Research preferences of the G20 countries: a bibliometrics and visualization analysis

Gege Lin, Zhigang Hu and Haiyan Hou*

The purpose of this study is to reveal the differences both in research output and research preferences of the G20 countries. The research outputs of the nineteen G20 countries (excluding the European Union) are measured based on their publications indexed in Web of Science. The research preferences of the G20 countries were studied by comparing their research output in each research subject. Clustering method was then employed to classify the countries according to their research preferences. Nineteen countries are classified into four clusters. Countries assigned to the same cluster are similar in distribution of research subjects. In the end, by VOSviewer, we showed the research pattern of each cluster. For example, USA in Cluster A is characterized by the emphasis on medical sciences and China in Cluster C is characterized by paying more attention to physical sciences.

Keywords: Bibliometrics, country-level studies, G20 countries, research preferences, VOSviewer.

THE G20 is initially an international economic cooperation forum established in 1999. After almost a twenty-year development, the G20 has become a major platform for international affairs and has played an increasingly important role in all kinds of global issues, including scientific research[1]. The G20 countries account for about 60% of the world's land area, 66.7% of the world population, and more than 90% of the sum of Gross Domestic Product. They are also the dominant producers of scientific research output, and serve as major engines to drive further development in science and technology.

However, the driving effects of the G20 countries vary sharply, both in their strength and their preferential research areas. Yang *et al.*[2] studied the research preference of the G7 countries, and found that life sciences are the main focus of these developed countries. Bouabid *et al.*[3] addressed the issue of scientific collaboration among the G7 countries; their research showed that the G7 countries had intensive intra-collaboration activities. Thomson Reuters (now named Clarivate Analytics) investigated the research and innovation performance of the G20 in a report[4] in 2014, and listed the amount and world share of each G20 country's publications in the select Organization for Economic Co-operation and Development (OECD) research fields. Hu *et al.*[5,6] explored research preferences of the provinces of China by the method of cosine similarity and hierarchical clustering, and mapped different provinces research hotspots using VOSviewer. Almeida *et al.*[7] analysed the way European countries are clustered according to their similarity.

In this study, we examine the research output and preference of the G20 countries. The following questions will be addressed below: (a) how great is the difference among the G20 countries in their research output? (b) what is the research preference for each country? (c) which countries have similar/different research preferences? (d) does a country's economic level have influence on its research preferences?

Data and methods

Data collection

The G20 countries' publications were retrieved from Web of Science (WoS), one of the world's most comprehensive bibliometric database. WoS Core Collection consists of three core journal databases (SCI, SSCI, A&HCI) and two important conference databases (CPCI-S and CPCI-SSH) that are the data sources of this study.

All journals and proceedings volumes are assigned to one or more research subjects, such as engineering electrical electronics, materials science, oncology, etc. In WoS, we retrieved the number of publications of the G20 countries for each research subject in 2015, and the statistical results showed that all publications counted covered more than 250 subjects. To exclude non-research articles, only article, proceedings article and review were included, and all the other data type, such as letter, note, and edictal martial were ignored.

The authors are in the WISE Lab, Dalian University of Technology, Dalian, China.
*For correspondence. (e-mail: huzhigang@dlut.edu.cn)

