

Original Paper

# Global Research on Coronaviruses: An R Package

---

Thierry Warin, PhD

HEC Montreal, Montreal, QC, Canada

---

**Corresponding Author:**

Thierry Warin, PhD

HEC Montreal

3000, chemin de la Côte-Sainte-Catherine

Montreal, QC, H3T 2A7

Canada

Phone: 1 514 340 6185

Email: [thierry.warin@hec.ca](mailto:thierry.warin@hec.ca)

## Abstract

---

**Background:** In these trying times, we developed an R package about bibliographic references on coronaviruses. Working with reproducible research principles based on open science, disseminating scientific information, providing easy access to scientific production on this particular issue, and offering a rapid integration in researchers' workflows may help save time in this race against the virus, notably in terms of public health.

**Objective:** The goal is to simplify the workflow of interested researchers, with multidisciplinary research in mind. With more than 60,500 medical bibliographic references at the time of publication, this package is among the largest about coronaviruses.

**Methods:** This package could be of interest to epidemiologists, researchers in scientometrics, biostatisticians, as well as data scientists broadly defined. This package collects references from PubMed and organizes the data in a data frame. We then built functions to sort through this collection of references. Researchers can also integrate the data into their pipeline and implement them in R within their code libraries.

**Results:** We provide a short use case in this paper based on a bibliometric analysis of the references made available by this package. Classification techniques can also be used to go through the large volume of references and allow researchers to save time on this part of their research. Network analysis can be used to filter the data set. Text mining techniques can also help researchers calculate similarity indices and help them focus on the parts of the literature that are relevant for their research.

**Conclusions:** This package aims at accelerating research on coronaviruses. Epidemiologists can integrate this package into their workflow. It is also possible to add a machine learning layer on top of this package to model the latest advances in research about coronaviruses, as we update this package daily. It is also the only one of this size, to the best of our knowledge, to be built in the R language.

(*J Med Internet Res* 2020;22(8):e19615) doi: [10.2196/19615](https://doi.org/10.2196/19615)

---

**KEYWORDS**

COVID-19; SARS-CoV-2; coronavirus; R package; bibliometric; virus; infectious disease; reference; informatics

## Introduction

---

The coronavirus disease (COVID-19) outbreak finds its roots in Wuhan with potential first observations in late 2019 [1]. On December 30, 2019, clusters of cases of pneumonia of unknown origin were reported to the China National Health Commission. On January 7, 2020, a novel coronavirus (2019-nCoV) was isolated. Two previous outbreaks have taken place since the year 2000 involving coronaviruses: (1) severe acute respiratory syndrome-related coronavirus (SARS-CoV) and (2) Middle East respiratory syndrome-related coronavirus (MERS-CoV) [2].

In the context of the global propagation of the virus, numerous initiatives have occurred mobilizing our current global technological infrastructure: (1) the internet and server farms; (2) the convergence in coding languages; and (3) the use of data, structured and unstructured.

First, the internet is used for exchanges between universities, research laboratories, or political leaders to cite a few examples. Second, the convergence in coding languages, in particular functional languages such as R (R Foundation for Statistical Computing), Python (Python Software Foundation), or Julia, has helped facilitate the communications between researchers. Reproducible research has accelerated in these past few years,

with principles such as open science, open data, and open code. The use of data has been amplified by the development of new methodologies within the field of artificial intelligence (AI), allowing researchers to generate and analyze structured and unstructured data in a supervised way, as well as in a semi- or unsupervised way [3]. Data initiatives have flourished across the world, collecting firsthand data, aggregating various data sets, and developing simulation-based models. The Johns Hopkins University of Medicine has made a great effort in terms of data visualizations, which has disseminated throughout the world [4]. In doing so, it has undoubtedly helped to raise citizens' awareness, helped policy makers inform their population and avoid some potential fake news, or helped correct some misconceptions or confusion. The Johns Hopkins initiative has been followed by several others across the world, creating a large variety and diversity of data sets. This creation process allows researchers to benefit from different levels of granularity when it comes to the data dimensions such as geography, indicators, or methodologies; the open science characteristic is also an interesting aspect of the current research contributions [1].

With better access to these new technologies and methodologies, the breadth of expertise is more extensive; epidemiologists are leading the core of the research, but data scientists, biostatisticians, researchers in humanities, or social scientists can contribute and leverage their domain expertise using converging methodologies such as decision trees, text mining, or network analysis [5].

It is with these hypotheses in mind that we propose to replicate the spirit of the Allen Institute for AI initiative and to design an R package, whose main objective is to integrate easily in a researcher's workflow. This package is named EpiBibR.

EpiBibR stands for an “epidemiology-based bibliography for R.” The R package is under the Massachusetts Institute of Technology license and, as such, is a free resource based on the open science principles (reproducible research, open data, and open code). The resource may be used by researchers whose domain is scientometrics but also by researchers from other disciplines. For instance, the scientific community in AI and data science may use this package to accelerate new research insights about COVID-19. The package follows the methodology put in place by the Allen Institute and its partners [6] to create the COVID-19 data set with some differences. The latter is accessible through downloads of subsets or a representational state transfer application programming interface. The data provide essential information such as authors, methods, data, and citations to make it easier for researchers to find relevant contributions to their research questions. Our package proposes 22 features for the 60,500 references (on June 26, 2020), and access to the data has been made as easy as possible to integrate efficiently in almost any researcher's pipeline.

Through this package, a researcher can connect the data to her research protocol based on the R language. With this workflow in mind, a researcher can save time on collecting data and can use an accessible language to perform complex analytical tasks, for instance, be it in R or Python. Indeed, it is usual that

researchers use multiple languages (functional or not) to produce specific outputs. This workflow opens these data to analyses from the most extensive spectrum of potential options, enhancing multidisciplinary approaches applied to these data (biostatistics, bibliometrics, and text mining, among others).

