Using the strength coefficient(SC) to evaluate the most number of publications in research affiliations involving COVID-19 till April 14, 2020: A bibliometric analysis

Wei-Chih Kan

Chi Mei Medical Center

Yu-Tsen Yeh

London Metropolitan University

Jui-Chung John Lin

Bayer HealthCare Pharmaceuticals US Innovation Center

Tsair-Wei Chien (✉ smile@mail.chimei.org.tw)

Chimei Medical center <https://orcid.org/0000-0003-1329-0679>

Research article

Keywords: research affiliation, strength coefficient, social network analysis, COVID-19, medical subject headings

Posted Date: April 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-24741/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: When the COVID-19 outbreak spreads to the world, many articles related to it have been published in academics. Since the largest quantity of confirmed cases was reported in China till April 14, 2020, whether the number of Chinese articles of research associated with the COVID-19 topped globally is required to be examined. Thus, an objective measure determining the dominant role in a group should be defined. This study aims to propose an index (strength coefficient, SC) to evaluate the most influential research affiliations in publications of COVID-19.

Methods: We simulated data to verify the separation index that can be viable in use for determining the dominant one that has the absolute advantage in a group. We selected 4,369 articles as of April 14, 2020, with abstracts from the Pubmed Central (PMC) based on the keywords COVID-19 or 2019-nCoV. An author-weighted scheme (AWS) was applied to quantify coauthor credits in the article byline. Social network analysis incorporated with SC(from 0 to 1.0 and cutoff=0/70) was applied to display the influential (1) article types, (2)countries, (3)medical subject headings(MeSH terms), and (4) research affiliations. Visual dashboards were created for displaying the results on Google Maps.

Results: We observed that the top one(SC) in each topic consists of (1) journal article(0.81), (2) China(0.61),(3) Acad Radol, (4) betacoronavirus (0.66), and (5) Hazhong University of Science and Technology(0.77) in article types, countries, journals, MeSH terms, and research affiliations, respectively.

Conclusion: We applied the SC to identify the strength of the top one over the next two. The SC was useful and viable in verifying the dominant role in a group. The implementation and application are worthy of further studies in the future.

Background

An index that can verify the dominant role in a group was required to obtain[1], such as the strength of the leading company in an industry. The Herfindahl index(known as HHI[2]) has been proposed to measure the strength of competition in industry[3]. However, The HHI defined as the sum of the squares of the market shares of the firms(i.e., $H = \sum_{i=1}^n s_i^2$, where the number in the industry; n as that sometimes limited to the 50 largest firms) has some disadvantages. For instance, the market shares are constructed by many fractions. The HHI is fluctuated without a common criterion (e.g., $H = 0.33 = 3 \times (0.33 \times 0.33)$ and $H = 0.25 = 4 \times (0.25 \times 0.25)$ for three and four equal- size companies, respectively).

Similarly, the Gini coefficient[4] was suggested to measure the author's research domain based on the top five medical subject headings(MeSH terms)[5, 6]. The major problem of these HHI and Gini is that all elements(or fractions) are equally considered(or weighted) in a formula instead of focusing on the top one with the dominant role (i.e., the 100% monopoly or say the strength of the competition). The scenario is rather similar to what we investigate one questionnaire forming a unidimensional construct using Eigenvalues to definition[7, 8]. The dimension coefficient(DC) is defined by the top three eigenvalues(λ_i)

with the formula $(= [(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)]/[1+[(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)])$ and criterion (≥ 0.70) , where λ_i stands for the number of Eigenvalues in descending order [9, 10], The ratio of DC is renamed as separation index (SI) [1] [or strength coefficient (SC) in this study] for representing the dominant extent to which the role plays in a set of entities. Whether the SC can also be applied to other fields (e.g., dominant research affiliations) is required for verification.

When the COVID – 19 spreads to the world, numerous articles related to COVID-19 have been published. As of April 14, 2020, more than 4,369 documents were released in search of keyword “COVID-19 or 2019-nCov” in Pubmed Central (PMC). We are motivated to investigate the most influential (1) article types, (2) countries, (3) MeSH terms, and (4) research affiliations and examine which ones have a high SC (≥ 0.70) enough to be a dominant role in a group.