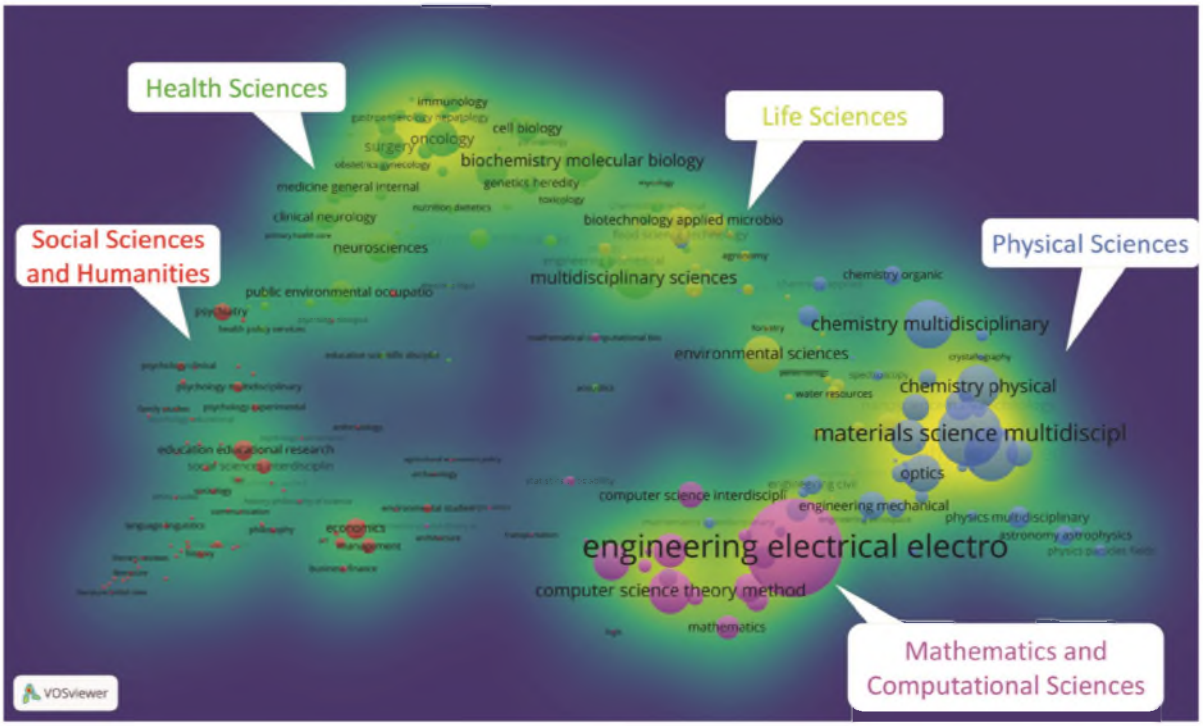

Figure 1. Research hotspot map of global researchers.

Similarity measurement among G20 countries

Two countries are similar in research preference if their cosine distance is close. As this study intended to compare the research preference instead of absolute research output, the cosine distance is more appropriate than Euclidean distance. Cosine distance refers to cosine of the angle between two vectors. Generally, the angle between two vectors is used as a measure of divergence between the vectors, and cosine of the angle is used as the numeric similarity (because cosine has the nice property that it is 1.0 for identical vectors and 0.0 for orthogonal vectors)[8]. The cosine similarity of two vectors X and Y is represented using a dot product and magnitude as

$$\text{similarity}\ (X, Y) = \cos(\theta) = \frac{X \cdot Y}{\| X \| \cdot \| Y \|}. \qquad (1)$$

The publications of each G20 country in various subjects were counted in 2015, and the disciplinary distribution vector of every country was constructed. We can calculate the degree of discipline similarity between any two countries based on formula (1).

After calculating the similarity of the G20 countries, we further classified them into four clusters using the method of hierarchical clustering. Hierarchical clustering[6,9] showed not only the result of clustering, but also the clustering process through a tree diagram. Further, the most similar two countries were merged. The process continued until all nineteen countries were merged.

Research hotspot of the G20 countries

To illustrate the research preference vividly, we drew research hotspots for several typical countries using the visualization technology in the end. VOS viewer, a software tool developed by Van Eck and Waltman for constructing bibliometrics networks[10–12], was employed.

Figure 1 is the research hotspot map, which provides a foundational research hotspot in the world today. The first layer (namely, network map) of the map is derived from the official website of VOSviewer. In the first layer, all the WoS research subjects are located based on their citation relations to each other. Five research fields are clustered and assigned with distinguished colours. They are social sciences and humanities in red, health sciences in green, life sciences in yellow, physical sciences and engineering in blue, and mathematics and computational sciences in purple.

Based on the first layer, we generated network maps for different G20 countries by keeping the layout of the map constant. The size of each node was redrawn according to the number of publications of the country in the corresponding research subject. The research network is able to show the relatedness of nodes and the weight of the item in the network. A new method was used to make the research hotspot map, which consists of two layers; one layer indicates research fields (namely, the first layer) and the other (namely, the base layer) indicates research hotspots. The base layer colours range from blue to green to yellow. The larger the number of items in the

Table 1. Publication counts of G20 countries in Web of Science

Rank	Country	Publications	GDP Rank	Rank	Country	Publications	GDP Rank
1	USA	447,685	1	11	South Korea	64,178	11
2	China	357,727	2	12	Brazil	48,782	9
3	UK	125,333	5	13	Russia	45,013	12
4	Germany	125,139	4	14	Turkey	32,309	18
5	Japan	91,150	3	15	Mexico	15,437	15
6	France	86,973	6	16	Saudi Arabia	15,096	20
7	India	81,608	7	17	South Africa	14,658	33
8	Italy	80,046	8	18	Argentina	9,889	24
9	Canada	76,045	10	19	Indonesia	5,530	16
10	Australia	68,241	13				