The goal of this package in this emergency context is to limit the references to the medical domain (here the PubMed repository) but to then leverage the methodologies used across different disciplines. As we will address this point later, a further extension could be to add references from other disciplines to not only benefit from the wealth of methodologies but also their theories and concepts. For instance, to assess the spread of the disease, the literature—and theories—from researchers in demography would certainly be relevant.

## Methods

### Motivation

Across the world, a couple of initiatives have emerged whose goals are primarily to provide access to medical references. The main objective is to disseminate, as much as possible, the extensive research that has been done in the past (and recent history) to save some time and to improve the efficiency of further investigation. Research processes need to be efficient, and the time spent to perform this research needs to be relevant in this emergency. In addition, by proposing a (as much as possible) comprehensive data set of medical references on the coronavirus topic, the wisdom of crowds principle may play a role. A broader community beyond university researchers may use it and help shorten the time to the vaccine production. Researchers from pharmaceutical companies or other organizations may tap into these data to fine-tune their research and research processes.

A nonexhaustive list of the current bibliographic packages comprises the LitCovid data set from the National Library of Medicine (NLM) [7], the World Health Organization (WHO) data set [8], the “COVID-19 Research Articles Downloadable Database” from the Centers for Disease Control and Prevention (CDC) - Stephen B. Thacker CDC Library [9], and the “COVID-19 Open Research Dataset” by the Allen Institute for AI and their partners [6]. All these resources are essential and serve various complementary purposes. They are disseminated to their respective channels (ie, to their respective audiences). They are tailored to their specific needs. The LitCovid data set comprises 6530 references and can be downloaded from the United States NLM's website in a format that suits bibliographic software. It deals primarily with research about 2019-nCoV. The WHO's data set has around 9663 references, also specifically on COVID-19. The CDC's database is proposed in a software format as well (Microsoft Excel [Microsoft Corp] and bibliographic software formats) and comprises 17,636 references about COVID-19 and the other coronaviruses. The Allen Institute for AI's data set proposes over 190,000 references about COVID-19 as well as references about the other coronaviruses. It is accessible through different subsets of the overall database and a dedicated search engine. It also taps into a variety of academic article repositories.

In this context, the contributions made by the EpiBibR package are fourfold. First, with more than 60,500 references, EpiBibR is among the most extensive reference databases and is updated daily. The sheer number of references may be more suitable for a broader audience. Second, EpiBibR collects the data exclusively from PubMed to propose a controlled environment. Third, EpiBibR matches the keywords from the Allen Institute for AI's database to offer some consistency for researchers. Last, it is an R package and, as such, can be integrated into a workflow a little more efficiently than a file necessitating a specific software. Research teams can install the package in their systems and tap into it without the risks of version issues.

Beyond these four differentiation elements, EpiBibR is not better or worse than any other existing database. It just serves its audience and its purpose, like the other databases. It has not been created to replace a current database but, to the contrary, to complement these databases. We do believe that we need more initiatives in this domain at the world stage to support and integrate all the potential audiences and various workflows across the world. As a result, these initiatives would help accelerate research on coronaviruses overall and COVID-19 in particular.

### Functionality

As previously mentioned, EpiBibR is an R package to access bibliographic information on COVID-19 and other coronaviruses references easily. The package can be found at [10]. The command to install it is `remotes::install_github('warint/EpiBibR')`. We advise making sure the latest version of the package has been installed on each researcher's system.

The installation procedure can be found on the README file of this Github account. A full website with the various functions and examples are accessible from this page as well. The vignette has been created based on this paper to extend the use cases as more data are collected.

The references were collected via PubMed, a free resource that is developed and maintained by the National Center for Biotechnology Information at the United States NLM, located at the National Institutes of Health. PubMed includes over 30 million citations from biomedical literature.

More specifically, to collect our references, we adopted the procedure used by the Allen Institute for AI for their COVID-19 project. We applied a similar query on PubMed ("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") to build our bibliographic data.

To navigate through our data set, EpiBibR relies on a set of search arguments: author, author's country of origin, keyword in the title, keyword in the abstract, year, and the name of the journal. Each of them can genuinely help scientists and R users to filter references and find the relevant articles.

To simplify the workflow between our package and the research methodologies, the format of our data frame has been designed to integrate with different data pipelines, notably to facilitate the use of the R package Bibliometrix with our data [11].

The package comprises more than 60,500 references and 22 features (see [Textbox 1](#)).

**Textbox 1.** Field tags and their descriptions.

<b>AU</b>	Authors
<b>TI</b>	Document title
<b>AB</b>	Abstract
<b>PY</b>	Year
<b>DT</b>	Document type
<b>MESH</b>	Medical Subject Headings vocabulary
<b>TC</b>	Times cited
<b>SO</b>	Publication name (or source)
<b>J9</b>	Source abbreviation
<b>JI</b>	International Organization for Standardization source abbreviation
<b>DI</b>	Digital Object Identifier
<b>ISSN</b>	Source code
<b>VOL</b>	Volume
<b>ISSUE</b>	Issue number
<b>LT</b>	Language
<b>C1</b>	Author address
<b>RP</b>	Reprint address
<b>ID</b>	PubMed ID
<b>DE</b>	'Authors' Keywords
<b>UT</b>	Unique Article Identifier
<b>AU_CO</b>	'Author's country of origin
<b>DB</b>	

Bibliographic database

EpiBibR allows researchers to search academic references using several arguments: author, author's country of origin, author + year, keywords in the title, keywords in the abstract, year, and source name. Researchers can also download the entire

bibliographic data frame comprised of around 60,500 references with 22 metadata each.

In [Textbox 2](#), we provide the descriptions of the functions available in the R language to collect the relevant information.