This study aims to apply the SC to evaluate the most influential research affiliations in publications of COVID-19 and to know the influent (1) article types, (2) countries, and (3) MeSH terms in publications.

Methods

2.1. Data Sources

We programmed Microsoft Excel visual basic for applications' (VBA) modules to extract abstracts containing MeSH terms, journals, coauthors' affiliations, and oration countries/regions as of April 14, 2020, published in PMC since January 1, 2020, in search of the keyword COVID-19 or 2019-nCov. (e.g., ((COVID-19) or (2019-nCov)) AND ("2020/01/01"[Date - Publication])). A total of 4,369 eligible abstracts were obtained from PMC.

Ethical approval was not necessary for this study because all the data were extracted from the website publicly available in PMC.

2.2 The strength coefficient

We define the strength coefficient as the formula $(= [(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)]/[1+[(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)])$ [7–10], where λ_i represents the values (e.g., count or frequency) from top one to three in descending order. The SC from 0 to 1.0 is the strength of concentration to the first one; the higher means, the stronger extent to which the monopoly is forward. The cutoff point is set at 0.7 [7–8], similar to the Cronbach Alpha in classical test theory. For instance, the first institute has publications (= 846), followed by the second (= 117) and the third (= 62). The SC equals 0.79 $(= [(846/117)/(117/62)]/[1+[(846/117)/(117/62)])$.

2.3 Social network analysis

Social network analysis (SNA) [11] was applied to explore the pattern of entities in a system using the software Pajek [in Koeln; PajekMan in Osoje (Ossiach, Austria)] [12]. In keeping with the Pajek guidelines, we defined an author research institute (or paper keyword) as a node (or an actor) that is connected to

other nodes through the edge (or the relation). The number of connections usually defines the weight between two nodes.

Centrality is a vital index for analyzing a network. Any individual or keyword in the center of a social network will determine its influence on the network and its speed at gaining information [13, 14]. In this study, we used the centrality degree as the number of weights connected to other nodes.

2/4 The Author-Weighted Scheme

The author-weighted scheme(AWS)[15] is applied to weight the author research institute:

$$W_j = \frac{\exp(\gamma_j)}{\sum_{j=0}^{m-1} \exp(\gamma_j)} = \frac{2.72^{\gamma_j}}{\sum_{j=0}^{m-1} 2.72^{\gamma_j}}$$

Where, considering a paper of $m + 1$ authors with the last being the corresponding author, W_j denotes the weight for an author on the order j in the article byline. The power, γ_j , is an integer number from $m-1$ to 0 in descending order. The sum of author weights in a byline is shown below:

$$= \sum_{k=0}^{m-1} \frac{\exp(\gamma_j)}{\sum_{j=0}^{m-1} \exp(\gamma_j)}$$

The sum of authorships for each paper equals 1.0. which is a basic concept ensuring that all papers have an equal weight irrespective of the number of coauthors [16]. As such, more importance is given to the first ($\exp[m]$, primary) and the last ($\exp[m-1]$, corresponding or supervisory) authors, whereas it is assumed that the others (the middle authors) have made smaller contributions [17, 18]. The smallest portion ($\exp(0) = 1$) is assigned to the last second author with the odds = 1 as the basic reference [19, 20].

2.5 Pattern of Authors' institute and countries o MeSH terms

We selected the topic of COVID-19 as the target article. A total of 4,369 articles (see Additional File 1) were collected. Five types of social networks were constructed, including (1) article types, (2)countries,(3) journals associated with counties, (4) MeSH terms, and (4) research affiliations. The top three were highlighted in each network to compute the SC for each dominant role.

2.6 Creating Dashboards on Google Maps

Author-made online modules were used to present the dominant role of each network. We created pages of HTML used for Google Maps. All the relevant information on the entities (i.e., research institutes, MeSH

terms, or countries) can be linked to dashboards on Google Maps.

