Global Research on Coronaviruses: An R Package

Thierry Warin

Submitted to: Journal of Medical Internet Research
on: April 24, 2020

Table of Contents

Global Research on Coronaviruses: An R Package

Thierry WarinPhD,

Corresponding Author:
Thierry WarinPhD,
Phone: +15146082106
Email: thierry.warin@skema.edu

Abstract

In these trying times, we developed an R package to simplify the workflow of interested researchers, with multidisciplinarity in mind. With more than 28,000 medical references, this package is among the largest ones about the medical references on coronaviruses. It is also the only one of this size - to the best of our knowledge - to be built in the R language. 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 organize the data in a dataframe. Then, we built functions to sort through the collection of references. Researchers can also integrate the data in their own pipeline and implement in R their own code libraries. We provide a short use case in this article based on a bibliometric analysis of the references made available by this package.

(JMIR Preprints 24/04/2020:19615)
DOI: https://doi.org/10.2196/preprints.19615

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Global Research on Coronaviruses: An R Package

Abstract

In these trying times, we developed an R package to simplify the workflow of interested researchers, with multidisciplinarity in mind. With more than 28,000 medical references, this package is among the largest ones about the medical references on coronaviruses. It is also the only one of this size - to the best of our knowledge - to be built in the R language. 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 organize the data in a dataframe. Then, we built functions to sort through the collection of references. Researchers can also integrate the data in their own pipeline and implement in R their own code libraries. We provide a short use case in this article based on a bibliometric analysis of the references made available by this package.

Keywords: covid-19, SARS-CoV-2, coronavirus, R package, bibliometrics

Introduction

The 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 (CoV) was isolated. Two previous outbreaks have taken place since 2000 involving coronaviruses: (1) SARS-Cov and (2) Middle East respiratory syndrome (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, political leaders, etc. Second, the convergence in coding languages, in particular functional languages such as R, Python 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, 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 florished across the world collecting first-hand data, aggregating various datasets, developing simulation-based models. 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 surely helped raising citizens' awareness, helped policymakers inform their population and avoid some potential fake news or helped correct some misconceptions or confusions. The Johns Hopkins initiative has been followed by several others across the world, creating a large variety and diversity of datasets. This creation process allows researchers to benefit from different levels of granularity when it comes to the data dimensions such as geography, indicators, methodologies, etc. the open science characteristic is also an interesting aspect of the current research contributions [1].

With a better access to these new technologies and methodologies, the breadth of expertise is wider: epidemiologists are leading the core of the research, but data scientists, bio-statisticians, 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 "epidemiology-based bibliography for R." The R package is under the MIT License and as such is a free resource based on the open science principles (reproducible research, open data, 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 artificial intelligence 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 CORD-19 dataset with some differences. The later is accessible through downloads of sub-sets or through a REST API. The data provide important 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 28,000 references (on April 21, 2020) and access to the data has been made as easy as possible in order to integrate efficiently in almost any researcher's pipeline.

Through this package, a researcher can connect the data to her analysis based on the R language. With this workflow in mind, a researcher can save time on collecting data, and can use a very accessible language to perform complex analytical tasks, be it R or Python for instance. Indeed, it is usual that researchers use multiple languages (functional or not) to produce certain outputs. This workflow opens these data to analyses from the largest spectrum of potential options, enhancing multidisciplinary approaches applied to these data (biostatistics, bibliometrics, text mining, etc.).

The goal of this package, in this emergency context, is to limit the references to the medical domain (here the PubMed repository), but then to 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 from their own theories and concepts. For instance, to assess the spread of the disease, the literature - and theories - from researchers in demography would certainly be relevant.

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 massive research that has been done in the past (and recent past) to save some time and to improve the efficiency of further research. Research processes need to be efficient and the time spent to perform this research needs to be relevant in this emergency situation. Also, by proposing a (as-much-as possible) comprehensive dataset of medical references on the coronavirus topic, it is also possible that the *wisdom of crowds* principle plays 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 non-exhaustive list of the current bibliographic packages comprises: the LitCovid dataset from the National Library of Medicine [7], the WHO dataset [8], the "COVID-19 Research Articles Downloadable Database" from the Centers for Disease Control and Prevention - 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. First of all, they are disseminated to their respective channels, i.e. to their respective audiences. They are tailored to their specific needs. The LitCovid dataset comprises 6,530 references and can be downloaded from the U.S. National Library of Medecine's website in a format that suits bibliographic softwares. It deals essentially with research about the 2019 novel Coronavirus. The WHO's dataset has around 9,663 references also specifically on Covid-19. CDC's database is proposed in software format as well (Microsoft Excel and bibliographic software formats) and comprises 17,636 references about Covid-19 and the other coronaviruses. The Allen Institute for AI's dataet proposes over 52,000 references about Covid-19 as well as references about the other coronaviruses. It is accessible through different subsets of the overall database and through a dedicated search engine. It also taps into a variety of academic articles repositories.