**Textbox 2.** Descriptions of the functions to collect the relevant references.

**EpiBib\_data <- EpiBib\_references()**

Download the entire bibliographic data frame

**colson\_articles <- EpiBib\_author("Colson")**

Search all the articles written by Philippe Colson

**canada\_articles <- EpiBib\_country("canada")**

Search by 'author's country of origin

**yang2019 <- EpiBib\_AU\_YE(author = "yang", year = 2019)**

Search by author and year

**covid\_articles <- EpiBib\_title("covid")**

Search by keywords in the title

**coronavirus\_articles <- EpiBib\_abstract("coronavirus")**

Search by keywords in the abstract

**A2020\_articles <- EpiBib\_year(2020)**

Search by year

**bio\_articles <- EpiBib\_source("bio")**

Search by source

## Results

In what follows, a use case about how we can use such a data set is proposed. This section is not intended to offer a complete systematic literature review of the 60,500 references. This is the purpose of another article. However, we want to illustrate some powerful techniques that can be applied to this collection of references (for instance, social network analysis) while remaining at a very high level [12].

A systematic literature review consists of mainly four stages: (1) planning, (2) conducting, (3) analysis, and (4) synthesis and reporting. In the first stage, a preliminary study aims to build a corpus of articles citing the most relevant articles in the domain. The second stage is about producing a general review of the main topics used in the corpus. The third stage here is about making a cocitation analysis of the references in each corpus article. The last stage is about proposing a keywords co-occurrence analysis [13].

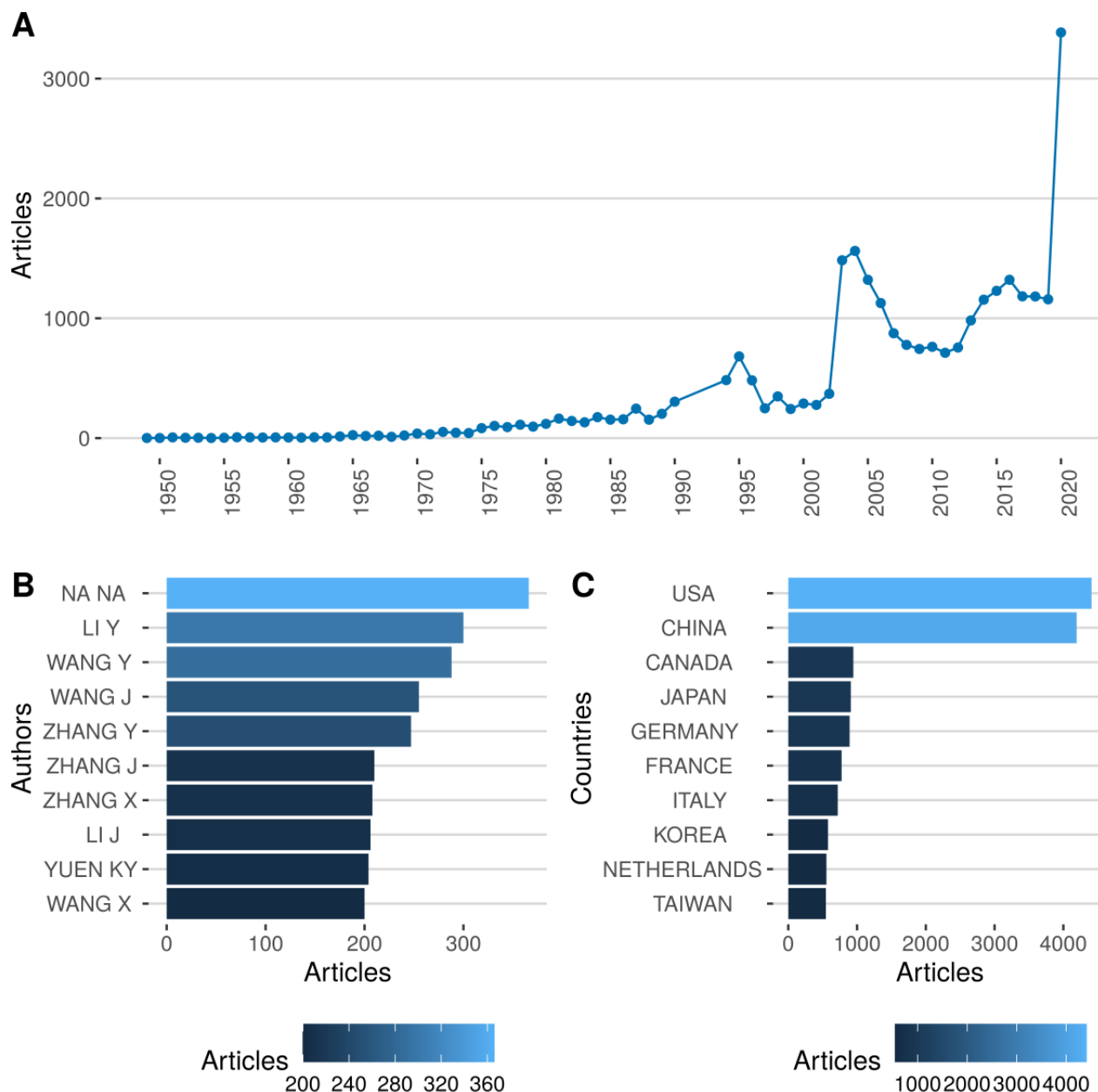
In our context, we propose a slightly modified perspective on a systematic literature review. The first stage is here at a different scale since we collect not just a few relevant articles to create the corpus but a vast, almost exhaustive, list of articles on a topic. It is worth noticing that the process of data selection

consists more of what we define as an "algorithmic systematic literature review," sometimes also referred to as "automation" [14]. An algorithmic systematic literature review comes with lots of benefits. The proposed modified systematic literature review improves the more classical approach since it does not rely on a manual search and extraction, with the potential biases and limitations it might create. An algorithmic systematic literature review combines the strength of both approaches: the power of big data with the academic soundness of the systematic literature review process. As such, it does not replace the expert's analysis of the literature. On the contrary, it should be used to augment the expert's study [3].

The *bibliometrix* package allows a thorough bibliometric analysis using R. Our EpiBibR data have been designed to integrate easily with the *bibliometrix* package. A shiny application is also available, called *biblioshiny()* [11]. This package has been used extensively for various exercises mobilizing massive amounts of data [15].

