



The swift knowledge development path of COVID-19 research: the first 150 days

Mei Hsiu-Ching Ho¹ · John S. Liu¹

Received: 16 June 2020 / Accepted: 9 December 2020 / Published online: 1 February 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract

Scholars all over the world have produced a large body of COVID-19 literature in an exceptionally short period after the outbreak of this rapidly-spreading virus. An analysis of the literature accumulated in the first 150 days hints that the rapid knowledge accumulation in its early-stage development was expedited through a wide variety of journal platforms, a sense and pressure of national urgency, and inspiration from journal editorials.

Keywords Knowledge network · Covid-19 · Main path analysis · Literature analysis

Introduction

The sudden appearance of the novel coronavirus 2019-nCov, which spawned the COVID-19 pandemic, has drawn scientists all around the world into a war against the threat it causes to human life. The scale and scope of this fight is so unprecedented that scientific communities, mostly health, medicine, and virology related, created an extremely large body of literature in a remarkably short period of time. The first COVID-19 case was reported in late December 2019 in China (Zhu et al. 2020) and by May 31, 2020 (just 5 months later), the Web of Science database had already compiled 4906¹ COVID-19 relevant publications into its databases, while the Scopus database had accumulated 13,197² articles. These sizes may seem to be quite impressive, yet they are dwarfed by LitCovid (Chen et al. 2020), which is a literature hub initiated just for the purpose of tracking up-to-date articles in PubMed that are relevant to COVID-19. Up to May 31, 2020, the hub

¹ This includes all documents of the types Early Access, Article, Editorial, Letter, Review, News Items, Proceeding Papers, etc. that contain COVID-19 and related terminologies in their title, abstract, and keywords.

² This includes all documents of the types Article, Letter, Editorial, Review, Note, Short Survey, Conference Paper, etc. that contain COVID-19 and related terminologies in their title, abstract, and keywords.

✉ Mei Hsiu-Ching Ho
mei.ho@mail.ntust.edu.tw

John S. Liu
johnliu@mail.ntust.edu.tw

¹ National Taiwan University of Science and Technology, No.43, Keelung Rd., Sec.4, Da'an Dist., Taipei 10607, Taiwan, ROC

has curated 17,892 relevant publications, equal to a production of almost 120 articles a day throughout the scientific community.

This sudden, grand emergence of the COVID-19 literature is a phenomenon deserving close observation and analysis. As of the date that this present paper was finalized, COVID-19 has still been affecting many countries around the world significantly, and the literature's unrelenting growth appears to have no end in sight. At this early development stage of COVID-19 research, several interesting questions arise. From the view of bibliometrics, what critical academic platforms accelerate knowledge sharing during a fast growing period of research output? What are the key issues on the knowledge developmental path and from where do they come? Is there any driver leading the development of the knowledge path at the early stage of COVID-19?

This report conducts bibliometric analysis upon a set of COVID-19 literature up to May 31, 2020. The data come from the Web of Science (WOS) service, which is selected based on two rationales. First, we rely on citation information to trace knowledge flows. Differing from the papers and documents published in LitCovid, WoS offers citation information. Second, it is better to focus on a set of papers that are selected by a qualified and strict scientific review process at this early stage of analysis. All papers in WoS belong to journals indexed as SCI, SCIE, or SSCI, which shows that the quality of papers are trustworthy as observations for further analysis. The data provided by WOS fit both criteria. In other words, we analyze a snapshot of the COVID-19 literature at an early development stage and from a selective sample. Although the literature assembled by WoS is less than that listed in LitCovid and Scopus, we conclude the analyses with interesting insights on the characteristics of this research strand.

General statistics of the COVID-19 literature

We obtain the dataset by employing keyword searches, such as "COVID-19", "2019 novel coronavirus", "sars-cov-2", "2019n-cov", etc.,³ in a document's title, abstract, and keyword list and for our analysis select those documents of the type Paper, Review, Early Access, Editorial, and Letter (as defined by WOS). After setting the period through May 31, 2020, we retrieve 4559 papers from more than 100 countries around the world. This section uses publication statistics to show the important knowledge sources. The third section builds up a citation network based on the retrieved dataset to explore the main knowledge diffusion paths of COVID-19 studies.

