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. 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. studied the research preference of the G7 countries, and found that life sciences are the main focus of these developed countries. Bouabid et al. 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 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. 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. 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.
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). The cosine similarity of two vectors $X$ and $Y$ is represented using a dot product and magnitude as

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

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 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.
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. 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)

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

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

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

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
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.
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 report13. 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 Glänzel14 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\textsuperscript{15}. In comparison, health sciences and life sciences
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
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.

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.

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