Let us first propose a simple count of the references on the coronaviruses literature. In [Figure 1](#), the historical development of research on coronaviruses can be analyzed as having three stages: exploration, initial development, and rapid development in the past year and the current year.

**Figure 1.** (A) Count of articles, (B) most productive authors, and (C) most productive countries.



In 2019, a little more than 1000 papers on coronaviruses had been produced, and in the first 5 months of 2020, close to 3400 papers were written on the topic. Those papers from the past 2 years seem to be an interesting, statistically representative sample. In Figure 1, we can also highlight the most productive authors as well as the most productive countries in terms of absolute counts. The most prolific authors provide interesting statistics since it is most likely to proxy research labs. In doing so, we can find which teams are working on which aspect of the coronaviruses. To illustrate our latter point, in Figure 2, we

first propose a compelling visualization, called a Sankey diagram. It links authors, keywords, and sources on a connected map. It is the first way to create groups of researchers. The results could be used by policy makers to identify areas of research on this topic.

We can also use powerful techniques such as Social Network Theory to find potential clusters of topics, clusters of researchers, and groups of country collaborations. Figure 3 is an example of the latter. The United States and China produce the bulk of the research on coronaviruses.

Figure 2. Sankey diagram: authors <math>\diamond</math> keywords <math>\diamond</math> sources.

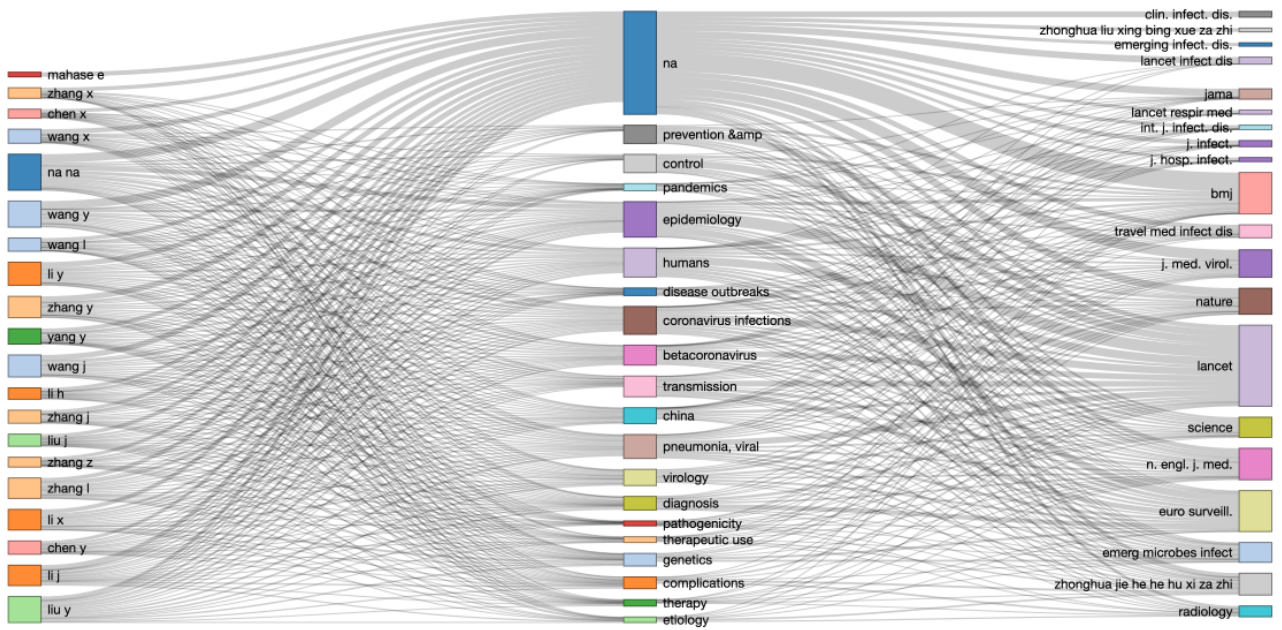


Figure 3. Country collaboration network.