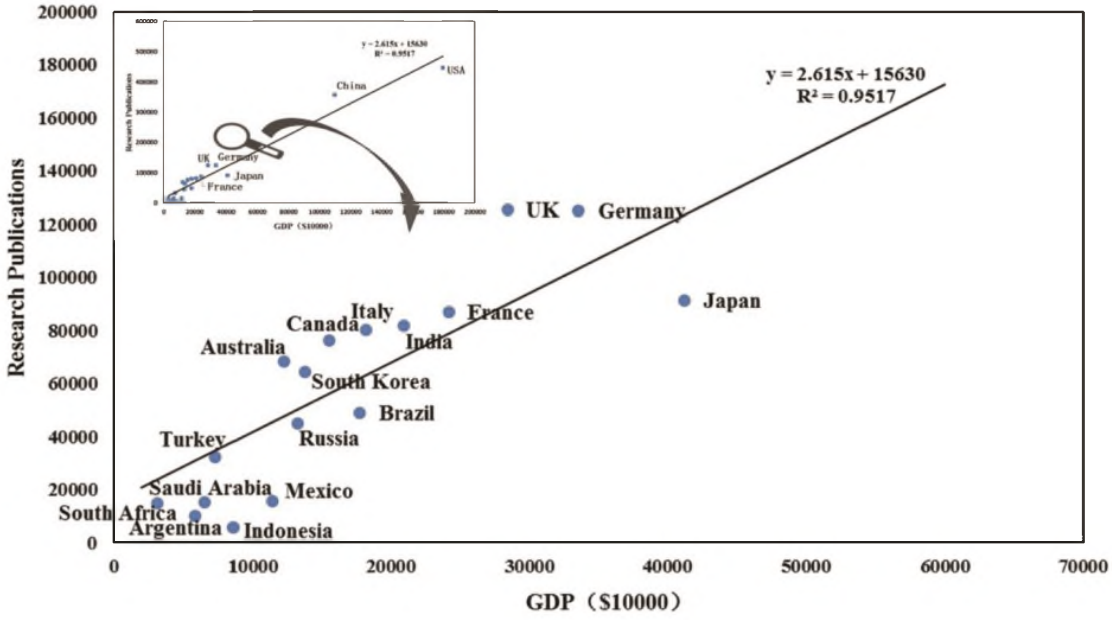

Figure 2. Relationship between GDP and the number of publications (2015).

neighbourhood of a point and the higher the weights of the neighbouring items, the closer the colour of area of the point is to yellow[10]. In this way, we can easily find out the research hotspot subjects in each research field.

Cluster analysis of the G20 countries

Research output

The number of publications of the G20 countries in 2015 is listed in Table 1. The United States, not surprisingly, ranks first with around 0.45 million publications. China ranks second with around 0.36 million publications. They are the undisputed leaders in research output, far beyond anyone else in the G20 countries. The sum of publications from the third ranked UK, the fourth ranked Germany and the fifth ranked Japan are still less than China's research output, let alone the US. Obviously, the G20 countries vary greatly in terms of counts of publications. The research output of the US is almost 81 times that of Indonesia, the lowest producer in the G20.

Generally, a country's scientific research output is proportional to its economic level. USA and China, which produced the most research publications, are also the bigger economic powers. UK, Germany and Japan rank the third to fifth in research publications; their GDPs also rank from third to fifth, just in reversed order. France, India and Canada rank sixth to eighth both in research publications and GDP. The top 13 research producers are exactly the same as the richest countries, only in a different sequence.

Figure 2 shows the correlation between the research output and the GDP of the G20 countries. For each country, coordinate X represents its GDP in 2015 and Y represents its publication count. It shows that a country's GDP is significantly ($R^2 = 0.9517$) correlated with its research output.

Table 2. Publications of research subjects in Web of Science (2015)

Web of Science categories	Pub.	Web of Science categories	Pub.
Engineering electrical electronic	163,833	.....	.....
Materials science multidisciplinary	97,143	Literature Slavic	580
Physics applied	77,730	Psychology psychoanalysis	566
Chemistry multidisciplinary	67,224	Andrology	502
Chemistry physical	59,842	Literature German Dutch Scandinavian	463
Computer science theory methods	59,024	Literature British Isles	421
Multidisciplinary sciences	56,745	Dance	367
Biochemistry molecular biology	53,522	Literature American	325
Optics	48,839	Folklore	306
Environmental sciences	47,639	Literature African Australian Canadian	154
.....	.....	Poetry	116

Table 3. Preferential research subjects of each G20 country in Web of Science TM (2015)

USA	China	UK	...	Indonesia
Engineering electrical electronic (28,547)	Engineering electrical electronic (41,016)	Engineering electrical electronic (6,782)	...	Engineering electrical electronic (984)
Multidisciplinary sciences (18,388)	Materials science multidisciplinary (33,912)	Multidisciplinary sciences (5,567)	...	Computer science information systems (575)
Materials science multidisciplinary (16,104)	Chemistry multidisciplinary (21,104)	Materials science multidisciplinary (3,866)	...	Physics applied (549)
Biochemistry molecular biology (15,395)	Physics applied (19,927)	Astronomy astrophysics (3,706)	...	Computer science theory methods (390)
Neurosciences (14,131)	Chemistry physical (17,192)	Neurosciences (3,648)	...	Environmental sciences (325)
Physics applied (14,038)	Energy fuels (14,283)	Biochemistry molecular biology (3,440)	...	Telecommunications (248)
Oncology (13,375)	Optics (13,162)	Physics applied (3,227)	...	Computer science interdisciplinary applications (240)
Public environmental occupational health (12,049)	Computer science theory methods (12,787)	Computer science theory methods (2,976)	...	Engineering industrial (234)
...	...	...	...	...
Dance (29)	Literature German Dutch Scandinavian (0)	Literature American (6)	...	Sport sciences (0)
Literature Slavic (20)	Literature romance (0)	Literature Slavic (6)	...	Transplantation (0)

Publications of research subjects

Table 2 lists the number of publications of 252 research subjects. 'Engineering electrical electronic' is the largest subject with 163,833 publications, and 68.7% higher than the second-placed 'materials science multidisciplinary'. The subjects of 'physics applied', 'chemistry multidisciplinary' and 'chemistry physical' rank third to fifth respectively. The top 5 research subjects contributed more than 1/4 of all publications. Correspondingly, the publications in the subjects of 'literature African Australian Canadian' and 'poetry', are only less than 200 and merely account for 0.0008%.

