



# Quantifying the publication preferences of leading research universities

Barbara S. Lancho-Barrantes<sup>1</sup> · Francisco J. Cantu-Ortiz<sup>2</sup>

Received: 10 August 2020 / Accepted: 10 November 2020 / Published online: 4 January 2021  
© Akadémiai Kiadó, Budapest, Hungary 2021

## Abstract

Research universities have a strong devotion and advocacy for research in their core academic mission. This is why they are widely recognized for their excellence in research which make them take the most renowned positions in the different worldwide university leagues. In order to examine the uniqueness of this group of universities we analyze the scientific production of a sample of them in a 5 year period of time. On the one hand, we analyze their preferences in research measured with the relative percentage of publications in the different subject areas, and on the other hand, we calculate the similarity between them in research preferences. In order to select a set of research universities, we studied the leading university rankings of Shanghai, QS, Leiden, and Times Higher Education (THE). Although the four rankings own well established and developed methodologies and hold great prestige, we choose to use THE because data were readily available for doing the study we had in mind. Having done that, we selected the twenty academic institutions ranked with the highest score in the last edition of THE World University Rankings 2020 and to contrast their impact, we also, we compared them with the twenty institutions with the lowest score in this ranking. At the same time, we extracted publication data from Scopus database for each university and we applied bibliometrics indicators from Elsevier's SciVal. We applied the statistical techniques cosine similarity and agglomerative hierarchical clustering analysis to examine and compare affinities in research preferences among them. Moreover, a cluster analysis through VOSviewer was done to classify the total scientific production in the four major fields (health sciences, physical sciences, life sciences and social sciences). As expected, the results showed that top universities have strong research profiles, becoming the leaders in the world in those areas and cosine similarity pointed out that some are more affine among them than others. The results provide clues for enhancing existing collaboration, defining and re-directing lines of research, and seeking for new partnerships to face the current pandemic to find was to tackle down the covid-19 outbreak.

**Keywords** Higher education institutions · Research universities · Research preferences · Scientific production · Cosine similarity · Hierarchical clustering

---

✉ Barbara S. Lancho-Barrantes  
b.s.lancho-barrantes@leeds.ac.uk

Extended author information available on the last page of the article

## Introduction

Higher education institutions (HEIs) differ in size, scope, history, mission and subject inclinations. This diversity manifests the extensive growth and development of this academic sector since the medieval period (European Commission 2019). Within these academic institutions are the research universities that emphasize the preeminence of high-level research.

Research universities are considered as the key higher education institutions that focus on advanced research as the core part of their missions with statements such as “research-intensive” or “research-based” (Mammadov and Aypay 2020; Taylor 2006).

These kind of universities are positioned on the higher education market sharing similar challenges, ambitions, incentives and motivations (Leydesdorff et al. 2019). In general, these institutions manifest a strong priority to discover new scientific knowledge in an extensive range of fields, to stand out with disruptive innovations, and to produce the future researchers through their wide-ranging PhD programs (Mohrman et al. 2008).

Research universities are a contemporary invention, having emerged in Prussia in the early nineteenth century, and in the United States only in the aftermath of the Civil War. Since then, they are contributing to the national economies as well as to local and regional economies (Atkinson and Blanpied (2008)).

The Carnegie Classification of Institutions of Higher Education, the framework for classifying colleges and universities in the United States, categorized 131 American research universities (public and private-non-profit) as “R1: Doctoral Universities—Very high research activity”: these universities have a very high level of research activity and per capita in such research activity. They are a small nucleus of highly productive “super research universities” (Carnegie Classification 2019; Carnegie Foundation 2001; Fernandez and Baker 2017).

The league of European research universities (LERU) is an established of consortium research-intensive universities across Europe. The League represents 23 leading universities pushing the frontiers of innovative research and sharing the values of high-quality teaching within an environment of internationally competitive research. According to LERU research universities are bringing cutting-edge knowledge and becoming the ultimate source of most innovation in the economy, society and culture of countries. They are providing a framework for an education through the skepticism, creativity and high-level capability (LERU 2019).

Europe has the oldest and leading research universities worldwide, such as the Paris-Sorbonne University (founded 1150), University of Heidelberg (1386) or Catholic University Leuven (1425) that produce large numbers of publications and are globally interconnected (Powell and Dusdal 2017a).

Furthermore, the Russell Group in the United Kingdom (UK) represents 24 leading research-intensive universities. These world-class universities play an important part in the intellectual and academic life of the UK and perform a huge social, economic and cultural impact across the country, and around the globe (Russell Group 2020). Some of these research universities have been recognized as the world’s most powerful and productive universities according to the number of publications per year. In fact some are very specialized in certain areas becoming the world leaders such as the strong scientific grounding of the University of Oxford in medical studies.

Powell and Dusdal (2017a, b) examined the research system of three countries, Germany, France and the United Kingdom, to find out that made them leaders in organizational

development of research and scientific innovation. The differences in productivity in these three countries could not be completely explained by the differences in global economic investments or the number of researchers committed to science. If not, it is a question of institutionalization and distribution organizational in which researchers produce science. The key findings of their study were that the institutionalization of the research universities in these countries support high levels of national scientific productivity.

