

Scale, style and spread of global institutional research

Gangan Prathap

Most research in the world is organized around institutions which can be categorized as belonging to different sectors, e.g. higher education, government, health, etc. In this article we have evaluated how countries share and spread their research efforts across these sectors. This is made possible because the Scimago Institutions Rankings (SIR) uses a five-sector classification of the leading 5362 research institutions in the world. It is shown that each country has its distinctive style of sharing and spreading its research efforts across major sectors. Some countries are similar to others (e.g. Spain and the Netherlands, or Iran and Turkey), while some are very different from each other (e.g. Spain and Iran). Various row-wise and column-wise operations on a country–sector matrix allow us to quantify and visualize these styles and spreads, and measure their shares of the global institutional research efforts.

Keywords: Bibliometrics, institutions rankings, national comparison, research efforts.

‘THE greatest invention of the nineteenth century was the invention of the method of invention’, remarked the philosopher and mathematician Alfred North Whitehead¹ in *Science and the Modern World*¹. While ‘the beginnings of the scientific movement was confined to a minority among the intellectual elite’, it was in the 19th century that science, technology and innovation began to be organized on an institutional scale. Here, we shall conflate ‘organizations’ with ‘institutions’, where organizations which are ‘formal structures with an explicit purpose [consciously] created’² are governed by ‘humanly devised informal constraints (sanctions, taboos, customs, traditions and codes of conduct) and formal rules (constitutions, laws, property rights) that structure political, economic and social interaction’^{3,4}.

The principal actors in performing and publishing research are authors, affiliated to institutions (departments, schools, universities, hospitals, autonomous government institutions, hospitals and medical research institutes, private industries, etc.). Since the introduction of the Science Citation Index, there are now multiple aggregators like Clarivate, Scopus, Google Scholar, etc. that curate scientific metadata and update them at regular intervals. Using this there are secondary agencies that rank universities and higher educational institutions (HEIs) where most research is carried out, e.g. ARWU, QS, THE, LEIDEN, etc. Unlike such ranking exercises, which deal only with universities and HEIs, the Scimago Institu-

tions Rankings (SIR) also covers research-focused institutions in the government, private and health sectors, as well as an ‘others’ category that falls outside these classifications. Thus, a five-way classification of global institutional research sector-wise, country-wise and region-wise is available in one resource.

The SIR report (<http://www.scimagoir.com/index.php>) which has recently appeared online has been used in the present study. SIR yields a composite indicator that combines three different sets of indicators based on research performance (50% of the total weight, using primary bibliometric data from Scopus), innovation outputs (30% of the total weight, based on PATSTAT) and societal impact measured by its web visibility (20% of the total weight, based on Google and Ahrefs).

The five-way sector-wise classification of global institutional research is based on an extensive manual process of disambiguation of the institutions names using the institutional affiliation of documents in the Scopus database. The methodology takes note of merging, segregating and denomination changes at the institutional level. Publications and citations are attributed to each institution taking into account the institutional affiliation of each author in the ‘affiliation’ field of the database. A mixed system (manual and automated) of affiliations to one or more institutions, if required, is also ensured. Institutions are then grouped country-wise and region-wise. Results are generated each year from the data retrieved over a period of five years ending two years before the edition of the ranking. For instance, rankings for 2018 are based on results from the five-year period 2012–16. The only exception is the case of web indicators which have only been calculated for the last year.

Gangan Prathap is in the A. P. J. Abdul Kalam Technological University, Thiruvananthapuram 695 016, India.
e-mail: gangan_prathap@hotmail.com

Table 1. Row-wise processing of a 21×5 dimensional representation of institutional research sector-wise for 20 leading countries (L20) and rest of the world (RoW)