In this context, the contributions made by the EpiBibR package are fourfold. First of all, with more than 28,000 references, EpiBibR is thus among the largest references databases and is updated on a daily basis. The sheer number of references may be more suitable for a broader audience. Second, EpiBibR collects the data exclusively from PubMed in order to propose a controlled environment. Third, EpiBibR matches the keywords from the Allen Institute for AI's database in order to offer some consistency for researchers. Last and not least, it is an R package and as such can be integrated into a researchers' 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 4 differenciation elements, EpiBibR is not better or worse than any other existing databases. It just serves its own audience and its own purpose, like the other databases. It has not been created to replace an existing database but to the contrary to complement these databases. We do believe that we need more initiatives in this domain at the world stage in order to support and integrate all the potential audiences and various workflows across the world. As a result, these intitiatives would help accelerate research on coronaviruses overall and Covid-19 in particular.

Functionality

As aforementioned, EpiBibR is an R package to easily access bibliographic data on Covid-19 and other coronaviruses references. The package can be found at the following url: https://github.com/warint/EpiBibR. The command to install it is remotes::install_github('warint/EpiBibR'). We advise to make sure the latest version of the package has been installed on your 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 references were collected via PubMed, a free resource that is developed and maintained by the National Center for Biotechnology Information (NCBI), at the U.S. National Library of Medicine (NLM), located at the National Institutes of Health (NIH). PubMed includes over 30 million citations from the biomedical literature.

More specifically, to collect our references, we adopted the procedure used by the Allen Institute for AI for their CORD-19 project. We apply 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 own bibliographic data.

To navigate through our dataset, 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 truly help scientists and R users to filter references and find the relevant articles.

In an effort to simplify the workflow between our package and the research methodologies, the format of our dataframe has been designed to integrate with different data pipelines, notably to facilitate the use of the R package Bibliometrix with our data [10].

The package comprises more than 28,000 references and 22 features (See Table 1).

Table 1. Features accessible through the package.

Field Tags	Descriptions	Field Tags	Descriptions
AU	Authors	ISSN	Source Code
TI	Document Title	VOL	Volume
AB	Abstract	ISSUE	Issue Number
PY	Year	LT	Language
DT	Document Type	C1	Author Address
MESH	Medical Subject Headings Vocabulary	RP	Reprint Address
TC	Times Cited	ID	PubMed ID
SO	Publication Name (or Source)	DE	Authors’ Keywords
J9	Source Abbreviation	UT	Unique Article Identifier
JI	ISO Source Abbreviation	AU_CO	Author’s Country of Origin
DI	Digital Object Identifier (DOI)	DB	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. You can also download the entire bibliographic dataframe cmprised of around 28,000 references with 22 metadata each.

In the next table (Table 2), we provide the descriptions of the functions available in the R language to collect the relevant information.

Table 2. Descriptions of the functions to collect the relevant references.

Functions	Descriptions
EpiBib_data <- EpiBib_references()	Download the entire bibliographic dataframe
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 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

Use Case: An Algorithmic Systematic Literature Review

In what follows, a use case about how we can use such dataset is proposed. This section is not intended to propose a systematic literature review of the 28,000 reference. 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 [11].

A systematic literature review consists in essentially four stages: (1) planning, (2) conducting, (3) analysis and

(4) synthesis and reporting. In the first stage, a preliminary analysis aims to build a corpus of articles citing the most relevant articles in the domain. The second stage is about producing a general analysis of the main topics used in the corpus. The third stage here is about making a co-citation analysis of the references in each corpus article. The last stage is about proposing a keywords co-occurrence analysis [12].

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 huge - almost exhaustive - list of articles on a topic. It is worth noticing that the process of selection of the data consists more in fact to what we define as an “algorithmic systematic literature review”, sometimes also referred as “automation” [13]. An algortihmic 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. To the contrary, it should be used to augment the expert’s analysis [3].

The *bibliometrix* package allows a thorough bibliometric analysis using R. Our EpiBib data have been designed to integrate easily with the *bibliometrix* package. A shinyapp is also available `biblioshiny()` [10]. This package has been used extensively for various exercises mobilizing massive amounts of data [14].

Let us first propose a simple count of the references on the coronaviruses literature (see Figure 1). 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.