Research preferences

Table 3 shows the research preferences of each G20 country. In 2015, the USA researchers published the most publications in 'engineering electrical electronic', 'multidisciplinary sciences', 'materials science multidisciplinary', 'biochemistry molecular biology' and 'neurosciences'. Chinese researchers did well in 'engineering electrical electronic', 'materials science multidisciplinary', 'chemistry multidisciplinary', 'physics applied' and 'chemistry physical'. Compared to USA and China, UK paid more attention to 'astronomy astrophysics', and Indonesia published more papers in 'computer science information systems', 'computer science theory methods' and 'telecommunications'.

Similarity in research preferences

The cosine distances between two countries were calculated based on the data in Table 3. As shown in Table 4, the resulting adjacency matrix represents the level of similarity between two countries. For example, the cosine similarity between Australia and China is 0.7991 whereas

Table 4. Adjacency matrix of G20 countries' cosine similarity in research areas

Country	Australia	China	France	India	……	Japan	Russia	UK	USA
Australia	1								
China	0.7991	1							
France	0.8906	0.9019	1						
India	0.7546	0.9224	0.8864	1					
……	……	……	……	……	……				
Japan	0.8345	0.9121	0.9519	0.9006	……	1			
Russia	0.6687	0.8344	0.8625	0.7952	……	0.8422	1		
UK	0.9634	0.802	0.9317	0.7707	……	0.8801	0.7238	1	
USA	0.9495	0.8311	0.9359	0.7883	……	0.9183	0.7223	0.976	1

Table 5. The cosine distance between G20 countries (in descending order)

Country A	Country B	Cosine distance	Country A	Country B	Cosine distance
France	Germany	0.9797	……	……	……
UK	USA	0.9760	Indonesia	USA	0.6733
Canada	USA	0.9736	Brazil	Russia	0.6718
Canada	UK	0.9683	Australia	Russia	0.6687
Australia	Canada	0.9653	India	South Africa	0.6531
Australia	UK	0.9634	Argentina	India	0.6475
France	Italy	0.9603	South Africa	South Korea	0.6445
Germany	Italy	0.9562	Indonesia	Russia	0.6435
Germany	Japan	0.9561	Russia	South Africa	0.6338
Germany	USA	0.9547	Indonesia	South Africa	0.6199
……	……	……	Argentina	Indonesia	0.5005

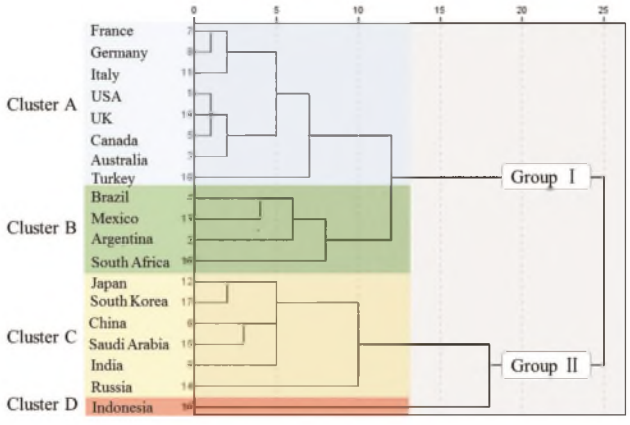

Figure 3. Dendrogram representing the clustering result of the G20 countries.

that between Australia and France is 0.8906. It means Australia is more similar to France than China in the distribution of research subjects.

Table 5 lists the country pairs with the most and the least similarity. France and Germany, two adjacent countries located in Western Europe, are closest in the vector space of research subjects. Their similarity is 0.9797. This similarity between them also shows in Almeida's research[7]. The next closest pair is UK and USA, both important developed countries in the world with similar economic status and structure. The third pair is Canada and USA. Although the absolute count of publications varies a lot between Canada and USA, both of them have a homogeneous disciplinary structure and research preferences. The most dissimilar country pair is Argentina and Indonesia. Their similarity is only 0.5005.

Clustering analysis

The dendrogram of the cluster analysis is shown in Figure 3. This is a pictorial representation of the data structure, indicating the merging objects and the merging distances.

Cluster A is composed of eight countries: France, Germany, Italy, USA, UK, Canada, Australia and Turkey. Except Turkey, all the others are developed countries and located in Europe and North America.

Cluster B is composed of four countries: Brazil, Mexico, Argentina and South Africa. They are all former European colonies, located in either Latin America or Africa.

Cluster C consists of six Asian countries namely Japan, South Korea, China, Saudi Arabia, India and Russia.

Cluster D contains only one country, namely, Indonesia, the smallest research producer in the G20. It is a unique country with more than seventeen thousand islands, and located far from the Asian continent.

At a highest level, the eleven countries in clusters A and B are merged into group I whereas the other seven countries in clusters C and D are merged into group II.