SIR 2018	Government	Health	Higher education	Private	Others	Total	Government	Health	Higher education	Private	Others	L1 distance
1	2	3	4	5	6	7	8	9	10	11	12	13
USA	61	204	432	36	26	759	0.08	0.27	0.57	0.05	0.03	1.00
China	154	26	414	14	6	614	0.25	0.04	0.67	0.02	0.01	1.00
France	231	41	122	2	3	399	0.58	0.1	0.31	0.01	0.01	1.00
India	62	12	197	0	0	271	0.23	0.04	0.73	0.00	0.00	1.00
Spain	87	110	62	1	0	260	0.33	0.42	0.24	0.00	0.00	1.00
Germany	130	24	85	7	8	254	0.51	0.09	0.33	0.03	0.03	1.00
Japan	25	31	176	9	2	243	0.10	0.13	0.72	0.04	0.01	1.00
Russian Federation	116	2	84	0	0	202	0.57	0.01	0.42	0.00	0.00	1.00
UK	14	62	114	5	2	197	0.07	0.31	0.58	0.03	0.01	1.00
Italy	56	48	68	1	2	175	0.32	0.27	0.39	0.01	0.01	1.00
Brazil	18	11	114	0	0	143	0.13	0.08	0.8	0.00	0.00	1.00
South Korea	28	19	86	2	0	135	0.21	0.14	0.64	0.01	0.00	1.00
Iran	8	3	115	0	0	126	0.06	0.02	0.91	0.00	0.00	1.00
Canada	10	43	53	0	3	109	0.09	0.39	0.49	0.00	0.03	1.00
Australia	10	48	41	1	3	103	0.10	0.47	0.4	0.01	0.03	1.00
Taiwan	5	22	67	1	2	97	0.05	0.23	0.69	0.01	0.02	1.00
Poland	30	5	58	0	0	93	0.32	0.05	0.62	0.00	0.00	1.00
Turkey	1	7	80	1	0	89	0.01	0.08	0.9	0.01	0.00	1.00
The Netherlands	15	13	17	0	0	45	0.33	0.29	0.38	0.00	0.00	1.00
Switzerland	6	12	15	3	1	37	0.16	0.32	0.41	0.08	0.03	1.00
L20	1067	743	2400	83	58	4351	0.25	0.17	0.55	0.02	0.01	1.00
RoW	204	159	830	83	5	1281	0.16	0.12	0.65	0.06	0.00	1.00
All	1271	902	3230	166	63	5632	0.23	0.16	0.57	0.03	0.01	1.00

Institutions are included if they have published at least 100 works in the Scopus database during the last year of the selected time period. In SIR 2018, 5632 institutions are ranked globally.

A country-wise comparison of national scientific publication performance based on simple counts of papers and citations was done by May⁵. King⁶ distinguished seven broad fields of research and computed productivity statistics by taking into account population and GDP. Since then many cross-country and cross-region comparisons have been made. An excellent and compact review can be found in Jurajda *et al.*⁷. However, a sector-wise classification of research at the institutional level showing the scale, style and spread globally is lacking. In this article, we use the SIR 2018 report as the basis for such a study using well-established matrix procedures and operations derived from graph-theoretic considerations⁸⁻¹⁰.

Methodology and results

Matrix representation of data

In Table 1, there are 21 rows corresponding to the number of institutions in each of 20 leading countries (L20) and rest of the world (RoW). Note that 17 countries from the economic G20 comprising the world's 20 biggest economies are in the L20 list. Spain, Poland and the

Netherlands, belonging to the L20 group, are not individual G20 countries but are represented by the European Union. Columns 2 to 6 in Table 1 show how the institutions in SIR 2018 are spread across five sectors (government, health, higher education, private and others). Note that institutions can vary enormously in size, e.g. Chinese Academy of Sciences (CAS) and Centre National de la Recherche Scientifique (CNRS) are parent organizations of a large number of constituent institutions with tens of thousands of individual authors, while some may have fewer than a hundred faculty (e.g. 65 at Indian Institute of Technology, Gandhinagar, India). This can be recognized as a 21×5 matrix representation of the 5362 institutions in SIR 2018. Following Krauze and McGinnis⁸, one can consider a 'space' where the basic elements are countries and their institutions. The space is constructed on the basis of connections among these elements (objects) and variables (in this case, the five sectors) in the form of matrices that reflect these connections through its metrics. Matrix transformations then allow various scientometric insights to be obtained⁸⁻¹⁰.