The Times Higher Education-World University Rankings (THE) is one of the leading global performance approaches that judges research-intensive universities across all of their core missions- following a set of five criteria in order to be included in the Overall Ranking. In THE rankings two criteria that have to do explicitly with the concept of research universities are: (1) Sufficient publications—An institution is required to publish more than 1000 papers over the previous 5 years, and more than 150 publications in any single year. Thresholds are also applied per subject for the subject rankings. (2) Subject breadth—An institution must not be focused on a single narrow subject area (more than 80% of their publication output is from one subject area) (THE-World University Rankings 2020) Besides this ranking there are other rankings of institutions such as the Academic Ranking of World Universities also known as Shanghai Ranking, originally compiled and issued by Shanghai Jiao Tong University in 2003 with an emphasis on Nobel Prize laureates and ground-breaking publications, the QS World University Rankings which is compiled using six simple metrics that capture university performance (Academic Reputation, Employer Reputation, Faculty/Student Ratio, Citations per faculty, International Faculty Ratio, International Student Ratio) (QS World University Rankings 2020), the SCImago Institutions Rankings (SIR) is a classification of academic and research-related institutions ranked by a composite indicator that combines three different sets of indicators based on research performance, innovation outputs and societal impact measured by their web visibility (SCImago Institutions Rankings (SIR), 2020), the CWTS Leiden Ranking 2020 offers a sophisticated set of bibliometric indicators that provide statistics at the level of universities on scientific impact, collaboration, open access publishing, and gender diversity (CWTS Leiden Ranking 2020). After studying these ranking methodologies, we choose THE because of the comprehensive nature of its indicators and because we had data available from the THE WUR, the Scopus database, and Scival (SciVal 2019; Scopus 2019a).

Several studies have examined research productivity of universities across scientific literature (Toutkoushian et al. 2003; Butler 2007; Chen et al. 2010; Cloete et al. 2011; Huang 2012; Jung 2012; McGill and Settle 2012; Abramo and D'Angelo 2014; Altbach 2015; Musiige and Maassen 2015; Budd 2017; Webber 2011; Bolli et al. 2016; Hu et al. 2017). Van Raan (2008) observed that the top research universities are about twice as efficient in production and citations as compared to the bottom performance universities based on the field normalized indicator.

Following this brief introduction to the framework of research universities and considering that we have not found studies that analyze the publication profiles of research universities, we ask ourselves a series of research questions: Q1: What could be making these universities to stand out scientifically? Q2: Are their research preferences different from those of other universities? Q3: Are research universities similar in research preferences? Q4: What are the most commonly used keywords in their production? Q5: Are research universities collaborating with each other?

Our main research hypothesis is that top-ranked research universities excel in certain areas becoming world leaders in these fields. In relation to this, research universities are similar in research preferences. However, there are some universities more analogous to

each other. We also study the question of how top universities are influencing the rest of the universities in the ranking, particularly, those in the lower positions.

## Data and method

We used the last edition of times higher education (THE) Ranking World University rankings 2020 to select a sample of research universities. Founded in 2004, the times higher education (THE) World University Rankings is one of the most globally recognized university ranking systems; they support students in decision making for studying, early career researchers deciding next steps for hosting institutions, research agencies that allocate research budgets, policymakers planning and checking scientific results for funding programs, etc. The ranking encompasses a total of 1400 institutions in its last edition 2020 (THE 2020).

We chose the first 20 leading research universities ordered by the general score, these are as follows (in descending order)

	Institution	Country
1	University of Oxford	United Kingdom
2	California institute of technology	United States
3	University of Cambridge	United Kingdom
4	Stanford University	United States
5	Massachusetts institute of technology (MIT)	United States
6	Princeton University	United States
7	Harvard University	United States
8	Yale University	United States
9	University of Chicago	United States
10	Imperial College London	United Kingdom
11	University of Pennsylvania	United States
12	Johns Hopkins University	United States
13	University of California, Berkeley	United States
14	ETH Zurich (Swiss Federal Institute of Technology in Zurich)	Switzerland
15	University College London (UCL)	United Kingdom
16	Columbia University	United States
17	University of California, Los Angeles	United States
18	University of Toronto	Canada
19	Cornell University	United States
20	Duke University	United States

The American institutions on the list: Harvard University, California Institute of Technology, Stanford University, Massachusetts Institute of Technology, Princeton University, Harvard University, Yale University, University of Chicago, University of Pennsylvania, Johns Hopkins University, University of California, Berkeley, Columbia University, University of California, Los Angeles, Cornell University, Duke University belong to the section of R1: Doctoral Universities—very high research activity in the Carnegie Classification. Furthermore, the British universities University of Oxford, University of Cambridge, Imperial College London, University College London belong

to the Russell Group and LERU and also form the Golden Triangle. The Golden Triangle is a term that was initially used to describe the grouping of elite, highly-funded universities located in the southern English cities of Oxford, Cambridge, and London. These universities are prominent in the world of education, research and innovation both in the UK and on an international scale. The institutions that form the Golden Triangle receive some of the highest research incomes, funding and grants from the UK government, European and global agencies, and the largest financial endowments of all British universities.