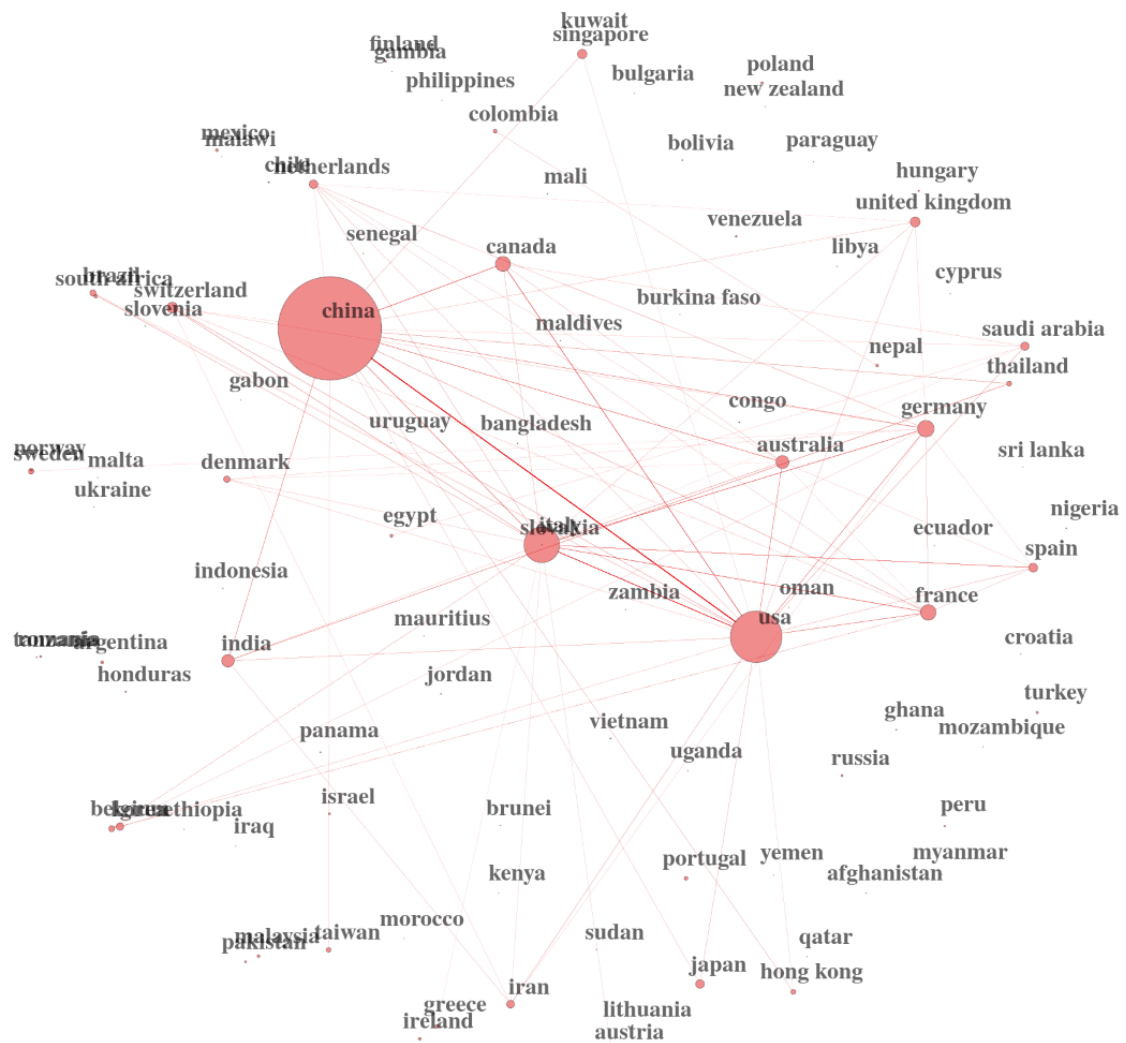


Figure 4 is an illustration of author collaboration networks. As previously mentioned, we remain at a very high level here.

However, policy makers or public health officers, for instance, could use these techniques to find more granular networks either

just within the 60,500 references or by crossing with other databases. We could even imagine crossing with unstructured data for some specific purposes [16].

Figure 5 is about finding clusters of topics. This technique can be applied to subsamples of the 60,500 references to provide a more granular analysis.

Figure 6 proposes indeed to go further on the topic dimension. For instance, we can study the evolution over time of the author's keywords usages.

Figure 4. Author collaboration network.

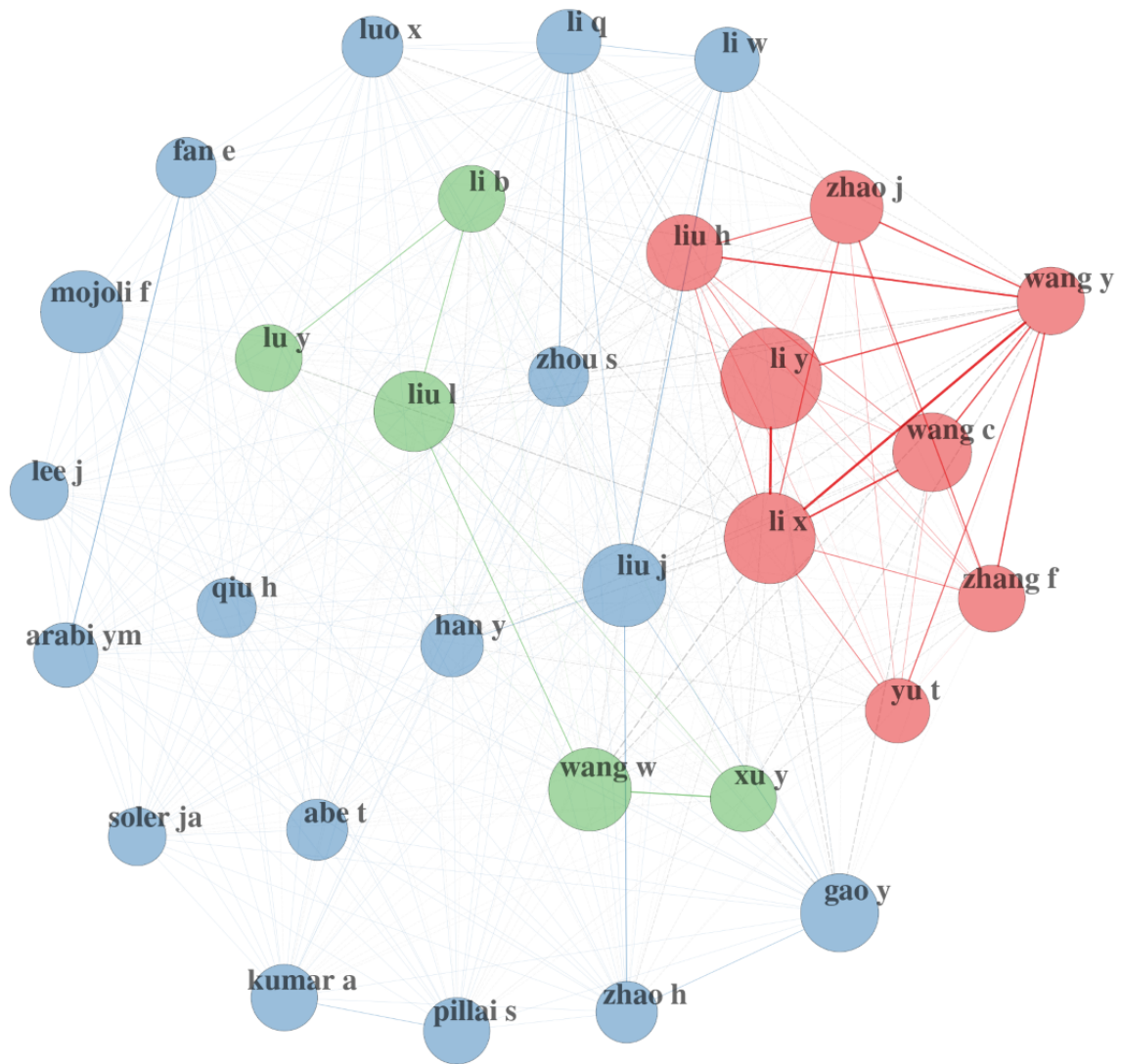




Figure 5. Thematic maps.