In our 21×5 dimensional object-variable space there are 21 objects (countries) and 5 variables (sectors). The 21×5 entries contain the actual number of institutions and provide a measure of the size or scale of how institutions are spread across sectors. An excellent visualization of this can be seen at <http://www.scimagoir.com/index.php?display=chart>. These entries can be interpreted

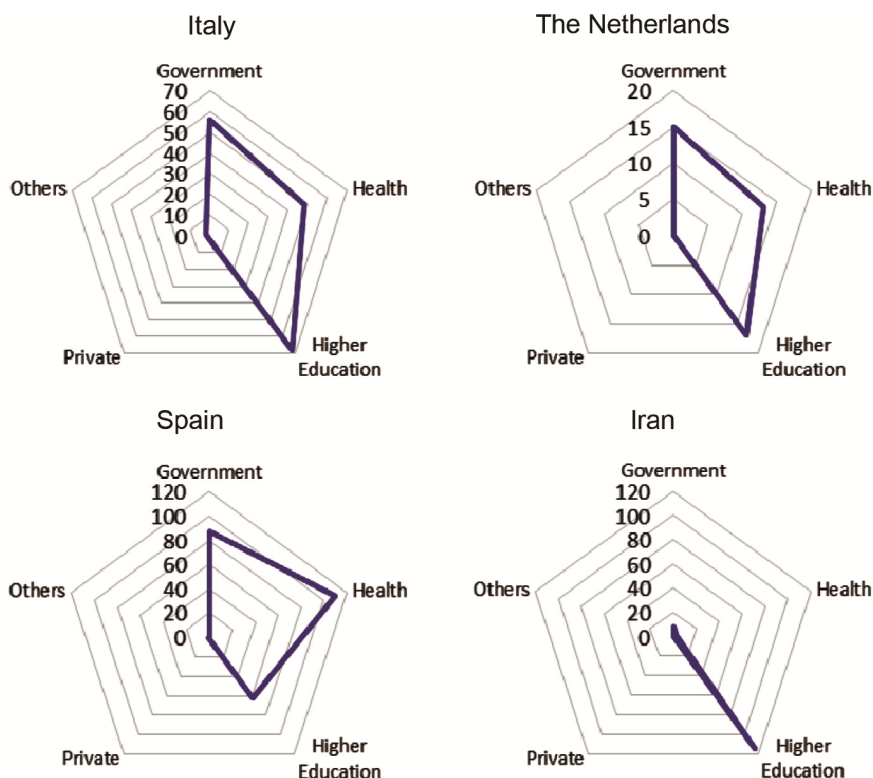


Figure 1. A radar-diagram visualization of the closest (Italy–The Netherlands) and the most distant (Spain–Iran) pairs in their styles of distributing research across sectors.

Table 2. Extract from 210 of the 10 closest and 10 most distant country-pairs using the concept of square of the L2 distances between the country vectors

Rank	Country-pair		Square of L2 distance
1	Italy	The Netherlands	0.001
2	USA	UK	0.003
3	China	India	0.004
4	South Korea	RoW	0.005
5	Iran	Turkey	0.006
6	Germany	France	0.006
7	China	Poland	0.008
8	Japan	RoW	0.010
9	Japan	Brazil	0.010
10	Canada	Australia	0.013
201	Australia	Iran	0.462
202	Spain	Brazil	0.476
203	Russian Federation	Iran	0.508
204	Germany	Iran	0.542
205	Russian Federation	Turkey	0.555
206	Germany	Turkey	0.570
207	France	Iran	0.640
208	Spain	Turkey	0.659
209	France	Turkey	0.675
210	Spain	Iran	0.688

row-wise or column-wise¹⁰, and each approach can give distinct insights into the spread or style of the way every country organizes its total research efforts. Also, each

approach can give different measures of how the shares of global research can be attributed to each country or cluster.

