Guangzhou, Guangdong, China

¹¹ Institute for Global Public Policy, Fudan University, Shanghai, China

¹² LSE-Fudan Research Centre for Global Public Policy, Fudan University, Shanghai, China

¹³ Faculty of Education, the University of Hong Kong, Hong Kong, China

¹⁴ Department of Internal Medicine, School of Medicine, University of New Mexico, Albuquerque, NM, USA

¹⁵ Business School, Hohai University, Nanjing, Jiangsu, China

¹⁶ School of Information Management, Wuhan University, Wuhan, Hubei, China

¹⁷ Department of Library, Information, and Archives, Shanghai University, Shanghai, China

-
- ¹⁸ Westlake High School, Austin, TX, USA
- ¹⁹ School of Information Management, Nanjing University, Nanjing, Jiangsu, China
- ²⁰ Department of Information Resource Management, Tianjin Normal University, Tianjin, China
- ²¹ Department of Library and Information Science, Yonsei University, Seoul, South Korea
- ²² Dell Medical School, University of Texas, Austin, TX, USA
- ²³ Department of Information Management, Peking University, Beijing, China