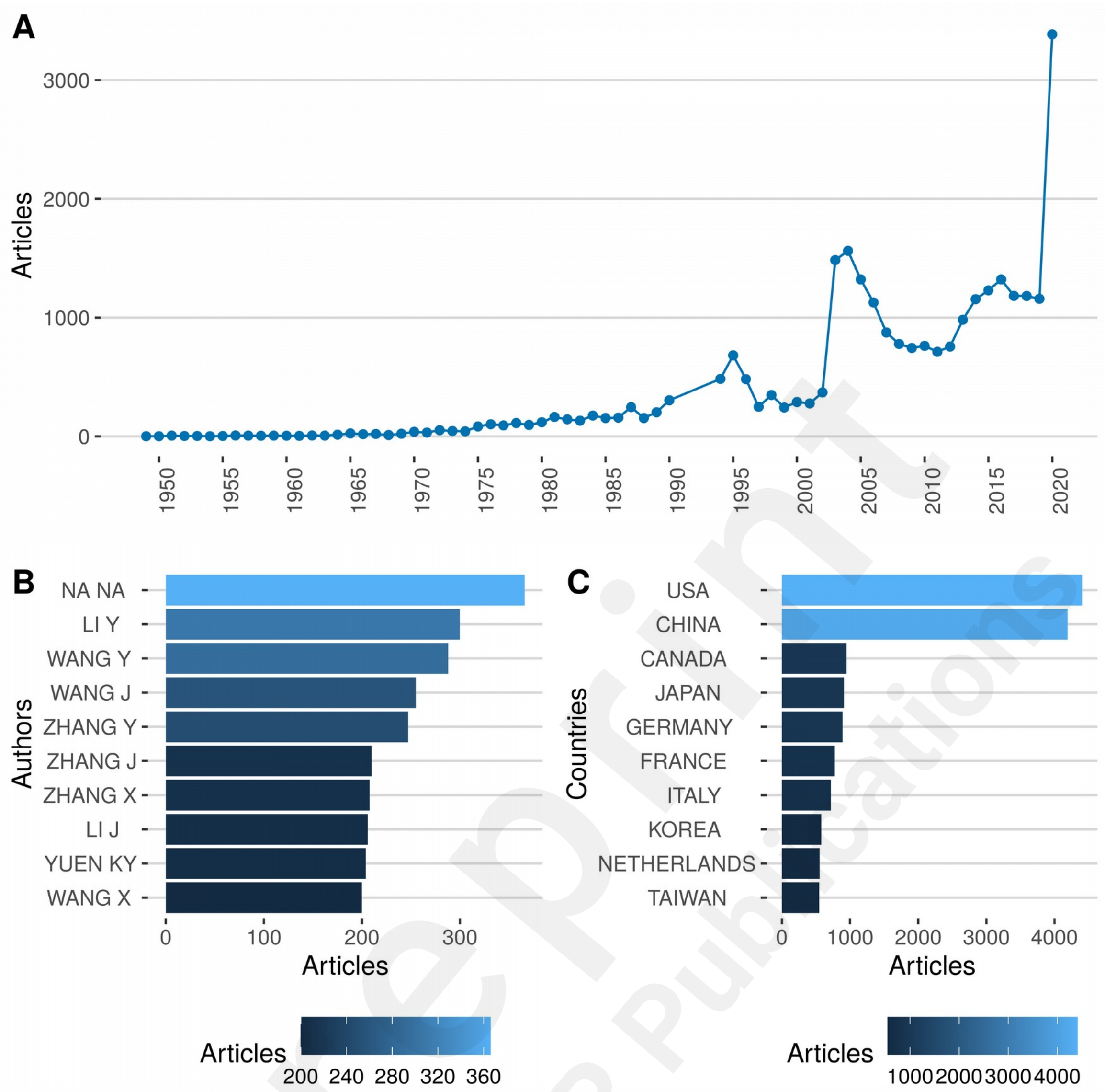

Figure 1. (A) Count of Articles, (B) Most Productive Authors and (C) Most Productive Countries.

In 2019, a little more than 1,000 articles on coronaviruses have been produced, and close to 3,400 articles just in the first 5 months of 2020. Those papers from the past two 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 productive authors provide an 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 the next figure, we first propose a very powerful visualization, called a Sankey diagram. It relates authors, keywords and sources on a connected map. It is a first way in creating groups of researchers. The results could be used by policymakers to identifiy areas of research in this particular topic.