Row-wise processing of matrix data

As an example one can think of the vector (61, 204, 432, 36, 26) from the first row of Table 1 as identifying how institutional research in the US is spread across five sectors (government, health, higher education, private, others). Column 7 in Table 1 is a simple row-wise total count of the institutions across sectors assuming that the same weight is given to them. For the US we get $61 + 204 + 432 + 36 + 26 = 729$ as the total count. In such simple counting, L20 accounts for 77% of the institutions that perform research that meets the SIR 2018 cut-off (4351 out of 5362 institutions). We shall see later that by invoking a column-wise approach using the concept of the five-dimensional space of variables we will get non-recursive and recursive shares that are based on uniform and differential weighting respectively.

Columns 8–12 in Table 1 normalize the counts sector-wise in a row-wise fractional form. Thus, for each country we can identify a five-dimensional vector as the fractional share sector-wise and row-wise. Across the five sectors the vector of fractional shares for the US is now (0.08, 0.27, 0.57, 0.05, 0.03), and this gives a total of 1.0

Table 3. Column-wise processing of a 21×5 dimensional representation of institutional research sector-wise for 20 leading countries and RoW

SIR 2018	Government	Health	Higher education	Private	Others	Total	Simple share	Government	Health	Higher education	Private	Others	Sum	L1 Share	Recursive share
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
USA	61	204	432	36	26	759	0.135	0.05	0.23	0.13	0.22	0.41	1.04	0.207	0.221
China	154	26	414	14	6	614	0.109	0.12	0.03	0.13	0.08	0.10	0.46	0.092	0.090
France	231	41	122	2	3	399	0.071	0.18	0.05	0.04	0.01	0.05	0.32	0.065	0.053
India	62	12	197	0	0	271	0.048	0.05	0.01	0.06	0.00	0.00	0.12	0.025	0.020
Spain	87	110	62	1	0	260	0.046	0.07	0.12	0.02	0.01	0.00	0.22	0.043	0.036
Germany	130	24	85	7	8	254	0.045	0.10	0.03	0.03	0.04	0.13	0.32	0.065	0.062
Japan	25	31	176	9	2	243	0.043	0.02	0.03	0.05	0.05	0.03	0.19	0.039	0.041
Russian federation	116	2	84	0	0	202	0.036	0.09	0.00	0.03	0.00	0.00	0.12	0.024	0.017
UK	14	62	114	5	2	197	0.035	0.01	0.07	0.04	0.03	0.03	0.18	0.035	0.036
Italy	56	48	68	1	2	175	0.031	0.04	0.05	0.02	0.01	0.03	0.16	0.031	0.028
Brazil	18	11	114	0	0	143	0.025	0.01	0.01	0.04	0.00	0.00	0.06	0.012	0.011
South Korea	28	19	86	2	0	135	0.024	0.02	0.02	0.03	0.01	0.00	0.08	0.016	0.015
Iran	8	3	115	0	0	126	0.022	0.01	0.00	0.04	0.00	0.00	0.05	0.009	0.008
Canada	10	43	53	0	3	109	0.019	0.01	0.05	0.02	0.00	0.05	0.12	0.024	0.022
Australia	10	48	41	1	3	103	0.018	0.01	0.05	0.01	0.01	0.05	0.13	0.025	0.025
Taiwan	5	22	67	1	2	97	0.017	0.00	0.02	0.02	0.01	0.03	0.09	0.017	0.017
Poland	30	5	58	0	0	93	0.017	0.02	0.01	0.02	0.00	0.00	0.05	0.009	0.007
Turkey	1	7	80	1	0	89	0.016	0.00	0.01	0.02	0.01	0.00	0.04	0.008	0.008
The Netherlands	15	13	17	0	0	45	0.008	0.01	0.01	0.01	0.00	0.00	0.03	0.006	0.005
Switzerland	6	12	15	3	1	37	0.007	0.00	0.01	0.00	0.02	0.02	0.06	0.011	0.013
L20	1067	743	2400	83	58	4351	0.773	0.84	0.82	0.74	0.50	0.92	3.83	0.765	0.735
RoW	204	159	830	83	5	1281	0.227	0.16	0.18	0.26	0.50	0.08	1.17	0.235	0.265
Total	1271	902	3230	166	63	5632	1	1	1	1	1	1	5	1.000	1.000