Results

The top ones(with SC) in each network are shown in figures from 1 to 5. It can be seen that (1) journal article(0.81), (2) China(0.61),(3) Journal of Acad Radol, (4) betacoronavirus (0.66), and (5) Hazhong Univeristy of Science and Technology(0.77) are highlighted in networks of article type, affiliated country, journal, MeSH term, and author research institutes, respectively.

All of those SCs, but the article type, were computed by the weights in their networks. The SC for article types are based on the on article counts weighted by the AWS(Table 1), different from the weights connected in the network in Fig. 1. The one of Huazhong University of Science and Technology(Hubei) ranks at the first place and gains $SC = 0.60(= [(12.57/7.74)/(7.74/7.08)] / (1+[(12.57/7.74)/(7.74/7.08)])$.

Table 1
The most number of articles in research institutes

Institute	Article
Huazhong University of Science and Technology(Hubei)	12.57
Fudan University(Shanghai)	7.74
Hokkaido University(Japan)	7.08
The University of Hong Kong(Hong Kong)	5.64
Zhongnan Hospital of Wuhan University(Hubei)	4.14
Chinese University of Hong Kong(Hong Kong)	4.12
Chinese Academy of Medical Sciences and Peking Union Medical College(Beijing)	4.09
Wuhan Institute of Virology(Hubei)	3.85
Capital Medical University(Beijing)	3.79
Zhejiang University School of Medicine(Zhejiang)	3.27
Note. articles are weighted by the AWS mentioned in Methods.	

It is worth noting that the journal of Acad Radol is closely associated with China and Adv Radiat Oncol related to the US. The method we combined counties and journals in a network. The highest weights for entities have been outlined in its network for easily understanding the association with others. Readers are invited to see the links in references[21–25] and examine the details about each network on a dashboard laid over Google Maps. ,

===Figures from 1 to 5 inserted about here===

Discussions

We observed that the top one(SC) in each topic consists of (1) journal article(0.81), (2) China(0.61),(3) Acad Radol, (4) betacoronavirus (0.66), and (5) Hazhong University of Science and Technology(0.77) in article types, countries, journals, MeSH terms, and research affiliations, respectively.

We applied the SC to represent the extent to which an entity plays a dominant role in a group. The SC concept is originated from test theory for identifying a one-dimension scale on an underlying construct[7–10].

The SC formula is defined as $(= [(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)]/[1+[(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)]])$. That can be divided into two parts: $F1 = (\lambda_1/\lambda_2)$ and $F2 = (\lambda_2/\lambda_3)$ referred to the analysis of variance(ANOVA), where F-value(= the sum of squares (SS) ÷ the error). The reliability equals $F1/(F1 + F2) [= 1 - F2/(F1 + F2) = (F1/F2)/(1 + (F1/F2))]$. As such, the cutoff-point is set at 0.70, equivalently equal to the simulation result[7].

We found that only two topics (i.e., journal type and research institute with SCs = 0.81 and 0.77) reached the acceptable level(= 0.70), indicating both(i.e., journal article and Hazhong University of Science and Technology) deserve the dominant roles played in their networks.

The latest public health information on COVID-19 has been released from Centers for Disease Control and Prevention (CDC) and National Institutes of health (NIH)[21–24]. Over 1,674 articles were found in PMC when the keyword “social network analysis” was searched as of April 14, 2020, and 444 with “social network analysis” in the title. We particularly applied SNA to extract information about the association among entities in publications on COVID-19, which is novel and rarely seen on the topic of COVID-19.

The second feature is about the AWS used for quantifying author contributions to the article. In this study, AWS was applied to authors’ origin countries and research affiliations. Others link MeSH terms and article types are assumed each entity with equal credit in an article.

The third feature is dashboards displayed on Google Maps. We provided hyperlinks [21–25] for readers to see details about the topic of interest(e.g., networks shown in Figures) on their own.

Although findings are based on the above analysis, there are still several potential limitations that may encourage further research efforts. First, all data were extracted from the PMC. There might be some biases of the author's institutes affiliated with countries/regions because of some blanks or typos in the article, which will affect the result of analysis by the accuracy of the indexing affiliated countries/regions.