For the purpose of comparing publication profiles between universities we decided to select the twenty universities that have the lowest scores:

	Institution	Country
1001 +	Warsaw University of Technology	Poland
1001 +	Wenzhou Medical University	China
1001 +	University of West Bohemia	Czech Republic
1001 +	Western Paraná State University (Unioeste)	Brazil
1001 +	West University of Timișoara	Romania
1001 +	Wrocław University of Science and Technology	Poland
1001 +	Yamagata University	Japan
1001 +	Yamaguchi University	Japan
1001 +	University of Yamanashi	Japan
1001 +	Yanshan University	China
1001 +	Yarmouk University	Jordan
1001 +	Yazd University	Iran
1001 +	Yeditepe University	Turkey
1001 +	Yıldız Technical University	Turkey
1001 +	Yokohama National University	Japan
1001 +	Yuan Ze University	Taiwan
1001 +	Zagazig University	Egypt
1001 +	University of Zagreb	Croatia
1001 +	University of Zanjan	Iran
1001 +	Zhejiang University of Technology	China

As we can observe they belong to China, Japan, Turkey, etc. The majority of them are based in Asia. There are at least 7 of 20 outside Asia, and 5 (25%) based in Europe. Just because they appear in the last ranking positions does not mean they do not have quality indicators. The lower ranking universities can be better choice for some students. Even though the academic position on the ranking can be essential to decide where to study, students consider other things when choosing their university: financial aspect, employment rate after graduation, professional training, learning environment or in general student life satisfaction (University World News 2013).

The Scopus data source has been used to extract the scientific outputs of the institutions. Scopus database remains the exclusive data source powering the flagship THE rankings. Scopus is the largest abstract and citation database of peer-reviewed research literature including over 24,000 titles, including 4200 Open Access journals from more than 5000 international publishers. This covers journals included in WOS, and its coverage is statistically balanced in terms of topics, countries, languages and publishers (de Moya-Anegon et al. 2007). In addition, SciVal has also been used to extract other metrics. SciVal is a

research evaluation tool based on Scopus data which allows us to visualize research performance, benchmark relative to peers, develop strategic partnerships, identify and analyze new, emerging research topics, and find out collaborators, etc. All data were downloaded cutoff date in December 2019.

We extracted the publications produced per each institution in a 5-year publication window from 2014 to 2018 (the same publication window used in THE 2020). The total number of documents downloaded were 1,160,626 (all type of documents is included). At the same time, we classified each title in its scientific area basing us on the ASJC (All Science Journal Classification) scheme in Scopus. Therefore, we count the number of publications that universities have in the different subject areas. There are over 300 sub-subject areas in Scopus, which aggregate up to 27 broad subject areas. The table below shows the broad subject area classification (Scopus 2019b):

Physical sciences	Chemical engineering, Chemistry, Computer science, Earth and Planetary sciences, Energy, Engineering, environmental science, Material science, Mathematics, Physics and Astronomy, Multidisciplinary
Health sciences	Medicine, Nursing, Veterinary, Dentistry, Health professions, Multidisciplinary
Social sciences	Arts and humanities, Business, Management and accounting, Decision sciences, Economics, Econometrics and finance, Psychology, Social sciences, Multidisciplinary
Life sciences	Agricultural and biological sciences, Biochemistry, Genetics and molecular biology, Immunology and Microbiology, Neuroscience, Pharmacology, Toxicology and pharmaceuticals, Multidisciplinary

The VOSviewer software tool was used for creating the term maps. VOSviewer was developed at Leiden University's Centre for Science and Technology Studies (CWTS) and provides techniques for constructing and visualizing bibliometric networks. These networks may for instance include journals, researchers, or individual publications, and they can be constructed based on citation, bibliographic coupling, co-citation, or co-authorship relations. VOSviewer also offers text mining functionality that can be used to construct and visualize co-occurrence networks of important terms extracted from a body of scientific literature. (VOSviewer 2020).

The data gathered was processed using the XLSTAT statistical program.

Needless to mention that scientific collaboration is an implicit and fundamental aspect of research universities considering that they participate in projects funded by international research agencies involving a large number of collaborating partners such as the EU-funded projects which involve collaboration between organizations from different EU countries or associated countries. For this reason, we considered to include an analysis of scientific collaboration in this study.

The Times Higher Education World University Rankings 2020 calibrated performance indicators to provide the most comprehensive and balanced comparisons, trusted by students, academics, university leaders, industry and governments. The performance indicators are grouped into five areas: Teaching (the learning environment); Research (volume, income and reputation); Citations (research influence); International outlook (staff, students and research); and Industry Income (knowledge transfer).

The variables relevant to this study are:

- Research (volume, income and reputation): 30%
- Reputation survey: 18%

- Research income: 6%
- Research productivity: 6%

The most prominent indicator in this category looks at a university's reputation for research excellence among its peers, based on the responses to our annual Academic Reputation Survey.