In the past 150 days, COVID-19 relevant articles have been published in 1159 academic journals, amounting to an average of 3.93 articles per journal. In fact, 25% (288) of these journals have published 4 or more articles. The scope of the journals covers a wide range of disciplines, including General/Internal Medicine, Public Health, Biology, Infection Disease, Pharmacology, Microbiology, Virology, Oncology, Immunology, Biochemistry, and Psychology, to name only a few.

³ The Keyword query used in WoS is listed below. TS: ("sars-cov-2" or "2019 novel corona virus" or "2019n-cov" Or "2019-cov" Or "2019-new coronavirus" Or "2019-novel coronavirus" Or "2019 ncov" Or "2019 ncov" Or "wuhan-2019-ncov" Or "wuhan virus" Or "2019 novel coronavirus" Or "2019 coronavirus disease" Or "2019 coronavirus disease" Or "2019 novel coronavirus disease*" Or "2019 novel coronavirus infection" Or "2019 novel coronavirus pneumonia" Or "wuhan pneumonia" Or "2019-coronavirus disease*" Or "2019-ncov disease*" Or "2019-ncov pneumonia" Or "2019-ncov" Or "covid-19").

Table 1 Top 20 journals by *g*-index

Rank	Name	<i>g</i> -index	<i>h</i> -index	Total papers	Total citations
1	<i>Lancet</i>	62	20	117	3906
2	<i>New England Journal of Medicine</i>	51	18	52	2643
3	<i>Nature</i>	25	5	36	655
4	<i>Journal of Medical Virology</i>	24	16	114	759
5	<i>Journal of the American Medical Association</i>	22	12	22	1255
6	<i>Lancet Respiratory Medicine</i>	20	9	34	439
7	<i>Lancet Infectious Diseases</i>	18	11	28	346
8	<i>Eurosurveillance</i>	18	8	55	334
9	<i>Science</i>	16	6	23	269
10	<i>International Journal of Infectious Diseases</i>	15	5	29	244
11	<i>Journal of Clinical Medicine</i>	14	9	29	209
12	<i>Intensive Care Medicine</i>	14	8	32	216
13	<i>Lancet Psychiatry</i>	13	7	15	175
14	<i>Emerging Microbes and Infections</i>	13	8	23	199
15	<i>British Medical Journal</i>	12	5	145	215
16	<i>Radiology</i>	11	8	11	290
17	<i>Journal of Korean Medical Science</i>	11	7	23	134
18	<i>European Radiology</i>	10	4	26	104
19	<i>Nature Medicine</i>	10	6	34	120
20	<i>Canadian Journal of Anesthesia</i>	10	5	38	129

Table 1 lists the top 20 journals in the order of their *g*-index⁴ and shows individual journals’ knowledge contribution. *Lancet*, *New England Journal of Medicine*, *Nature*, *Journal of Medical Virology*, and *Journal of the American Medical Association* are the top 5 journals with a *g*-index of over 20. *Lancet* has the highest *g*-index (62) as well as the highest citation number (3906) among all the journals. *New England Journal of Medicine*, although publishing far less than half the amount of papers as *Lancet*, has 2643 citations and the 2nd highest *g*-index at 51. Similarly, 22 papers in *Journal of the American Medical Association* obtain more than 1,200 citations, which is in third place among all journals. Compared to the top 5 journals, *British Medical Journal* has published the largest number of studies (145), but its citation number and *g*-index are much less than those in the top 5. Among the top 5 journals, some contribute toward publishing various scientific discussions on COVID-19 (e.g., *Lancet* and *Journal of Medical Virology*), and some provide critical scientific results for further extensions (e.g., *Lancet*, *New England Journal of Medicine*, and *Journal of the American Medical Association*). Academic journals always take on the role as a platform for scholars to absorb knowledge, share new findings, and push scientific knowledge accumulation forward. Particularly, at this early development stage of COVID-19 studies, the research findings or editorial ideas in these top journals in the medical field listed in Table 1 seem to have attracted much attention for scientific research.

⁴ A journal’s *g*-index measures its contribution based on received citations (Egghe 2006). It is a variation of the *h*-index.