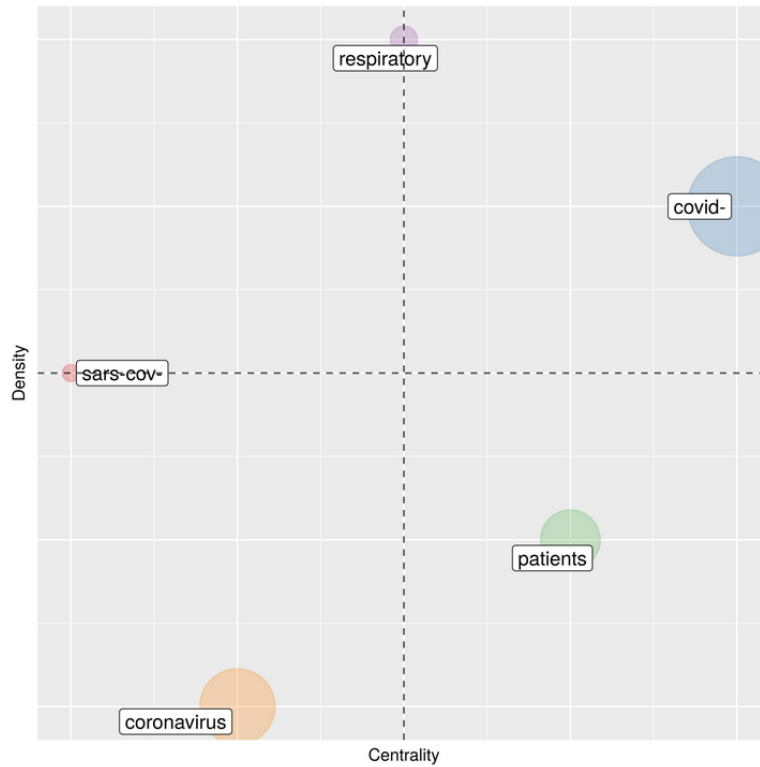
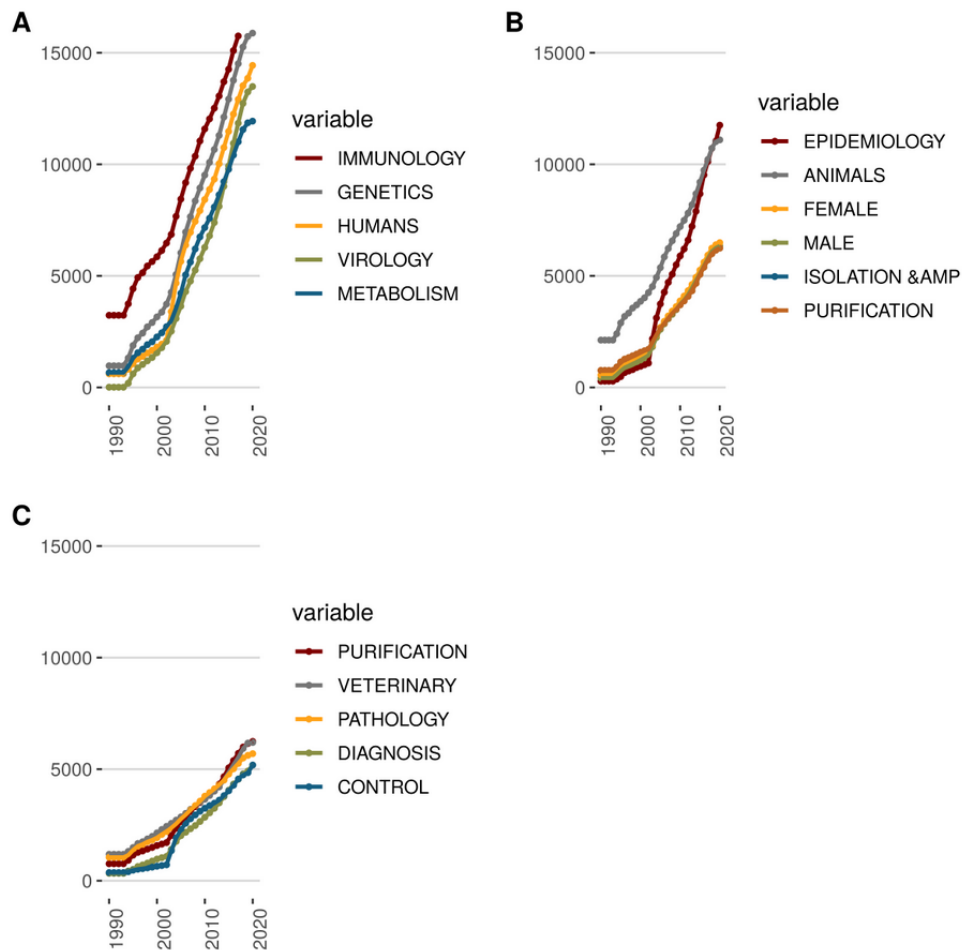


Figure 6. Author's keywords usage evolution over time.



As previously mentioned, this section is not intended to cover a systematic literature review but rather to illustrate some of the potential uses of powerful techniques or highlight some of the interesting questions that can be addressed. With a data set of 60,500 references, we have access to a wealth of information. Combined with the state of the technological development we have access to today, multiple questions can be answered. To go further, we could imagine analyzing document cocitations or reference burst detections [17].