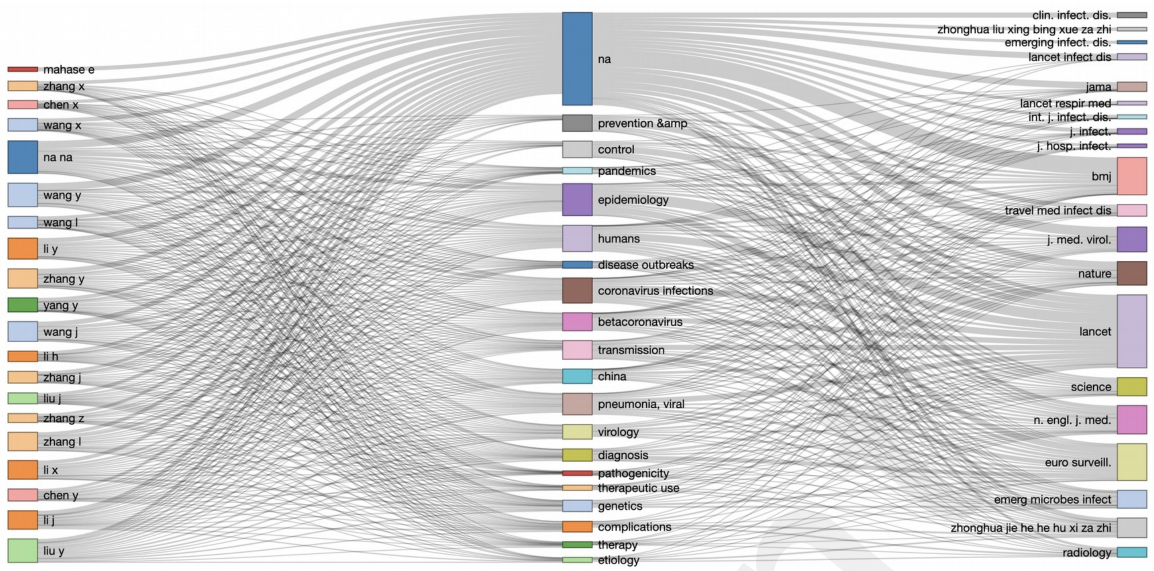

Figure 2. Sankey Diagram: Authors <> Keywords <> Sources.

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

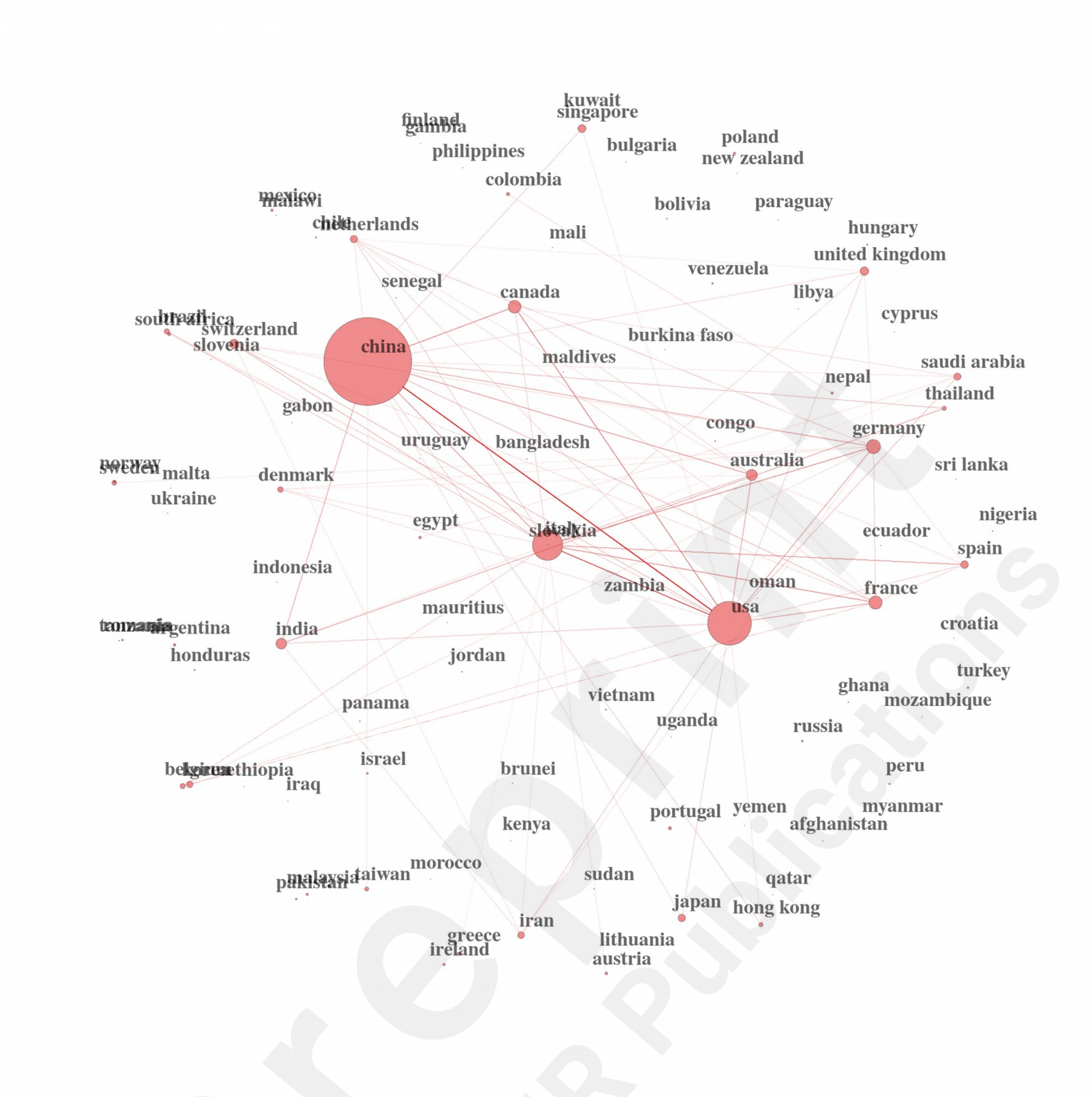

Figure 3. Country Collaboration Network.