Second, the proposed SC has several limitations, such as at least three entities are required to compute the SC. Whether the SC is more suitable for comparing the dominant role than those with HHI[1–3] and/or Gini[4–6] is worthy of further investigations in the future.

Third, the data extracted from PMC cannot be generalized to other major citation databases—such as the Scientific Citation Index (SCI; Thomson Reuters, New York, NY, USA) and Scopus (Elsevier, Amsterdam, The Netherlands). As such, the most influential (or dominant) roles for research institutes and others are determined by the paper selections on journals indexed in databases.

In conclusion, We applied SC to identify the strength of the top one over the next two. The SC was useful and viable in verifying the dominant role in a group. The implementation and application are worthy of further studies in the future.

Abbreviations

DC
dimension coefficient
HHI
Herfindahl index
MeSH
medical subject headings
SC
strenght coefficient
SI
separation index
PMC
Pubmed Central

Declarations

Ethics approval and consent to participate

Not applicable.

All data were downloaded from the website database in Pubmed

Consent to publish

Not applicable.

Availability of data and materials

All data used in this study are available in Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

There are no sources of funding to be declared.

Authors' Contributions

WC developed the study concept and design. JCJL and YT analyzed and interpreted the data. TWC monitored the process of this study and helped in responding to the reviewers' advice and comments. TWC drafted the manuscript, and all authors provided critical revisions for important intellectual content. The study was supervised by TWC. All authors read and approved the final manuscript.

Acknowledgments

We thank Enago (www.enago.tw) for the English language review of this manuscript. All authors declare no conflicts of interest.

References

1. Lee CJ, Chou W, Chien TW, Yeh YT, Jen TH.(2020). Using the Separation Index for Identifying the Dominant Role in an Organization: A case of Publications in Organizational Innovation, Journal of Organizational Innovation 2020;12(4)"135–145.
2. Hirschman AO. **The Strategy of Economic Development**. (Yale Studies in Economics: 10.) New Haven: Yale University Press, 1958.
3. Catherine LH, Alan P. Inventive Concentration: An Analysis of Fuel Cell Patents. Science Public Policy. 2004;31(1):15–25.
4. Gini C. Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica. 1997;87:769–89.
5. Chien TW, Chang Y, Wang HY. Understanding the productive author who published papers in medicine using National Health Insurance Database: A systematic review and meta-analysis. Medicine. 2018;97(8):e9967. doi:10.1097/MD.0000000000009967.
6. Chien TW, Wang HY, Kan WC, Su SB. Whether article types of a scholarly journal are different in cited metrics using cluster analysis of MeSH terms to display: A bibliometric analysis. Medicine. 2019;98(43):e17631. doi:10.1097/MD.00000000000017631.
7. Chien TW. Cronbach's Alpha with the Dimension Coefficient to Jointly Assess a Scale's Quality. Rasch Measurement Transactions. 2012;26(3):1379.
8. Chien TW, Wang HY, Hsu CF, Kuo SC. Choropleth map legend design for visualizing the most influential areas in article citation disparities: A bibliometric study.Medicine 2019; 98: 41, e1752.