## Discussion

### Further Developments

Recent developments in computing power, as well as data accessibility, offer new tools to develop policies to promote new capabilities or enhance existing skills as a way to encourage the further coevolution of new capabilities, echoing ideas put forward by Hirschman [18] more than 50 years ago. The difference is that now researchers, policy makers, and business analysts can analyze them in practice. It is capitalizing on the principles from the “wisdom of crowds” [19-22].

In this global pandemic, knowledge sharing and open data can have an impact on the solutions as well as the pace to discover the answers. With such a package, the easy dissemination through such an integrated workflow and low-level pipeline of tools may also help public policies. It allows the use of research evidence in health policy making [23].

Developing easy access to data and data modelization is of great importance for evidence-based policy making. In the past, there were lots of areas in public policy making where data were not accessible. As a result, decisions were made on assumptions coming from theoretical foundations or benchmarks from other sources. In our day and age, with more and more access to data across the world, being open data initiatives or not, evidence-based decisions are more and more possible. Numerous authors have demonstrated the role of data in informing better evidence-based policies [23-26].

This R package is updated daily when it comes to collecting the references and their metadata, and it will also be updated regularly to propose different use cases and new functionalities. We will update the modeling contribution of the package. For instance, we will integrate some of the *bibliometrix* package’s functions directly in our package to ease the scientometrist’s workflow. We will also include some models for network analysis and natural language processing–based studies.

### Acknowledgments

The author would like to thank Marine Leroi and Martin Paquette for their help in collecting the data and the maintenance of the package. The usual caveats apply.

### Conflicts of Interest

None declared.

### References

1. Wu T, Ge X, Yu G, Hu E. Open-source analytics tools for studying the COVID-19 coronavirus outbreak. medRxiv 2020:A. [doi: [10.1101/2020.02.25.20027433](https://doi.org/10.1101/2020.02.25.20027433)]

### Conclusion

In these trying times, we believe that working with reproducible research principles based on open science, disseminating scientific information, providing easy access to scientific production on this particular issue, and offering a rapid integration in researchers’ workflows may help save time in this race against the virus, notably in terms of public health. In this context, we believe the bibliometric packages made available by research institutions, nongovernmental organizations, or individual researchers complement the other data packages and help provide a more comprehensive understanding of the pandemics. One of the objectives is to reduce “the barrier for researchers and public health officials in obtaining comprehensive, up-to-date data on this ongoing outbreak. With this package, epidemiologists and other scientists can directly access data from four sources, facilitating mathematical modelling and forecasting of the COVID-19 outbreak” [1].

This package aims at providing this easy access and integration in a researcher’s workflow. It is specially designed to collect data and generate a data frame compatible with the *bibliometrix* package [11]. Such data sets may facilitate access to the right information. Moreover, the use of massive data sets crossed with robust data analyses may foster multidisciplinary perspectives, raising new questions and providing new answers [27-29]. Classification techniques can be used to go through the large volume of references and allow researchers to save time on this part of their research. Network analysis can be used to filter the data set. Text mining techniques can also help researchers calculate similarity indices and help them focus on the parts of the literature that are relevant for their research.

The package collects references that are interesting, for the most part, for the medical domain and allows multidisciplinary perspectives on this data set. It could be interesting to get views from other disciplines, for instance, mathematics, computer science, political science, economics, and environmental science. This is the result of the emergency in which humanity finds itself right now. We could also envisage later to add references from other disciplines such as social sciences and augment, or open, the perspectives on the issue. Not only would we benefit from a multidisciplinary perspective through the methodology dimension, as the goal is with our EpiBibR package, but we would also benefit from the multidisciplinary perspectives through the ontological concepts and theories of these added domains.