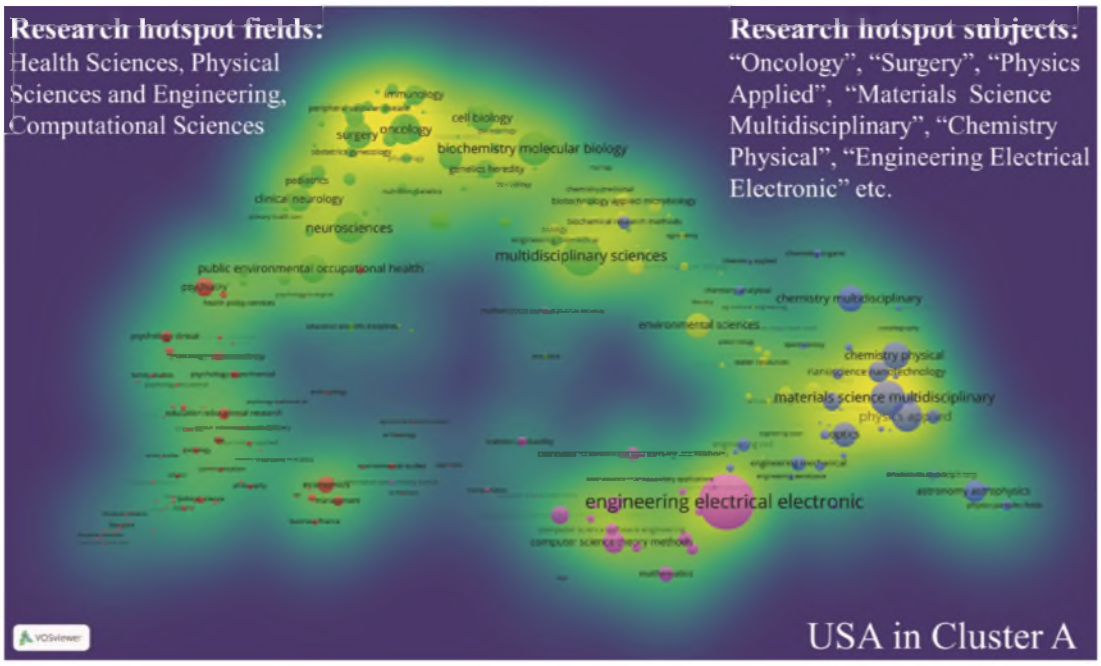

Figure 4. Research hotspot map of USA in Cluster A.

Group I is composed of developed countries mainly from the west. Compared to those in Group I, the countries in Group II are almost located in Asia, and most of them are developing countries.

Although economic level has a significant effect on discipline structure and subject development, geographical locations might have a greater influence on research preferences. For example, Japan and South Korea, two developed countries, belong to Group II instead of Group I.

Research hotspot map of the clusters

Each cluster is featured by its unique research preference. In the following section, we selected one typical country of each cluster. They are the USA in cluster A, Brazil in cluster B, China in cluster C, and Indonesia in cluster D.

Research hotspot map of the USA

The research hotspot map of USA is shown in Figure 4. Health sciences, physical sciences and engineering are the hotter fields in USA. Their hotspot subjects include 'oncology', 'immunology', 'surgery', etc., 'materials science multidisciplinary', 'physics applied', 'chemistry physical', etc. respectively.

Besides, USA is also an important producer of research papers in the subject of 'engineering electrical electronic'. Furthermore, there are no obvious hotspot subjects in social sciences and humanities and life sciences.

USA shows a typical research hotspot map for North American and European countries, called 'Western Model' by a report published in 1997 (ref. 13). In this report, 'Western Model' is defined by focusing on research in clinical medicine and biomedical sciences.

Research hotspot map of Brazil

Brazil is featured by its advantage in life sciences (Figure 5). It mainly focuses on 'plant sciences', 'agronomy', 'food science technology', 'environmental sciences' and so on. Abundant natural resources and agricultural production provide an essential necessity and condition for Brazil and other countries in Cluster B.

Health sciences is also a preferential research field of Brazil. In this field, the hotspot subjects are 'biochemistry molecular biology', 'genetics heredity', 'dentistry oral surgery medicine', 'parasitology', etc. In the research fields of physical sciences and engineering or mathematics and computational sciences, Brazil also had good performance, and the subjects of 'materials science multidisciplinary', 'engineering chemical', 'engineering electrical electronic', etc. received more attention. But the Brazilian researchers seemingly pay less attention to social sciences and humanities, in which field the only primary subject of Brazil is psychiatry.

Brazil's research map is in accordance with the result of the report[13]. As we find, Brazil still follows the 'bio-environmental model' with biology, earth and space sciences in the main focus. The finding is also confirmed by a study of Glänzel[14] in 2006.