9. Divgi DR. Dimensionality of binary items: Use of a mixed model. Paper presented at the annual meeting of the National Council on Measurement in Education. Boston, MA, 1980/.
10. Lord FM. Applications of item response theory to practical testing problems. Hillside: Erlbaum; 1980.
11. Bright CF, Haynes EE, Patterson D, Pisu M. The Value of Social Network Analysis for Evaluating Academic-Community Partnerships and Collaborations for Social Determinants of Health Research. *Ethn Dis* 2017 Nov 09;27(Suppl 1):337.
12. deNooy W, Mrvar A, Batagelj V. Exploratory Social Network Analysis With Pajek: Revised and Expanded. 2nd ed. New York: Cambridge University Press; 2011.
13. Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*. 1978 Jan;1(3):215–39.
14. deNooy W, Mrvar A, Batagelj V. Exploratory Social Network Analysis With Pajek: Revised and Expanded. 2nd ed. New York: Cambridge University Press; 2011.
15. 10.2196/11015
Chien TW, Wang HY, Lai FJ Applying an Author-Weighted Scheme to Identify the Most Influential Countries in Research Achievements on Skin Cancer: Observational Study *JMIR Dermatol* 2019;2(1):e11015. URL: <https://derma.jmir.org/2019/1/e11015> DOI: 10.2196/11015.
16. Vavryčuk V. Fair ranking of researchers and research teams. *PLoS ONE* 2018 Apr 5;13(4):e0195509.
17. Egghe L, Rousseau R, Van Hooydonk G. Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *J Am Soc Info Sci*. 2000;51(2):145–57. [FREE Full text] [CrossRef].
18. Batagelj V, Mrvar A. Pajek - Analysis and Visualization of Large Networks. In: International Symposium on Graph Drawing. Berlin, Germany: Springer Nature; 2003:477–478.
19. Chien T, Chow JC, Chang Y, Chou W. Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names: A bibliometric study. *Medicine*. 2018 Sep;97(39):e12418.
20. Chien T, Wang H, Chang Y, Kan W. Using Google Maps to display the pattern of coauthor collaborations on the topic of schizophrenia: A systematic review between 1937 and 2017. *Schizophr Res*. 2019 Feb;204:206–13.
21. Chien TW. Dashboard of article types in publications on COVID-19. 2020/4/14 available at <http://www.healthup.org.tw/gps/covidjournal.htm>.
22. Chien TW. Dashboard of affiliated countries in publications on COVID-19. 2020/4/14 available at <http://www.healthup.org.tw/gps/covidnation2.htm>.
23. Chien TW. Dashboard of journals in publications on COVID-19. 2020/4/14 available at <http://www.healthup.org.tw/gps/covidnation23.htm>.
24. Chien TW. Dashboard of MeSH terms in publications on COVID-19. 2020/4/14 available at <http://www.healthup.org.tw/gps/covidnation24.htm>.

25. Chien TW. Dashboard of research institutes in publications on COVID-19. 2020/4/14 available at <http://www.healthup.org.tw/gps/covidnation25.htm>.
26. CDC. The latest public health information on COVID-19. 2020/4/17 available at <https://www.coronavirus.gov>.
27. NHI. The latest public health information on COVID-19. 2020/4/17 available at <https://www.coronavirus.gov>.
28. COVID-19 Data Curation Group. Open access epidemiological data from the COVID-19 outbreak [published online ahead of print, 2020 Feb 19]
10.1016/S1473-3099(20)30119-5
Xu B, Kraemer MUG, Open. COVID-19 Data Curation Group. Open access epidemiological data from the COVID-19 outbreak [published online ahead of print, 2020 Feb 19]. *Lancet Infect Dis*. 2020;S1473-3099(20)30119-5. doi:10.1016/S1473-3099(20)30119-5.
29. Wolkewitz M, Puljak L. Methodological challenges of analysing COVID-19 data during the pandemic. *BMC Med Res Methodol*. 2020;20(1):81. doi:10.1186/s12874-020-00972-6. Published 2020 Apr 14.

Figures

Figure 1

Distribution of article types on COVID-19

Figure 2

Distribution of author origin affiliations on COVID-19

Figure 3

Distribution of author origin affiliations associated with journals on COVID-19

Figure 4

Distribution of Mesh Terms on COVID-19

Huazhong University of Science and Technology(China)	724.4
Renmin Hospital of Wuhan University(China)	209.9
The University of Hong Kong(Hong Kong)	208.0
Fudan University(China)	178.1
Capital Medical University(China)	169.8
the First Affiliated Hospital of Guangzhou Medical University(China)	156.5
Zhejiang University(China)	144.2
Second Affiliated Hospital of Nanchang University(China)	3.0
Baghdad Teaching Hospital(Iraq)	1.0
Birmingham City University(UK)	1.0

Figure 5

Distribution of research institutes on COVID-19