Table 2 Top 20 countries according to three publication indicators

Rank	Quantity		Quality		Country	<i>m</i> -index
	Country	No. of papers	Country	No. of citations		
1	China	995	China	11,551	China	0.9862
2	U.S.	848	U.S.	2,350	U.S.	0.5141
3	Italy	440	England*	816	Italy	0.2152
4	England*	356	Germany	668	England*	0.1585
5	Canada	127	Italy	563	Singapore	0.1194
6	India	121	Singapore	361	Canada	0.1173
7	France	117	Canada	293	Germany	0.0902
8	Australia	113	France	197	Taiwan	0.0815
9	Singapore	100	Switzerland	185	India	0.0735
10	Germany	91	South Korea	171	France	0.0653
11	Iran	89	Japan	156	Iran	0.0550
12	Switzerland	81	Taiwan	136	Vietnam	0.0539
13	Turkey	60	Australia	132	Australia	0.0477
14	South Korea	54	Scotland	98	Brazil	0.0465
15	Brazil	51	India	95	Netherlands	0.0447
16	Japan	49	Netherlands	91	Switzerland	0.0404
17	Spain	46	Sweden	91	Turkey	0.0403
18	Netherlands	37	Belgium	77	Japan	0.0382
19	Thailand	33	Vietnam	75	Norway	0.0307
20	Taiwan	32	Thailand	63	Spain	0.0273

*We do not combine all the regions in the United Kingdom of Great Britain and Northern Ireland (England, Scotland, and Wales), because we follow the address information offered by WOS

Based on the address of corresponding authors, we trace the locations of knowledge sources. As shown in Table 2, this paper adopts both quantity and quality views to highlight the contributions from different countries. Quantitatively, the top 5 countries in terms of publication are China, United States, Italy, England, and Canada. China and the U.S. are in the first tier with publications of around 800–1000 papers, while Italy and England follow up with around 300–500. In addition, to have a proper understanding of the contributions of all scientists from the global world, we further extract more than 20,000 addresses in 104 countries/regions provided by all papers in this dataset. After excluding unknown information (2.8%), we use a pie chart for the geographical distribution (Fig. 1). The result is consistent with the statistics of Table 2. These countries actively and quickly responded to the COVID-19 threat in academic research. Canada, France, Australia, Singapore, and India are others in the table with at least 100 publications.

Citation count and *m*-index⁵ qualitatively show us the importance of the studies in different countries (Table 2). China has the highest citation number (11,551) with a *m*-index

⁵ The *m*-index is originally designed to measure the influence of a scholar and is defined as the percentage of all knowledge dissemination paths that contain a scholar's works within a target scientific field (Liu et al. 2012). Here, the concept is extended from a measure for scholars to an index for countries, thus indicating the level of association of a country's publications with the mainstream knowledge within a target scientific field. The *m*-index ranges from 0 to 1.