Figure 4 is an illustration of author collaboration networks. As aforementioned, we remain at a very high level here. However, policymakers, public-health officers for instance, could use these techniques to find more granular networks either just within the 28,000 references or by crossing with other databases. We could even imagine crossing with unstructured data for some specific purposes [15].

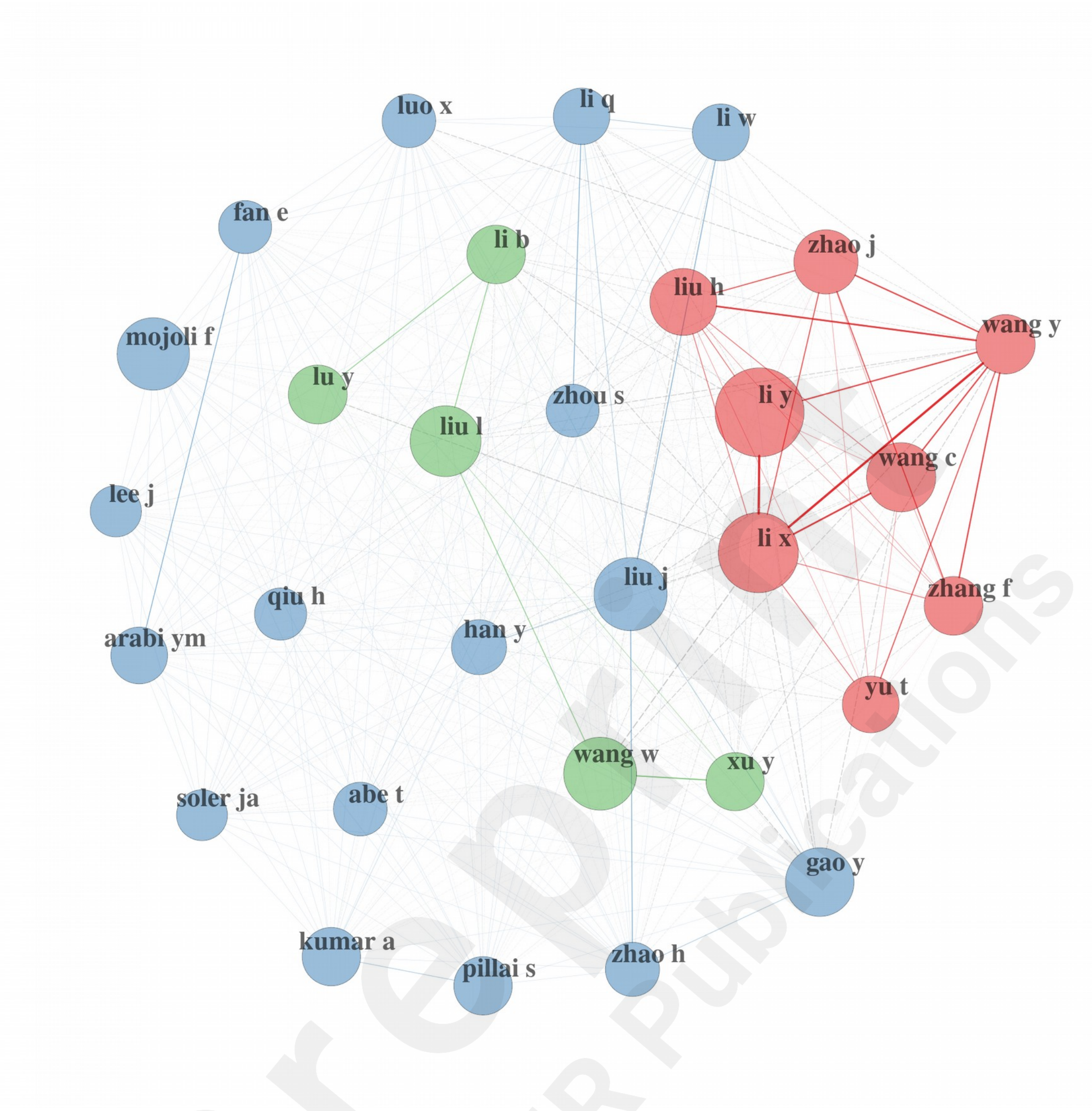

Figure 4. Author Collaboration Network.

Figure 5 is about finding clusters of topics. This technique can be applied to sub-samples of the 28,000 references in order to provide a more granular analysis.