Research income is scaled against academic staff numbers and adjusted for purchasing-power parity (PPP). This indicator is fully normalised to take account of each university's distinct subject profile, reflecting the fact that research grants in science subjects are often bigger than those awarded for the highest-quality social science, arts and humanities research.

To measure productivity, they count the number of publications published in the academic journals indexed by Elsevier's Scopus database per scholar scaled for institutional size and normalized for subject.

- Citations (research influence): 30%

This indicator examines the research influence capturing the average number of citations received by the university's published work. The data include all indexed publications between 2014 and 2018 and the citations made in the 6 years from 2014 to 2019 to these publications. The data are normalized to reflect variations in citation volume between different subject areas. This means that institutions with high levels of research activity in subjects with traditionally high citation counts do not gain an unfair advantage.

Variables extracted from Scopus and SciVal:

- Scientific production: Total documents produced by a research university during the study period from 2014 to 2018. All types of documents are considered (articles, conference proceedings, reviews, books, book chapters...). The scientific production of a research university is counted as the documents with at least one author from that research university (according to the affiliation that appears in the credits of authorship).
- Scientific production in the different subject areas: Publications produced by these universities in the 27 subject areas and 300 sub-subject areas.
- Citations: Number of citations to the set of publications from 2014 to 2018.
- Citations per Publication: It is calculated by dividing the total number of citations by the total number of publications.
- Collaboration %: Percentage of publications made in collaboration: international, national, institutional or single.
- Collaboration Impact: It is the citations obtained by publications made with international, national or institutional collaborators.
- Field-weighted citation impact (FWCI) is the ratio of the total citations received by the denominator's output, and the total citations that would be expected based on the average of the subject field.

A FWCI of:

- Exactly 1 means that the output performs just as expected for the global average.
- More than 1 means that the output is more cited than expected according to the global average. For example, 1.23 means 23% more cited than expected.
- Less than 1 means that the output is cited less than expected according to the global average.
- Outputs in top citation percentiles: This metric shows how many publications of each research universities are in the top 1%, 5%, 10% or 25% most cited documents globally.
- Publications in top journal percentiles: This indicator calculates how many publications of each research university are in the top 1%, 5%, 10% or 25% most-cited journals.
- Academic-corporate collaboration: Indicates the degree of collaboration between research universities and corporate affiliations or industrial sectors.
- Academic-corporate collaborationImpact: Calculates the average citations received by the outputs that have been co-authored by researchers from both academic and corporate affiliations.

Analysis of scientific collaboration between research universities:

- We have calculated the scientific collaboration between the twenty research universities in that period of time. For example: how many publications do Cambridge and Oxford have in common? At the same time, we have calculated the percentage of collaboration with other universities not included in the sample.

### Measuring through cosine similarity and agglomerative hierarchical clustering the research preferences of universities.

We have applied cosine similarity and agglomerative hierarchical clustering to analyze the research similarities among research universities.

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.

This is the mathematical definition of the cosine similarity:

$$A \cdot B = \|A\| \|B\| \cos\theta$$

Cosine similarity:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Because this analysis compares the research preferences in subject areas instead of the absolute research output of universities consequently 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. A cosine value of 0 means that the two vectors are at 90° to each other (orthogonal) and have no



match. The closer the cosine value to 1, the smaller the angle and the greater the match between vectors. (Singhal 2001; Zhao and Suzuki 2015; Zhigang et al. 2017; Lin et al. 2018).

The publications of each research university in several subject areas were counted, and the disciplinary distribution vector of every research university was constructed to calculate the degree of discipline similarity between any two research universities based on cosine similarity formula. After calculating the similarity of the research universities, we applied the method of Agglomerative Hierarchical Clustering.

## Agglomerative hierarchical clustering

Cluster analysis allows to identify the natural structure of a dataset and also to categorize the objects into different clusters where objects within a same cluster are quite similar to each other (Arbelaitz et al. 2013). Hierarchical clustering (Zhang et al. 1996) shows a hierarchical structure of the input data set instead of only one partition. It successively seeks to build a hierarchy of clusters (presented in a dendrogram) via two types of strategies: Agglomerative hierarchical clustering algorithm or AGNES (agglomerative nesting) (bottom-up) and Divisive Hierarchical clustering algorithm or DIANA (divisive analysis) (top down).

Such categorization is inspired from the mechanism of grouping the objects whether bottom up or top-down approach. AHC is considered as a bottom-up hierarchical approach where each object is set in a separated cluster (Bouguettaya et al. 2015) then AHC will merge such clusters into larger clusters. The process continues until a specific termination has been reached. A complete linkage algorithm aims to identify the similarity between two clusters by measuring two nearest data points that are located in different clusters. Hence, the merge will be done between the clusters that have a minimum distance (most similar) between each other (Barirani et al. 2013; Sammour 2019).

The text mining function of VOSviewer (van Eck and Waltman 2011) was chosen to create a map based on the text used in the title and abstract fields of research universities publications. We chose fractionalization and binary counting and selecting 5 as the minimum number of documents in which a term is required to occur in order to remain in the analysis. The number of terms that meet the threshold was 2500. Terms that do not occur in a sufficiently large number of documents have now been removed from the analysis. Among the remaining terms, we proposed to keep 60% of the terms that should remain in the analysis. VOSviewer removed from the analysis the terms that appear to be least relevant (van Eck and Waltman 2017; Waltman and van Eck 2013; Waltman et al. 2010). At the same time, we use a thesaurus file to remove non-informative terms and to merge terms that are synonyms. A thesaurus file is a plain text file consisting of two tab-delimited columns.