(an L1 distance). This allows us to compare the spread and style of research sector-wise between two countries using the concept of an L2 distance (here we actually use square of the L2 distance) between the row-wise fractionalized vectors. Since there are 21 rows (L20 + RoW), we can get 210 country-pairs for which the squares of the L2 distances can be computed. Table 2 shows an extract from 210 (i.e. $21 \times 20/2$) of the 10 closest and 10 most distant pairs. Figure 1 is a radar-diagram visualization of the closest (Italy–The Netherlands) and the most distant (Spain–Iran) pairs in their styles of distributing research across sectors. It is not surprising to see pairings such as the US–UK, China–India, Iran–Turkey and Germany–France in the list of most closely related styles of functioning. In the other list, we find Iran and Turkey appearing frequently. A simple explanation is that Iran and Turkey are similar to each other, and both are equally unlike others such as France, Germany and Spain.

Column-wise processing of matrix data

Table 3 is the same 21×5 matrix representation of the 5362 institutions in SIR 2018 in columns 2–6. However, subsequently, we will follow a column-wise processing to see what additional insights can be obtained¹⁰.

Column 7 in Table 3 remains the same as in Table 1, and is a simple row-wise total count of the institutions across sectors assuming that the same weight is given to

Table 4. Percentage change of share of institutions in research from row-wise to column-wise counting and from non-recursive to recursive counting for the 20 leading countries and RoW

SIR 2018	Percentage change row-wise to column-wise	Percentage change – non-recursive to recursive
1	2	3
USA	64.2	6.7
China	-17.6	-1.9
France	-25.3	-18.5
India	-58.1	-18.2
Spain	-21.8	-16.3
Germany	38.0	-4.1
Japan	-4.0	6.5
Russian federation	-51.4	-27.0
UK	2.0	0.9
Italy	-11.1	-11.5
Brazil	-58.4	-14.3
South Korea	-36.7	-7.3
Iran	-64.3	-11.6
Canada	16.0	-6.1
Australia	34.1	-3.8
Taiwan	-0.9	-1.7
Poland	-54.8	-20.8
Turkey	-50.5	-0.5
The Netherlands	-36.1	-18.9
Switzerland	90.3	10.4
L20	-4.8	-3.9
RoW	16.4	12.8

them. We have added a new column 8 so that we can see how the row-wise counts are shared between countries. The US seems to account for only a 13.7% share of institutions, although it is well known that nearly a quarter of the world’s publications comes from the US. In such simple counting, L20 accounts for 77% of the institutions that perform research, leaving only 23% share for RoW numbering more than 170 countries.

We now invoke a column-wise approach to get non-recursive and recursive shares that are based on uniform and differential weighting respectively⁹. Columns 9–13 in Table 3 normalize the counts country-wise in a column-wise fractional form for each sector. Thus, for each sector, we can identify a 21-dimensional column vector as the country-wise fractional shares. Across the 21 countries and 5 sectors we get a 21×5 matrix, which we shall call the **A** matrix. For each country we can identify a row: e.g. for the US this is now (0.05, 0.23, 0.13, 0.22, 0.41) and gives a total of 1.04 (an L1 distance), assuming that each sector has a unit weight. Column 14 in Table 3 gives these totals (we shall call these the non-recursive totals for reasons that will become clear soon) for L20 and RoW. For all 21 countries, these distances add up to exactly 5.0 as expected from conservation of count considerations⁹. We can now get a column-wise share for each country as shown in column 15 of Table 3. Comparing this with column 8 of the table we see that the US share has gone up to 20.7% (a change of 64.2%). Switzerland has recorded the highest increase (90.3%), while Iran (-64.3%), India (-58.1%) and Brazil (-58.4%) registered the largest decrease in shares when one uses column-wise operations instead of row-wise operations as the basis for counting.