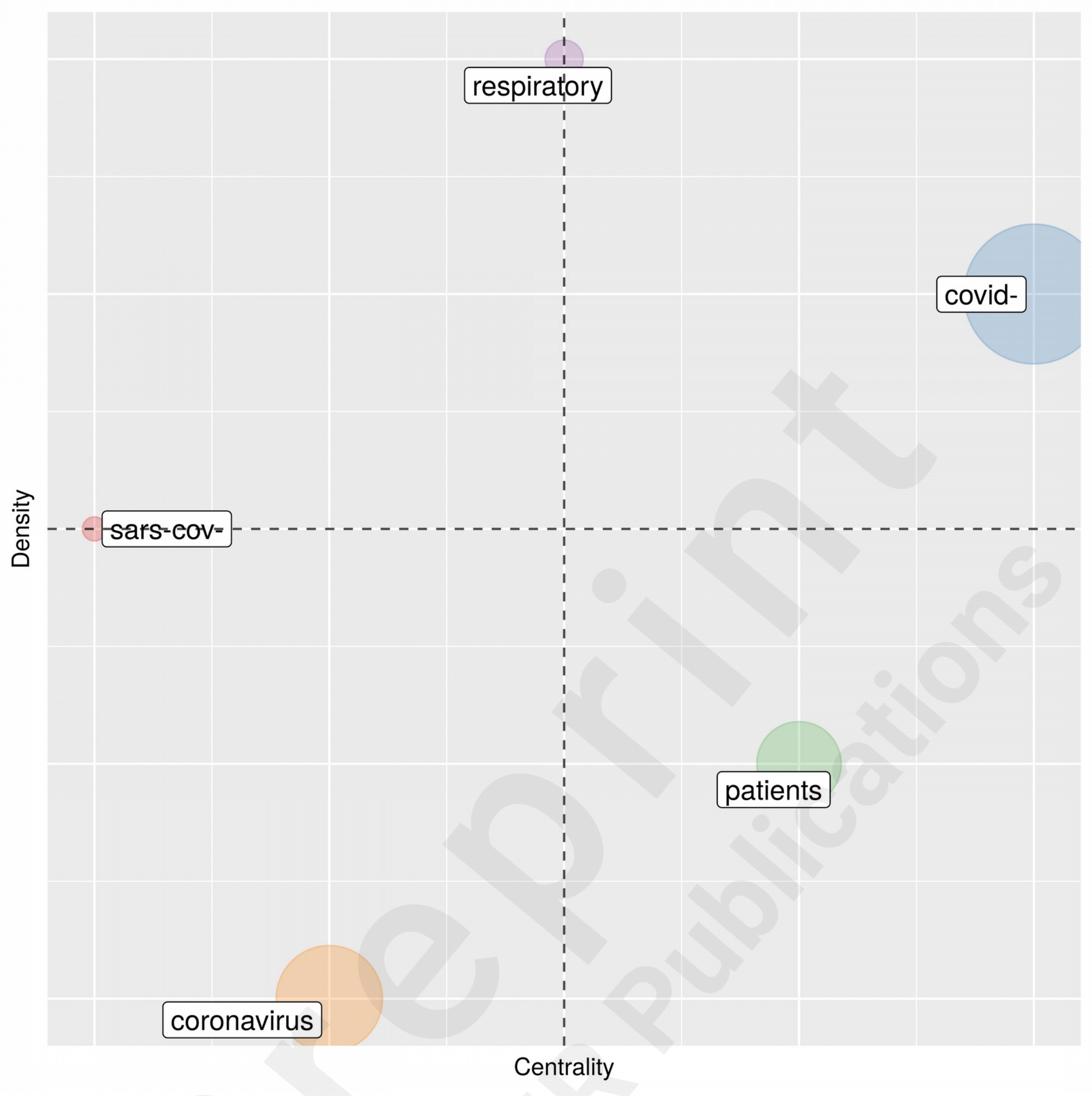

Figure 5. Thematic Maps.