## Results

### Overview scientific performance research universities

In Table 1 we can observe indicators to measure research performance of the top 20 research universities. We extracted indicators from THE 2020 (overall, research, citations) and we applied metrics from SciVal (scholarly output, citation count, collaboration, etc.)

The three universities with the highest THE general score are University of Oxford, California Institute of Technology and University of Cambridge. According to the THE

**Table 1** Indicators applied to top 20 research universities

Overall score	Countries	Overall	Research	Citations	Scholarly output	Citation count	Citations per publication	Collaboration (%)	Collaboration impact (%)	Field-weighted citation impact	Output in top 10% citation percentiles (%)	Publications in top 10% journal percentiles	Academic-corporate collaboration impact	Academic-corporate collaboration (%)
	United Kingdom	95.4	99.6	98.4	69,463	12,17,856	17.5	58	22.5	2.35	26.3	47.8	5.8	42.1
	Oxford													
	California institute of technology	94.5	97.2	97.9	22,797	4,60,835	20.2	51.6	25.8	2.22	32.7	46.5	5.4	30
	United Kingdom	94.4	98.7	95.8	55,572	9,57,157	17.2	59.2	22.1	2.18	27.7	49.1	6.7	33.5
	Cambridge													
	Stanford university	94.3	96.4	99.9	68,916	1,4,14,899	20.5	40.9	27.4	2.66	29.5	50.6	7.2	40.7
	Massachusetts institute of technology	93.6	92.4	99.5	46,260	9,55,252	20.6	48.5	23.9	2.44	31	53.8	8.7	27.1
	Princeton university	93.2	96.3	98.8	22,881	4,10,205	17.9	45.7	22.8	2.33	28.2	49.3	6.9	21.6
	Harvard university	93	98.6	99.1	1,49,867	28,43,045	19	43.4	24.7	2.36	28.8	48.8	5.7	45.1
	Yale university	91.7	94.8	97.3	45,240	7,81,048	17.3	38.2	25	2.2	27.2	48.4	4.6	57.4

**Table 1** (continued)

Overall score	Countries	Overall	Research	Citations	Scholarly output	Citation count	Citations per publication	Collaboration (%)	Collaboration impact (%)	Field-weighted citation impact	Output in top 10% citation percentiles (%)	Publications in top 10% journal percentiles	Academic-corporate collaboration impact (%)	Academic-corporate collaboration impact
University of Chicago	United States	90.2	91.4	96.7	44,049	7,83,696	17.8	40.5	25.7	2.15	27.2	49.7	4.7	43.9
Imperial college London	United Kingdom	89.8	87.6	97	58,853	9,81,155	16.7	59.7	21.1	2.24	26.7	48.1	7.7	32.9
University of Pennsylvania	United States	89.6	90.4	98.2	59,108	9,53,214	16.1	31.5	24.5	2.2	25.2	45.2	5.3	48.1
Johns Hopkins university	United States	89.2	91.4	98.3	65,895	11,34,765	17.2	39.2	23.9	2.26	25.9	43.8	5.8	46.5
ETH Zurich	Switzerland	88.3	92.8	90.3	39,205	6,27,946	16	65	18.3	2.05	28.1	52	7.5	19.4
University of California, Berkeley	United States	88.3	90.6	99.2	45,254	8,70,875	19.2	44.6	23.8	2.38	29.7	51	5.8	32.1
UCL	United Kingdom	87.1	88.7	96.1	75,038	11,72,831	15.6	55.7	20.4	2.17	25	44.2	5.6	37.2

Table 1 (continued)

Overall score	Countries	Overall	Research	Citations	Scholarly output	Citation count	Citations per publication	Collaboration (%)	Collaboration impact (%)	Field-weighted citation impact	Output in top 10% citation percentiles (%)	Publications in top 10% journal percentiles	Academic-corporate collaboration impact (%)
	Columbia University	87	82.6	98.2	55,607	9,38,345	16.9	40.1	24.2	2.25	25.2	47	41.1
	United States												
	University of California, Los Angeles	86.8	88.6	97.3	59,847	9,63,022	16.1	39	22.5	2.15	24.8	46.1	37.5
	United States												
	University of Toronto	85.5	89.5	93.6	87,177	13,12,741	15.1	50.4	21.4	2.05	22.2	40.4	44.8
	Canada												
	Cornell University	85.1	86	96.6	45,008	7,42,256	16.5	38.8	2.3	2.1	25.4	46.6	40.8
	United States												
	Duke University	84	76.8	97	44,589	7,29,460	16.4	36.6	24	2.19	25	47	39.6
	United States												

indicators we can confirm that University of Oxford had the highest score in Research (99.6) and Stanford university excelled others in the Citations metric (99.9). Regarding to SciVal metrics Harvard university produced the largest number of publications (149,867) and also obtained the largest amount of citations (2,843,045). Massachusetts Institute of Technology (MIT) achieved the highest average of citations per publication (20.6).