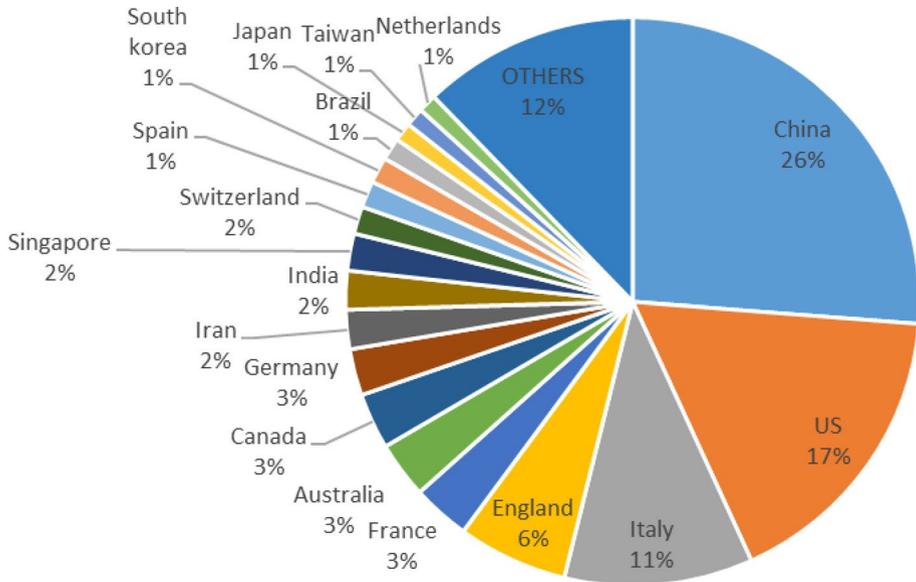


Fig. 1 Geographical distribution of the addresses of all authors

of 0.98. In other words, most COVID-19 studies at the current stage cite studies with the corresponding author from China, likely because these early studies by Chinese scientists offer some reference information for follow-up studies. The U.S. is another important knowledge source, presenting studies (848 papers) with more than 2000 citations. The *m*-index for the U.S. is higher than 0.5, implying that U.S.-located studies also play a critical role in the overall development of the COVID-19 literature. Italy is in third place with *m*-index at 0.21. Aside from countries with high publication numbers, Germany is worth mentioning as it has published 91 papers (in 10th place), yet its high citation number (668) ranks it 4th in that category. One early report on cases there (Rothe et al. 2020) significantly increased Germany’s citation count by 190. We also find that several Asia countries (e.g., Japan, South Korea, and Taiwan) illustrate their importance in this wave of research. The statistics can simply show us the high involvement of scientists from certain countries during the expedited period of publication in the early stage of the topic’s development.

The countries at the top of Table 2 represent those that suffered earlier and the most under COVID-19. The emergence of this unpredictable virus in a country has made it think of new ways to block or reduce any harmful threats to its residents. As scientists, the best way they contribute to fighting COVID-19 is to understand the virus as quickly as possible. China, the first country in December 2019 to confront such challenges, responded much earlier than other countries. Severe health situations appeared soon after in South Korea, the U.S., Italy, France, Germany, and UK, forcing them to find ways to fight the rapid growth of infected cases. These countries exhibit high scientific contributions either for their quantitative publication number or qualitative citations. Like a facilitator, the uncertain and unpredictable outbreak of COVID-19 has quickly spurred scholars and scientists to make concerted efforts at finding solutions through various research and experiments.

The rapid growth of publications in the past 150 days shows how these efforts have driven scholarly research toward a boundless world of knowledge.

Knowledge path and the drivers from journal editorials

In order to observe the early development of COVID-19 studies, we apply main path analysis (MPA) to explore the knowledge flow during this period. MPA has been widely used to explore the knowledge diffusion trajectory of a certain technology or theory (Verspagen 2007; Liu and Lu 2012; Liu et al. 2016). This study adopts the SPLC algorithm (Hummon and Doreian 1989) to determine the significance of a citation link when applying MPA. The choice is based on the suggestion that SPLC “fits the knowledge diffusion model better than the other traversal weights” (Liu et al. 2016). After the significance of each citation link is determined, the key-route approach (Liu and Lu 2012) helps search for the main paths. This approach not only guarantees the inclusion of the most significant citation links in the main paths, but also allows for controlling the main path details through the inclusion parameter, which specifies the number of top citation links to be included. This study sets the inclusion parameter to 10 as the knowledge diffusion paths are best demonstrated under this parameter number. In other words, the main paths either reduce to a monotonous line or lose their simplicity in structure when decreasing or increasing this number. We calculate all the above-mentioned MPA procedures through in-house software.

When applying MPA to identify the development of knowledge in a certain technology field, researchers always obtain a certain amount of papers or patent documents, and the data are accumulated over a long time; e.g., years. Though we only collect publication data of COVID-19 studies for the first 150 days, the size of the literature is big enough to employ MPA as a proper tool for exploring the topic in its early stage. In addition, the analysis provides an early observation of the fast changing path in order to explore the implications for both scholars of COVID-19 and bibliometrics studies. Thus, we construct a citation network from 4,559 publications, wherein the largest component consists of 3510 papers. Based on MPA, we identify the key topics and show the critical role that the editorial literature plays for knowledge diffusion in the past five months.

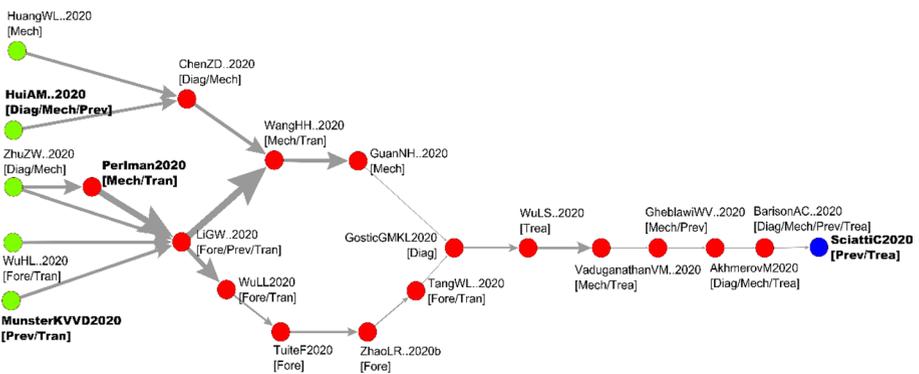


Fig. 2 Main path of COVID-19 studies (Key-route 10)