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

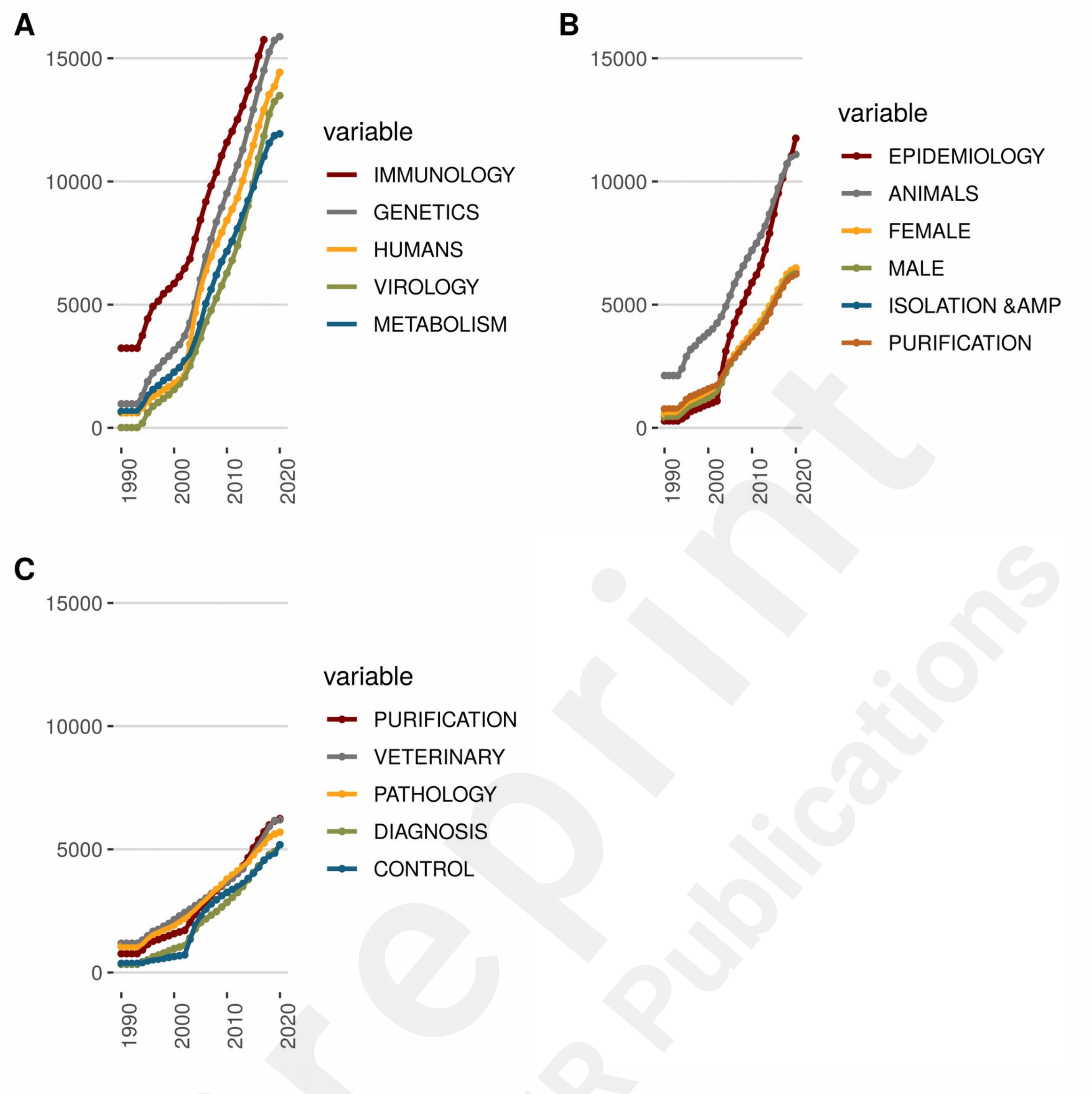

Figure 6. Author’s Keywords Usage Evolution Over Time.

As aforementioned, 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 dataset of 28,000 references, we have indeed access to a wealth of information. Combined with the state of the technological development we have access to in this day and age, there are multiple questions that can be answered. To go furher, we could imagine analyzing document co-citations, or reference Burst-detections [16].

Discussion and 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 capabilities as a way to encourage the further coevolution of new capabilities, echoing ideas put forward by Albert Hirschman more than 50 years ago [17]. The difference is that now, researchers, policy makers and business analysts have the capacity to analyze them in practice. It is capitalizing on the principles from the “wisdom of crowds” [18–21].

In this global pandemics, knowledge sharing and open data can have an impact on the solutions as well as the pace to discover the solutions. 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 policymaking [22].

Developing an easy access to data and data modelisation 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 asssumptions coming from theoretical foundations or from 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 [22–25].

This R package is updated on a daily basis when it comes to collecting the references and their metadata, and it will also be updated on a regular basis to propose different use cases and new functionalities. We will update the modelling contribution of the package. For instance, we will integrate some of the *bibliometrix* package's functions directly in our package in order to ease the scientometrist's workflow. We will also include some models for network analysis and Natural Language Processing (NLP) -based analyses.

Conclusion

In these trying times, we believe that working with reproducible research principles based on open science, disseminating scientific information, providing an 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, NGOs 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 scietnists can directly access data from four sources, facilitating mathematical modeling 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 especially designed to collect data and generate a dataframe compatible with the *bibliometrix* package [10]. Such datasets may facilitate the access to the right information. Moreover, the use of massive datasets crossed with powerful data analyses may foster multidisciplinary perspectives, raising new questions and providing new answers [26–28]. 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 dataset. 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.

Again, here, the package collects references - for the most part - interesting for the medical domain, and allowing multidisciplinary perspectives on this dataset. It could be interesting to get the perspectives from other disciplines, for instance mathematics, computer science, political science, economics, and environmental science. This is the result of the emergency situation in which humanity finds itself right now. We could also envisage later on 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.

Ackowledgements

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

The authors have declared no competing interest.