ETH Zurich was the most collaborative university with a 65% of its papers in collaboration. Stanford university succeeded in collaboration impact (27.4) and in the FWCI (2.66). California Institute of Technology published the highest percentage in output in top 10% citation percentiles (32.7). It is maybe worthwhile to mention that the Jet Propulsion Laboratory, a federal owned lab funded by NASA, is operated as a division of California Institute of Technology partially it may explain the high position of the university in metrics. Massachusetts Institute of Technology (MIT) had the highest percentage of publications in top 10% journal percentiles (53.8).

Regarding to Academic-Corporate Collaboration and Academic-Corporate Collaboration Impact, Massachusetts Institute of Technology (8.7) and Yale University (57.4) highlighted in these indicators from others.

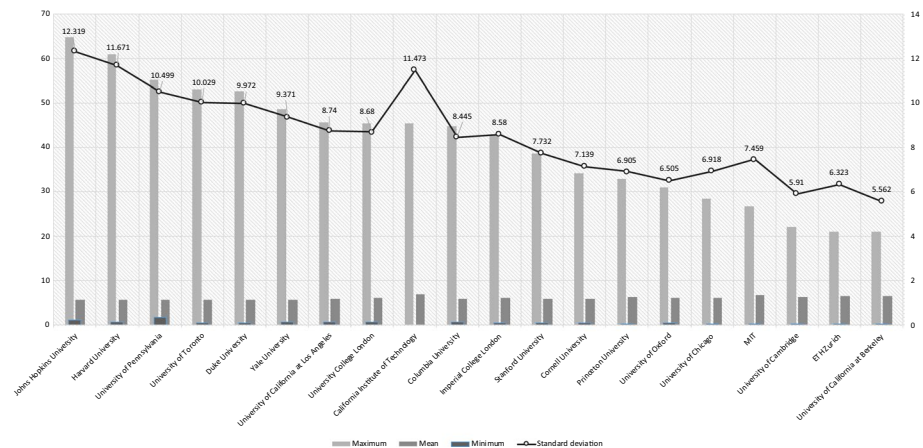
As we can appreciate according to the metrics, they all share high levels of productivity, collaboration and impact levels. All of them have high level of publications in top 10% journals, and a FWCI higher than 2 indicating that their publications have been cited twice more than would be expected based on the world average for similar publications.

The following Table 2 shows the Spearman correlation among all the indicators from the research universities.

The overall score present high statistical correlations with Research (0.883), Citations per publication (0.719), FWCI (0.559), and Output in Top 10% Citation Percentiles (0.712).

Research is highly correlated with Output in Top 10% Citation Percentiles (0.666) and with Citations per Publication (0.586). Thus, it is not surprising that Citations metric is related with Citations per publication (0.764), Collaboration Impact (0.530), FWCI (0.937) and Output in Top 10% Citation Percentiles (0.511).

The results seem to indicate that the main research indicators in the ranking have high correlation with Scival impact indicators which leads to the conclusion that high levels of impact cause universities to excel in positions in the ranking compared to others.



**Fig. 1** Descriptive statistics minimum (Min), maximum (Max), mean, and standard deviation (SD) applied to the research preferences

**Table 2** Correlation matrix (Spearman) of indicators applied to research universities

Variables	Overall	Research	Citations	Scholarly output	Citation count	Citations per publication	Collaboration (%)	Collaboration impact (%)	Field-weighted citation impact	Output in top 10% citation percentiles (%)	Publications in top 10% journal percentiles	Academic-corporate collaboration rate (%)	Academic-corporate collaboration impact
Overall	<b>1</b>	0.883	0.451	-0.043	0.117	0.719	0.404	0.303	0.559	0.712	0.433	0.178	-0.129
Research	0.883	<b>1</b>	0.322	0.019	0.141	0.586	0.408	0.199	0.390	0.666	0.410	0.047	-0.058
Citations	0.451	0.322	<b>1</b>	0.154	0.196	0.764	-0.221	0.530	0.937	0.511	0.322	0.249	0.031
Scholarly output	-0.043	0.019	0.154	<b>1</b>	0.959	-0.212	0.047	-0.169	0.181	-0.384	-0.396	-0.108	0.489
Citation count	0.117	0.141	0.196	0.959	<b>1</b>	-0.042	0.162	-0.113	0.260	-0.211	-0.228	-0.002	0.408
Citations per publication	0.719	0.586	0.764	-0.212	-0.042	<b>1</b>	0.048	0.611	0.790	0.851	0.602	0.203	-0.169
Collaboration	0.404	0.408	-0.221	0.047	0.162	0.048	<b>1</b>	-0.583	0.011	0.336	0.284	0.377	-0.597
Collaboration impact (%)	0.303	0.199	0.530	-0.169	-0.113	0.611	-0.583	<b>1</b>	0.418	0.367	0.126	-0.310	0.405
Field-weighted citation impact (%)	0.559	0.390	0.937	0.181	0.260	0.790	0.011	0.418	<b>1</b>	0.588	0.396	0.363	-0.062
Output in top 10% citation percentiles (%)	0.712	0.666	0.511	-0.384	-0.211	0.851	0.336	0.367	0.588	<b>1</b>	0.747	0.310	-0.443