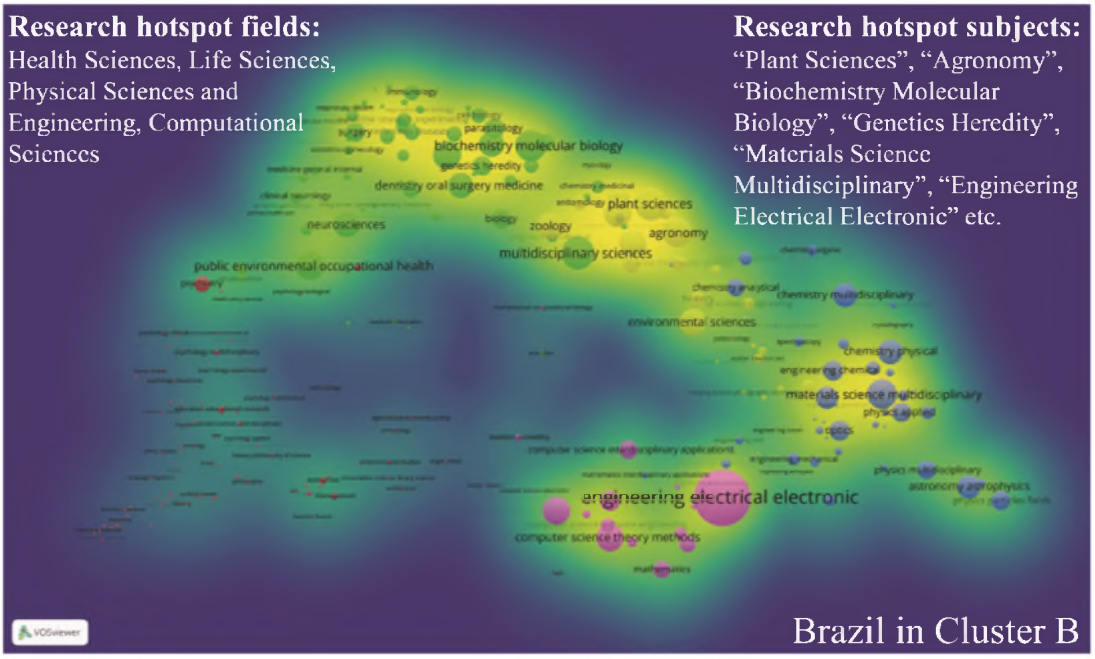

Figure 5. Research hotspot map of Brazil in Cluster B.

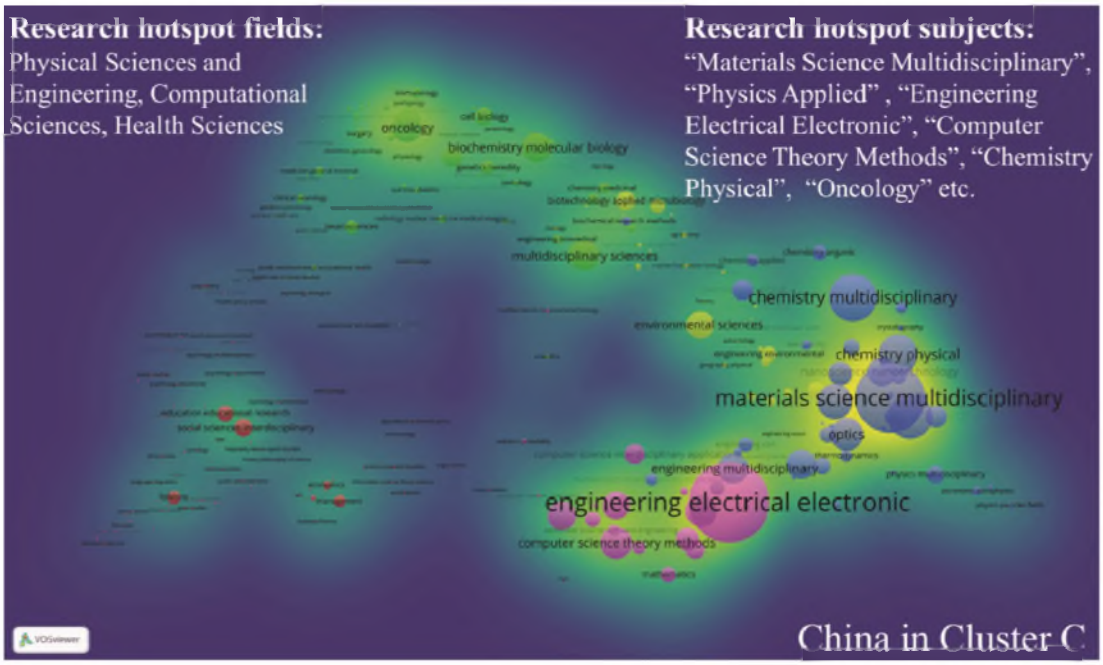

Figure 6. Research hotspot map of China in Cluster C.

Research hotspot map of China

China is a typical developing country. It shows entirely different research preference compared to other clusters (Figure 6). Chinese researchers have been more active in the field of physical sciences and engineering, as well as mathematics and computational sciences. 'Materials science multidisciplinary', 'chemistry physical', 'physics applied', etc. and 'engineering electrical electronic', 'computer science theory methods', 'engineering multidisciplinary' in the two fields respectively are the most productive subjects in China. The emphasis on physical sciences is also stated in a previous study[15]. In comparison, health sciences and life sciences