2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* 2020 Feb;395(10223):470-473. [doi: [10.1016/s0140-6736\(20\)30185-9](https://doi.org/10.1016/s0140-6736(20)30185-9)]
3. Warin T, Duc R, Sanger W. Mapping innovations in artificial intelligence through patents: a social data science perspective. 2017 Dec Presented at: International Conference on Computational Science and Computational Intelligence (CSCI). . pp. 252-257; 2017; Las Vegas, NV p. 252-257. [doi: [10.1109/csci.2017.40](https://doi.org/10.1109/csci.2017.40)]
4. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins Coronavirus Resource Center. URL: <https://coronavirus.jhu.edu/map.html> [accessed 2020-04-21]
5. Warin T, Sanger W. Connectivity and closeness among international financial institutions: a network theory perspective. *IJCM* 2018;1(3):225. [doi: [10.1504/ijcm.2018.094479](https://doi.org/10.1504/ijcm.2018.094479)]
6. COVID-19. Semantic Scholar. URL: <https://pages.semanticscholar.org/coronavirus-research> [accessed 2020-04-21]
7. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. *Nature* 2020 Mar;579(7798):193. [doi: [10.1038/d41586-020-00694-1](https://doi.org/10.1038/d41586-020-00694-1)] [Medline: [32157233](https://pubmed.ncbi.nlm.nih.gov/32157233/)]
8. Global research on coronavirus disease (COVID-19). World Health Organization. 2020. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov> [accessed 2020-04-24]
9. COVID-19 research articles downloadable database. Centers for Disease Control and Prevention. 2020. URL: <https://www.cdc.gov/library/researchguides/2019novelcoronavirus/researcharticles.html> [accessed 2020-04-24]
10. EpiBibR. GitHub. URL: <https://github.com/warint/EpiBibR>
11. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetrics* 2017 Nov;11(4):959-975 [FREE Full text] [doi: [10.1016/j.joi.2017.08.007](https://doi.org/10.1016/j.joi.2017.08.007)]
12. Camacho D, Panizo-Lledot Á, Bello-Orgaz G, Gonzalez-Pardo A, Cambria E. The four dimensions of social network analysis: an overview of research methods, applications, and software tools. *arXiv* 2020:A [FREE Full text] [doi: [10.1016/j.inffus.2020.05.009](https://doi.org/10.1016/j.inffus.2020.05.009)]
13. Tani M, Papaluca O, Sasso P. The system thinking perspective in the open-innovation research: a systematic review. *J Open Innovation Technol Market Complexity* 2018 Aug 18;4(3):38. [doi: [10.3390/joitmc4030038](https://doi.org/10.3390/joitmc4030038)]
14. Pulsiri N, Vatananan-Thesenvitz R. Improving systematic literature review with automation and bibliometrics. 2018 Presented at: 2018 Portland International Conference on Management of Engineering and Technology (PICMET); August 2018; Honolulu, HI. [doi: [10.23919/PICMET.2018.8481746](https://doi.org/10.23919/PICMET.2018.8481746)]
15. Zhao L, Tang Z, Zou X. Mapping the knowledge domain of smart-city research: a bibliometric and scientometric analysis. *Sustainability* 2019 Nov 25;11(23):6648. [doi: [10.3390/su11236648](https://doi.org/10.3390/su11236648)]
16. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res* 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: [10.2196/19016](https://doi.org/10.2196/19016)] [Medline: [32287039](https://pubmed.ncbi.nlm.nih.gov/32287039/)]
17. Kleinberg J. Bursty and hierarchical structure in streams. 2002 Presented at: KDD02: The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; July 2002; Edmonton, Alberta. [doi: [10.1145/775047.775061](https://doi.org/10.1145/775047.775061)]
18. Hirschman AO. *The Strategy of Economic Development*. New Haven, CT: Yale University Press; 1958.
19. Gal MS. The power of the crowd in the sharing economy. *Law Ethics Hum Rights* 2019;13:59. [doi: [10.1515/lehr-2019-0002](https://doi.org/10.1515/lehr-2019-0002)]
20. Cai CW, Gippel J, Zhu Y, Singh AK. The power of crowds: grand challenges in the Asia-Pacific region. *Aust J Manage* 2019 Sep 18;44(4):551-570. [doi: [10.1177/0312896219871979](https://doi.org/10.1177/0312896219871979)]
21. Wang PDC, Soares VS, de Souza JM, Esteves MGP, Esteves NCL, Duarte FR. A Crowd Science framework to support the construction of a Gold Standard Corpus for Plagiarism Detection. 2019 Presented at: 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD); May 2019; Porto, Portugal. [doi: [10.1109/cscwd.2019.8791853](https://doi.org/10.1109/cscwd.2019.8791853)]
22. Avasilcai S, Galateanu E. Co - creators in innovation ecosystems. Part II: crowdsprings 'crowd in action. *IOP Conf Ser: Mater Sci Eng* 2018 Sep 18;400:062001. [doi: [10.1088/1757-899x/400/6/062001](https://doi.org/10.1088/1757-899x/400/6/062001)]
23. Payán DD, Lewis LB. Use of research evidence in state health policymaking: menu labeling policy in California. *Prev Med Rep* 2019 Dec;16:101004 [FREE Full text] [doi: [10.1016/j.pmedr.2019.101004](https://doi.org/10.1016/j.pmedr.2019.101004)] [Medline: [31709136](https://pubmed.ncbi.nlm.nih.gov/31709136/)]
24. Wolffe TA, Whaley P, Halsall C, Rooney AA, Walker VR. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environ Int* 2019 Sep;130:104871 [FREE Full text] [doi: [10.1016/j.envint.2019.05.065](https://doi.org/10.1016/j.envint.2019.05.065)] [Medline: [31254867](https://pubmed.ncbi.nlm.nih.gov/31254867/)]
25. Giménez-Bertomeu VM, Domenech-López Y, Mateo-Pérez MA, de-Alfonseti-Hartmann N. Empirical evidence for professional practice and public policies: an exploratory study on social exclusion in users of primary care social services in Spain. *Int J Environ Res Public Health* 2019 Nov 20;16(23):4600 [FREE Full text] [doi: [10.3390/ijerph16234600](https://doi.org/10.3390/ijerph16234600)] [Medline: [31756963](https://pubmed.ncbi.nlm.nih.gov/31756963/)]
26. Villumsen S, Faxvaag A, Nøhr C. Development and progression in Danish eHealth policies: towards evidence-based policy making. *Stud Health Technol Inform* 2019 Aug 21;264:1075-1079. [doi: [10.3233/SHTI190390](https://doi.org/10.3233/SHTI190390)] [Medline: [31438090](https://pubmed.ncbi.nlm.nih.gov/31438090/)]
27. Warin T, Leiter DB. An empirical study of price dispersion in homogenous goods markets. Middlebury College, Department of Economics 2007:A [FREE Full text]
28. Warin T, Leiter D. Homogenous goods markets: an empirical study of price dispersion on the internet. *IJEER* 2012;4(5):514. [doi: [10.1504/ijebr.2012.048776](https://doi.org/10.1504/ijebr.2012.048776)]

29. De Marcellis-Warin N, Sanger W, Warin T. A network analysis of financial conversations on Twitter. *IJWBC* 2017;13(3):1. [doi: [10.1504/ijwbc.2017.10004118](https://doi.org/10.1504/ijwbc.2017.10004118)]

## Abbreviations

**AI:** artificial intelligence  
**CDC:** Centers for Disease Control and Prevention  
**COVID-19:** coronavirus disease  
**MERS-CoV:** Middle East respiratory syndrome–related coronavirus  
**NLM:** National Library of Medicine  
**SARS-CoV:** severe acute respiratory syndrome–related coronavirus  
**WHO:** World Health Organization  
**2019-nCoV:** novel coronavirus

*Edited by G Eysenbach; submitted 24.04.20; peer-reviewed by CM Dutra, J Yang, S Abdulrahman; comments to author 12.06.20; revised version received 26.06.20; accepted 26.07.20; published 11.08.20*

*Please cite as:*

Warin T

*Global Research on Coronaviruses: An R Package*

*J Med Internet Res* 2020;22(8):e19615

URL: <http://www.jmir.org/2020/8/e19615/>

doi: [10.2196/19615](https://doi.org/10.2196/19615)

PMID: [32730218](https://pubmed.ncbi.nlm.nih.gov/32730218/)

©Thierry Warin. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 11.08.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.