Figure 2 illustrates that the main path at key-route 10 includes 21 papers. Each paper is labeled with the 1st author's name followed by the names of the 2nd, 3rd authors, etc. The thickness of a line and arrow on the main path is based on the SPLC value counted by citation information. Inside the square brackets are the research topics as classified by LitCovid.⁶ Mechanism and Transmission are the two top topics on the main paths.⁷ Eleven of the 21 papers cover issues focusing on Mechanism ([Mech] in Fig. 2), six on Transmission ([Tran]), five on Diagnosis ([Diag]), and four each on Treatment ([Trea]), Prevention ([Prev]), and Forecasting ([Fore]). Many of the very early studies target Mechanism, Diagnosis, and Transmission of 2019n-CoV, whereas the latter studies write about Forecasting and Treatment.⁸

Figure 2 presents the critical issues discussed in the early stage. One of the five knowledge sources, labeled ZhuZW.0.2020 (Zhu et al. 2020), particularly focuses on virus mechanism and explains the genomic characteristics and structure of 2019n-CoV. Another source, labeled HuangWL.0.2020 (Huang et al. 2020), presents the clinical features, possible symptoms, diagnosis methods, and some treatment outcomes. In addition, studies on the early transmission in Wuhan, China as well as the potential global outbreak and impact before March 2020 help scientists to learn from the experiences and management of this particular virus transmission (Li et al. 2020; Wu et al. 2020; Munster et al. 2020; Gostic et al. 2020). Because of these early observations, the unpredictable risk from asymptomatic carrier transmission has become critical information for preventing virus outbreak. The transmission issue then extends to some discussion on the role the public health system has. Papers close to the end of the main path investigate virus receptor/regulator and the COVID-19 and heart (cardiovascular disease) nexus (Akhmerov and Marbán 2020; Barison et al. 2020). We note that many of the more recent studies have studied patients with severe cardiovascular disease, seeking treatments that can help fight the virus mechanism inside the human body.

Editorials interestingly seem to play an essential communication role in the early germination stage of COVID-19 research. Four out of the 21 papers on the main paths are editorials (shown in bold font in Fig. 2). While appearing in different parts of the main paths, their function at facilitating knowledge diffusion is the same. Among the early studies, Perlman (2020), an editorial in *New England Journal of Medicine*, recapitulates several earlier findings in China (Zhu et al. 2020; Huang et al. 2020; Zhou et al. 2020) and advocates the importance of virus identification for further virus detection. From the virological view, it also mentions that the genomic sequence helps to further develop vaccines and antiviral therapies. A very strong link from Perlman2020 to LiGW.0.2020 (Li et al. 2020) highlights the essential connecting role of the editorials to later studies. Another two editorials, HuiAM.0.2020 (Hui et al. 2020) and MunsterKVVD2020 (Munster et al. 2020), position themselves as the knowledge source on the main paths. The former is penned by a global team led by a scholar in Hong Kong, while the latter is authored by a team in the U.S. and the Netherlands. They both point out the possible global threat of virus transmission,

⁶ LitCovid classifies papers into 8 research topics: General, Mechanism, Transmission, Diagnosis, Treatment, Prevention, Case Report, and Forecasting. Each paper is categorized into one or more topics.

⁷ This differs from the assessment by the number of publications. In LitCovid, up to May 31, 2020, the numbers of publications for each category are Prevention (7,333), Treatment (3,791), Diagnosis (2,608), Mechanism (1,947), General (1042), Transmission (753), Forecasting (266), and Case Report (115).

⁸ We classify each paper into one or more than one classification by LitCovid, but the content of each paper may cover different coronavirus issues. Therefore, the sum of the number of all classifications is more than the number of papers (i.e., 21) on the main path.

motivating more observations from Chinese studies. SciattiC2020 (Sciatti and Ceconi 2020), appearing at the end of the main paths, is a more recent editorial, specifically commenting on a previous work by Barison et al. (2020) and echoing the authors' view over rethinking in-hospital management of the human-virus cohabitation situation. The inevitable position of editorials on the main paths demonstrates that they communicate/update/summarize the crisis and provide insights for scholars to conduct studies critical for human beings, especially in the early stage of a crisis such as this COVID-19 pandemic.