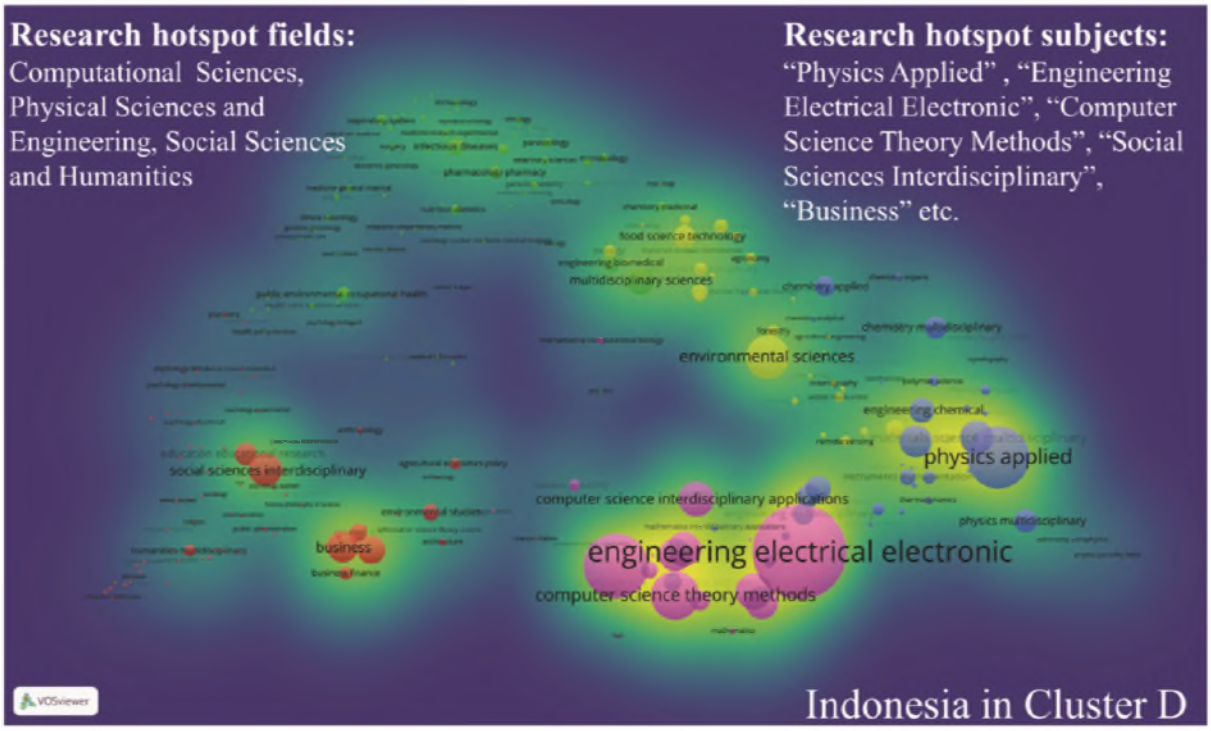

Figure 7. Research hotspot map of Indonesia in Cluster D.

are not so preferred in China compared to USA and Brazil.

Besides, China also focuses on the subjects of 'oncology' in health sciences. But the fields of social sciences and humanities and life sciences have not formed obvious hotter subjects.

Research hotspot map of Indonesia

Indonesia published the highest share of papers in mathematics and computational sciences (Figure 7). In this field, 'engineering electrical electronic', 'computer science theory methods', 'automation control systems', 'computer science information systems' and 'computer science artificial intelligence' are the most preferred subjects.

Physical sciences and engineering is another important research field for Indonesia where 'physics applied', 'materials science multidisciplinary' and 'energy fuels' are the most dominant subjects. Indonesia also focused on social sciences and humanities, such as 'social sciences interdisciplinary', 'business', 'management', 'education educational research', etc. compared to health sciences, Indonesia prefers to conduct research in life sciences.

In the end, we drew radar charts for each of the above countries, and revealed the difference in their research maps, as shown in Figure 8. By comparing the shape of radar charts, we are able to distinguish the research spotlight of each country or cluster more clearly. The different characteristics of clusters in research preferences could be identified obviously.

Conclusions

We compared the research performances of the nineteen G20 countries in this study. We found that different countries vary greatly not only in research outputs, but also in research preferences. The research publications of USA could be 80 times that of Indonesia. The difference is in terms of research preferences as well. By clustering method, the nineteen G20 countries are first classified into four clusters, and then two groups at a higher level.

Each cluster is featured with a particular research preference. For example, the countries in cluster A, e.g., the USA, prefer the research area of health sciences whereas those in cluster C, e.g., China, lay greater emphasis on physical sciences and engineering.

The study also found significant correlation between national research performance and economic level or geographic location of a country. If some countries have similar economic levels or their locations are close to each other, their research preferences tend to be similar. The developed countries conduct more research in biology and medical sciences whereas the developing countries emphasize on physical sciences and computer sciences.

Using VOSviewer and radar charts, we provide research maps of four typical countries in each cluster. It