**Table 2** (continued)

Variables	Overall	Research	Citations	Scholarly	Citation	Citations	Collabo-	Collabo-	Field-	Output	Publica-	Academic-corporate	
				output	count	per pub-	ration	ration	weighted	in top	tions in	col-	
						lication	(%)	(%)	citation	10%	top 10%	laboration	
							impact	impact	impact	citation	journal	impact	
							(%)	(%)	percent-	percent-	percen-	laboration	
									titles (%)	titles (%)	titles	impact	
Publications in top 10% journal percentiles	0.433	0.410	0.322	-0.396	-0.228	0.602	0.126	0.284	0.396	0.747	<b>1</b>	0.533	-0.472
Academic-corporate collaboration (%)	0.178	0.047	0.249	-0.108	-0.002	0.203	-0.310	0.377	0.363	0.310	0.533	<b>1</b>	-0.672
Academic-corporate collaboration impact	-0.129	-0.058	0.031	0.489	0.408	-0.169	0.405	-0.597	-0.062	-0.443	-0.472	-0.672	<b>1</b>

**Table 3** Research preferences in publications of the Top 20 Research Universities

	Cali- for- nia institute of tech- nology	Colum- bia univer- sity	Cornell univer- sity	14.13	6.68	4.29	5.41	3.72	3.46	5.59	4.53	8.59	4.66	9.29	5.41	7.16	4.32	7.41	3.92	4.55	5.53
	4.74	4.54	5.04	3.35	3.16	2.64	0.58	2.13	2.02	7.67	2.93	1.34	5.56	5.06	4.03	8.53	4.37	9.05	3.86	4.42	5.03
	7.92	13.65	20.05	17.35	20.24	15.22	18.55	13.27	8.72	16.61	13.47	13.53	16.68	18.32	14.90	16.84	18.17	15.884	18.99		
Agricul- tural and biologi- cal sci- ences																					
Arts and humani- ties	0.65	5.04	3.35	3.16	2.64	0.58	2.13	2.02	7.67	2.93	1.34	5.56	5.06	4.03	8.53	4.37	9.05	3.86	4.42	5.03	
Biochem- istry, genet- ics and molecu- lar biol- ogy	7.92	13.65	20.05	17.35	20.24	15.22	18.55	13.27	8.72	16.61	13.47	13.53	16.68	18.32	14.90	16.84	18.17	15.884	18.99		
Business, manage- ment and account- ing	0.24	1.81	2.42	1.32	1.18	0.97	0.61	2.10	1.43	1.50	1.68	1.38	2.13	1.03	2.05	1.28	1.88	2.06	1.37	1.08	
Chemical engi- neering	3.99	1.89	292	1.46	1.57	4.20	1.49	5.99	4.20	3.22	6.15	2.40	4.75	2.30	3.81	4.07	2.58	1.69	2.03	1.92	
Chemistry	8.79	3.44	5.41	3.14	2.80	6.91	2.78	9.69	7.99	5.88	12.57	4.68	9.08	4.19	7.92	10.02	6.17	3.71	3.81	4.22	



**Table 3** (continued)

California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins University	MIT	Stanford university	Swiss Federal Institute of Technology Zurich	University of California Berkeley	University of Chicago	University of Pennsylvania	University of Toronto	Yale University					
12.11	6.23	7.53	5.32	2.87	9.60	4.05	19.68	13.22	9.60	16.46	7.04	13.02	6.70	7.07	6.10	7.31	5.19	5.72	3.23
0.45	1.16	1.02	0.96	0.45	0.84	0.50	1.81	1.25	0.95	1.44	0.65	1.48	0.76	1.04	0.96	0.82	0.92	0.72	0.56
0.00	0.54	0.12	0.11	0.61	0.12	0.22	0.03	0.01	0.14	0.07	0.88	0.03	0.74	0.04	0.12	0.12	0.73	0.54	0.15
39.57	6.73	3.86	1.84	3.93	4.72	2.91	8.41	12.50	5.14	13.09	4.66	10.53	6.52	8.53	4.57	5.76	1.07	3.47	4.05
0.76	2.53	2.54	1.90	1.52	0.80	0.77	2.30	3.22	1.94	2.08	1.53	2.91	1.15	2.25	2.41	2.75	2.10	1.15	2.00
2050	0.95	1.35	0.63	0.32	3.52	0.48	4.89	3.58	2.28	4.25	1.41	3.45	1.04	2.26	4.99	1.47	0.35	1.00	0.73
22.27	7.21	8.33	6.73	4.28	15.69	5.14	26.07	14.09	10.94	21.09	8.07	17.07	8.70	12.10	9.43	7.49	5.28	7.57	4.23
5.14	4.07	5.45	4.17	1.86	4.69	1.71	4.11	5.28	3.44	9.47	3.09	7.71	3.51	4.31	3.16	4.36	1.12	2.85	3.41
0.0	1.23	0.78	1.86	1.66	1.10	1.55	0.43	0.21	1.20	0.65	1.46	0.39	1.10	0.72	0.54	0.78	1.39	2.40	1.12