References

1. Wu T, Ge X, Yu G, Hu E. Open-source analytics tools for studying the COVID-19 coronavirus outbreak. medRxiv. 2020;2020.02.25.20027433. doi:10.1101/2020.02.25.20027433.
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet. 2020;395:470–3. doi:10.1016/S0140-6736(20)30185-9.
3. Warin T, Duc RL, Sanger W. Mapping Innovations in Artificial Intelligence through Patents: A Social Data Science Perspective. In: 2017 International Conference on Computational Science and Computational Intelligence (CSCI). 2017. pp. 252–7.
4. Systems Science and Engineering C for. COVID-19 Map. Johns Hopkins Coronavirus Resource Center. 2020. https://coronavirus.jhu.edu/map.html. Accessed 21 Apr 2020.
5. Warin T, Sanger W. Connectivity and closeness among international financial institutions: A network theory perspective. International Journal of Comparative Management. 2018;1:225–54. doi:10.1504/IJCM.2018.094479.
6. Scholar S. CORD-19 | Semantic Scholar. https://pages.semanticscholar.org/coronavirus-research. Accessed 21 Apr 2020.
7. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature. 2020;579:193–3. doi:10.1038/d41586-020-00694-1.
8. Organization WH. Global research on coronavirus disease (COVID-19). 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov. Accessed 24 Apr 2020.
9. Disease Control and Prevention C for. COVID-19 Research Articles Downloadable Database. 2020. https://www.cdc.gov/library/researchguides/2019novelcoronavirus/researcharticles.html. Accessed 24 Apr 2020.
10. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017;11:959–75. https://ideas.repec.org/a/eee/infome/v11y2017i4p959-975.html. Accessed 21 Apr 2020.
11. Camacho D, Panizo-LLedot À, Bello-Orgaz G, Gonzalez-Pardo A, Cambria E. The Four Dimensions of Social Network Analysis: An Overview of Research Methods, Applications, and Software Tools. 2020. http://arxiv.org/abs/2002.09485. Accessed 24 Apr 2020.
12. Tani M, Papaluca O, Sasso P. The System Thinking Perspective in the Open-Innovation

Research: A Systematic Review. Journal of Open Innovation: Technology, Market, and Complexity. 2018;4:38. doi:10.3390/joitmc4030038.
13. Pulsiri N, Vatananan-Thesenvitz R. Improving Systematic Literature Review with Automation and Bibliometrics. In. 2018. pp. 1–8.
14. Zhao L, Tang Z-y, Zou X. Mapping the Knowledge Domain of Smart-City Research: A Bibliometric and Scientometric Analysis. Sustainability. 2019;11:6648. doi:10.3390/su11236648.
15. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. Journal of Medical Internet Research. 2020;22:e19016. doi:10.2196/19016.
16. Kleinberg J. Bursty and Hierarchical Structure in Streams. Data Mining and Knowledge Discovery. 2003;7:373–97. doi:10.1023/A:1024940629314.
17. Hirschman A. The strategy of economic development. Yale University Press; 1958. https://books.google.ca/books?id=wls-AAAAYAAJ.
18. Gal M. The Power of the Crowd in the Sharing Economy. Law and Ethics of Human Rights. 2019;13:29–59.
19. Cai C, Gippel J, Zhu Y, Singh A. The power of crowds: Grand challenges in the Asia-Pacific region. Australian Journal of Management. 2019;44:551–70.
20. De C. Wang P, Soares V, De Souza J, Esteves M, Schots N, Duarte F. A crowd science framework to support the construction of a gold standard corpus for plagiarism detection. In. 2019. pp. 440–5.
21. Avasilcai S, Galateanu E. Co-creators in innovation ecosystems. Part II: Crowdsprings 'Crowd in action. In. 2018.
22. Payán D, Lewis L. Use of research evidence in state health policymaking: Menu labeling policy in California. Preventive Medicine Reports. 2019;16.
23. Wolffe T, Whaley P, Halsall C, Rooney A, Walker V. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environment International. 2019;130.
24. Giménez-bertomeu V, Domenech-lópez Y, Mateo-pérez M, 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. International Journal of Environmental Research and Public Health. 2019;16.
25. Villumsen S, Faxvaag A, Nøhr C. Development and progression in Danish eHealth policies: Towards evidence-based policy making. Studies in Health Technology and Informatics. 2019;264:1075–9.
26. Warin T, Leiter DB. An Empirical Study of Price Dispersion in Homogenous Goods Markets. Middlebury College, Department of Economics; 2007. https://ideas.repec.org/p/mdl/mdlpap/0710.html. Accessed 18 Apr 2020.
27. Warin T, Leiter D. Homogenous goods markets: An empirical study of price dispersion on the internet. International Journal of Economics and Business Research. 2012;4:514–29. doi:10.1504/IJEBR.2012.048776.
28. Marcellis-Warin ND, Sanger W, Warin T. A network analysis of financial conversations on Twitter. International Journal of Web Based Communities. 2017;13:281–310. doi:10.1504/IJWBC.2017.086588.