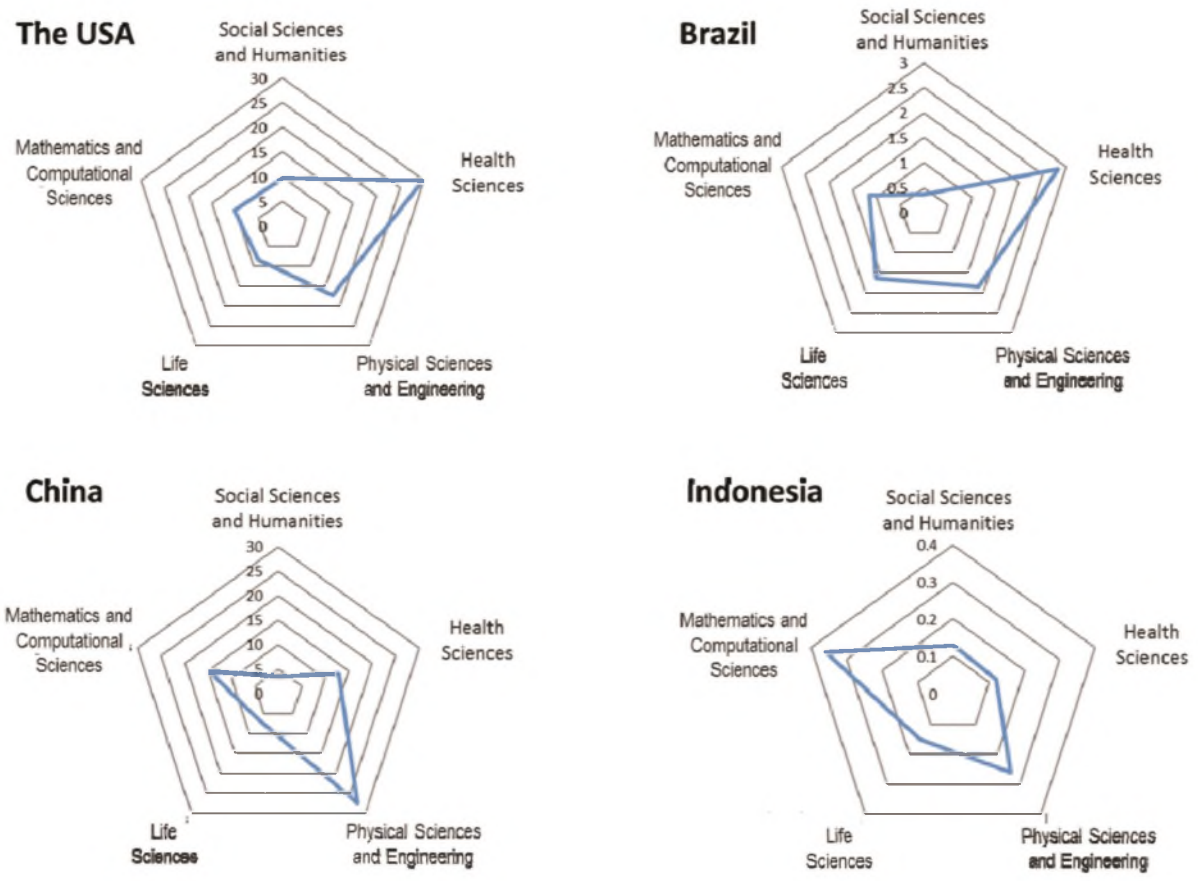

Figure 8. Radar charts of the selected four countries in different clusters.

allows us to compare research preferences of different clusters more intuitively. In the future, the method and visualizations could be extended to the international level for comparison among different universities or regions.

1. Callaghan, M., Ghate, C., Pickford, S. and Rathinam, F., Global Cooperation Among G20 Countries: Responding to the Crisis and Restoring Growth. 2014, 1–21.
2. Yang, L. Y., Yue, T., Ding, J. L. and Han, T., A comparison of disciplinary structure in science between the G7 and the BRIC countries by bibliometric methods. *Scientometrics*, 2012, **93**, 497–516.
3. Bouabid, H., Paul-Hus, A. and Larivière, V., Scientific collaboration and high-technology exchanges among BRICS and G-7 countries. *Scientometrics*, 2016, **106**, 873–899.
4. Reuters, T., The research and innovation performance of the G20. Thompson Reuters, London, 2014.
5. Hu, Z., Lin, G., Sun, T. and Wang, X., An EU without the UK: mapping the UK's changing roles in the EU scientific research. *Scientometrics*, 2018, **115**(3), 1185–1198.
6. Hu, Z., Guo, F. and Hou, H., Mapping research spotlights for different regions in China. *Scientometrics*, 2017, **110**, 779–790.
7. Almeida, J. A. S., Pais, A. A. C. C. and Formosinho, S. J., Science indicators and science patterns in Europe. *J. Informetr.*, 2009, **3**, 134–142.
8. Singhal, A., Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 2001, **24**, 35–43.
9. Zhao, K. and Suzuki, E., Clustering classifiers learnt from local datasets based on cosine similarity. 2015, 150–159.
10. Eck, N. J. V. and Waltman, L., *VOSviewer Manual*, version 1.6.8, CWTS, Leiden Google Scholar, 27 April 2018, 1–51.
11. Ding, Y., Rousseau, R. and Wolfram, D., Measuring Scholarly Impact, Springer International, Pu, 2016.
12. van Eck, N. and Waltman, L., Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 2009, **84**, 523–538.
13. European Commission Directorate–General for Telecommunications IMAE. The European Report on Science and Technology Indicators. 1994.
14. Glänzel, W., Leta, J. and Thijs, B., Science in Brazil. Part 1: A macro-level comparative study. *Scientometrics*, 2006, **67**, 67–86.
15. Fu, H., Chuang, K., Wang, M. and Ho, Y., Characteristics of research in China assessed with essential science indicators. *Scientometrics*, 2011, **88**, 841–862.

ACKNOWLEDGEMENTS. The research was supported by the Natural Science Foundation of China (NSFC) under Grant 71503031 and China Postdoctoral Science Foundation under Grant 2016M591435.

Received 7 May 2018; revised accepted 4 July 2018

doi: 10.18520/cs/v115/i8/1477-1485