Column 15 in Table 3 is based on a count of shares assuming each sector has a weight of unity, where **A** is a 21×5 vector in the country–sector space⁸. Using the **AA^T** operation^{8,9}, it is possible to transform these relationships from the country–sector space to a country–country space. Let us call this new matrix **O** ($= \mathbf{AA}^T$). This is now a 21×21 matrix where the totalized output is distributed as a symmetric matrix of terms which all add up to 5. It is a real symmetric matrix, and its eigenvalues are all real and its eigenvectors are mutually orthogonal. If we introduce a 21×1 vector of unit terms called **u** (i.e. each country is given an identical weight), then **Ou** gives a totalized output vector (21×1) which is exactly the same as that obtained earlier using the country–sector space considerations. As before, the terms add up to 5, as is to be expected, as there are 5 sectors. The matrix transformations from country–sector space to country–country space using stochastic matrices lead to a transformation matrix that is orthogonal and this ensures that the conservation law is upheld in the scheme proposed here⁹. At this stage (we call this the non-recursive stage), when each country is given an identical weight, we get share vectors which are the same as before (i.e. operating with the matrix from the country–sector space).

Note that the totalized output measures (or distances) we have found so far have assumed that each country is given an identical weight and this leads to a non-recursive measure. It is known from graph-theoretic procedures that the fundamental eigenvector of the \mathbf{O} matrix provides a more accurate and appropriate evaluation of the countries' output than an approach based on a simple addition of all the terms in the row¹⁰. Each country will be weighted according to its relative strength or weakness. It is possible to carry out this recursive iterated improvement using the eigenvalue relationship $\mathbf{O}\mathbf{s} = \lambda\mathbf{s}$, where \mathbf{s} is the recursively weighted improvement of the totalized country share score. Each country is then weighted recursively according to its standing and this leads to an eigenvalue calculation^{8,9} that produces a converged share vector shown in column 16 of Table 3. There are noticeable changes. RoW (12.8%) and Switzerland (10.4%) have gained the most when recursive weighting is used; the Russian Federation (−27%) and Poland (−20.8%) are the biggest losers. Table 4 summarizes the percentage change of share of institutions in research from row-wise to column-wise counting and from non-recursive to recursive counting for L20 and RoW.

Concluding remarks

Here we have evaluated how countries apportion their research sector-wise. We have used SIR 2018, where a five-sector classification of the leading 5362 research institutions is available. Countries have their unique styles of sharing and spreading their research efforts across sectors and it is seen that some countries are simi-

lar to others, while some are very different from the others. Various row-wise and column-wise operations allow us to quantify and visualize these styles and spreads, and measure their shares of the global institutional research efforts.

1. Whitehead, A. N., *Science and the Modern World*, The Macmillan Company, London, UK, 1925.
2. Edquist, C. and Johnson, B., Institutions and organizations in systems of innovation. In *Systems of Innovation: Technologies, Institutions and Organizations* (ed. Edquist, C.), Pinter/Cassell Academic, London, UK, 1997, p. 47.
3. North, D. C., *Structure and Change in Economic History*, Norton and Company Inc., New York, USA, 1981.
4. North, D. C., Institutions. *J. Econ. Perspectives*, 1991, **5**(1), 97–112.
5. May, R. M., The scientific wealth of nations. *Science*, 1997, **275**(5301), 793–796.
6. King, D. A., The scientific impact of nations. *Nature*, 2004, **430**(6997), 311–316.
7. Juraida, S., Kozubek, S., Munich, D. and Skoda, S., Scientific publication performance in post-communist countries: still lagging far behind. *Scientometrics*, 2017, **112**(1), 315–328.
8. Krauze, T. K. and McGinnis, R., A matrix analysis of scientific specialties and careers in science. *Scientometrics*, 1979, **1**(5–6), 419–444.
9. Prathap, G., Totalized input–output assessment of research productivity of nations using multi-dimensional input and output. *Scientometrics*, 2018, **115**(1), 577–583.
10. Ramanujacharyulu, C., Analysis of preferential experiments. *Psychometrika*, 1964, **29**(3), 257–261.

Received 16 October 2018; accepted 22 November 2018

doi: 10.18520/cs/v116/i4/530-535