Conclusion

The rapid growth of COVID-19 publications implies that scholars have made tremendous efforts to find solution through different resources during its outbreak period. Various studies on the characteristics of COVID-19 have paved the way for speeding up vaccine developments and pharmaceutical solutions to fight the virus. Research about virus mechanisms and transmission found on the main path of COVID-19 studies scientifically indicates the importance of these topics in the past 150 days.

We summarize three findings to shed light on the tremendous growth of publications in the past 150 days. First, under the context of this topic's urgency, top academic journals in medical field provide a wide variety of platforms for scholars to present and learn scientific results. Their fast turnaround time has sped up knowledge sharing. Second, virus mechanism and its transmission control are the key issues in the early stage. The countries affected in the initial period (e.g., China, South Korea) or those with serious outbreaks (e.g., U.S., Italy, Germany, France, England, and Singapore) have all been in a race against time and produced a relatively large number of COVID-19 publications, reflecting how the pressure to fight the virus accelerated scientific research. Third, editorials have a special position on the main paths of COVID-19 studies, summarizing the latest observations and inspiring further investigations. Although these editorials may be short, they are a bridge to communicate the latest information and research direction. The editorial messages insightfully point toward a path for scientists to further work out trustworthy research evidence.

The data we use in this analysis cover only the first five months of the outbreak, yet the development of COVID-19 research is continuously evolving. The analysis from a selective sample in WoS provides part of the development in the early stage, and a further study can extend into covering a wider range of time and database. COVID-19 relevant publications have grown so fast that we should keep our eyes on their development to know the whole story. Furthermore, the statistics we use to show research participation from different nations are based on the corresponding author's address, but not from the content or sample of each paper. We do have some limitations when using the results to represent the scientific contributions of some certain nations. The progress of all countries' efforts to overcome the negative effects from coronavirus will help form the holistic scientific path of COVID-19. These forthcoming studies also determine what certain research topics will eventually dominate the main paths.

Acknowledgements We thank two anonymous reviewers for their constructive comments which have greatly improved the accuracy and readability of this article. This work is partially supported by Taiwan's Ministry of Science and Technology grants MOST 109-2410-H-011-020-MY2, 109-2410-H-011-018, 108-2410-H-011-025, and 108-2410-H-011-021.

References

- Akhmerov, A., & Marbán, E. (2020). COVID-19 and the Heart. *Circulation Research*, 126(10), 1443–1455.
- Barison, A., Aimo, A., Castiglione, V., Arzilli, C., Lupón, J., Codina, P., et al. (2020). Cardiovascular disease and COVID-19: Les liaisons dangereuses. *European Journal of Preventive Cardiology*. <https://doi.org/10.1177/2047487320924501>.
- Chen, Q., Allot, A., & Lu, Z. (2020). Keep up with the latest coronavirus research. *Nature*, 579(7798), 193–193.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
- Gostic, K., Gomez, A. C., Mummah, R. O., Kucharski, A. J., & Lloyd-Smith, J. O. (2020). Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *Elife*, 9, e55570.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223), 497–506.
- Hui, D., Azhar, E., Madani, T., Ntoumi, F., Kock, R., Dar, O., et al. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health: The latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases*, 91, 264–266.
- Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39–63.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus—infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207.
- Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, 63(3), 528–542.
- Liu, J. S., Lu, L. Y., & Lu, W.-M. (2016). Research fronts in data envelopment analysis. *Omega*, 58, 33–45.
- Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D., & de Wit, E. (2020). A novel coronavirus emerging in China—key questions for impact assessment. *New England Journal of Medicine*, 382(8), 692–694.
- Perlman, S. (2020). Another decade, another coronavirus. *New England Journal of Medicine*, 382(8), 760–762.
- Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., et al. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*, 382(10), 970–971.
- Sciatti, E., & Ceconi, C. (2020). Les liaisons dangereuses and the danger of deductions: The interplay between cardiovascular disease and COVID-19. *European Journal of Preventive Cardiology*. <https://doi.org/10.1177/2047487320925622>.
- Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(01), 93–115.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet*, 395(10225), 689–697.
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *BioRxiv*. <https://doi.org/10.1101/2020.01.22.914952>.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382, 727–733.