**Table 3** (continued)

California institute of technology	2.78	4.80	4.21	4.75	5.02	5.38	3.14	2.05	3.53	2.64	3.73	2.95	3.27	3.47	2.92	4.80	4.64	3.13	4.17
Immunology and microbiology	1.53	11.71	3.91	3.17	2.78	8.01	2.90	9.55	7.13	12.05	4.75	9.90	5.20	8.62	11.03	5.34	3.25	3.95	3.00
Materials science	4.46	4.31	4.03	4.03	2.19	6.66	2.69	10.79	5.55	10.61	3.86	8.37	4.45	5.91	5.35	6.43	8.43	3.20	2.86
Mathematics	9.01	4.46	4.46	4.46	2.19	6.66	2.69	10.79	5.55	10.61	3.86	8.37	4.45	5.91	5.35	6.43	8.43	3.20	2.86
Medicine	2.55	44.75	34.19	52.60	60.84	42.72	64.69	8.46	4.66	38.57	8.06	45.49	10.64	45.60	22.07	30.97	55.22	53.12	48.56
Multidisciplinary	4.25	2.66	2.74	2.41	3.18	2.29	2.18	4.61	3.93	3.76	3.00	2.44	4.30	2.55	3.69	3.28	2.08	1.51	3.15
Neuroscience	2.00	6.62	3.83	5.54	6.88	3.19	7.10	3.36	2.85	5.27	2.87	9.60	3.54	6.69	5.27	5.46	6.26	6.81	7.23
Nursing	0.07	2.43	1.87	3.74	2.87	1.84	3.64	0.13	0.12	1.33	0.51	1.78	0.54	2.17	1.20	1.63	3.54	2.74	2.76
Pharmacology and pharmaceuticals	0.47	2.50	1.66	3.00	2.39	2.20	3.14	1.22	0.34	1.48	1.56	2.59	0.84	2.01	1.70	1.80	2.70	2.47	3.05

**Table 3** (continued)

California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins university	MIT	Stanford university	Swiss Federal Institute of Technology Zurich	University of California Berkeley	University of Chicago	University of Cambridge	University of Oxford	University of Pennsylvania	University of Toronto	Yale university								
45.37	8.94	10.68	7.73	7.97	13.84	7.43	26.62	32.89	15.19	20.26	10.01	20.92	10.74	19.06	24.75	13.24	6.20	7.01	9.26					
Physics and astronomy	Psychology	Social sciences	Veterinary	0.41	6.05	3.06	4.27	3.75	0.97	3.37	1.10	2.14	3.28	1.19	5.15	3.32	5.05	2.93	3.20	3.40	4.60	4.68	6.38	9.26
0.13	2.17	0.23	0.15	0.22	0.34	0.16	0.10	0.12	0.26	0.15	0.11	0.13	0.61	0.07	0.27	1.49	0.12	0.17						

## Research preferences at top 20 research Universities

On average, the areas where research universities publish the most are Medicine, Physics and Astronomy and Engineering. However, there are some which have a strong and defined research profile such as California Institute of Technology in Physics and Earth and Planetary Sciences. Harvard university, Johns Hopkins University and University of Pennsylvania in Medicine. And MIT in Physics and Engineering among others (Fig. 1).

The smallest data value is 0.03 in the area of Veterinary, the maximum 64.68 in Medicine. On average, the research universities contribute to the different areas a 6.113 on average, and the standard deviation of the data is around 8.511. As shown in the above Table 3, some universities have higher standard deviation than others, for example, Johns Hopkins University, Harvard University and California Institute of Technology. These universities have the defined publication profile in the field of Medicine and Physics and Astronomy. University of California at Berkeley is the institution with the lowest standard deviation and is one of the universities whose contribution to the disciplines is more distributed.

## Cosine similarity

The cosine distances between two research universities were calculated based on the data in Table 3. Table 4 shows the research universities with the most and the least similarity in the publication in subject areas. Princeton University and University of California at Berkeley, both important American universities in the world with similar economic status and structure, are the closest in the vector space of research subjects (0.906).

The next closest pair is Princeton University and University of Cambridge, two prominent research universities located in the different countries (0.902). The third pair is Harvard University and University College London (0.895). Although the absolute count of publications varies between them both have similar disciplinary structure and research preferences. The most dissimilar research universities pair is Imperial College London and Stanford University (0.675).

As shown in Table 5, the resulting adjacency matrix represents the level of similarity between two research universities in sub-subject areas. For example, the cosine similarity between massachusetts institute of technology (MIT) and ETH Zurich is 0.901 whereas that between MIT and University of Pennsylvania is 0.728. It means MIT is more similar to ETH Zurich than University of Pennsylvania in the contribution in publication to the sub-subject areas.

## Scientific similarity among 20 leading universities and lowest one in THE ranking

We created the next Table to find out if there are research similarities between the first twenty leading research areas and the last twenty research universities in THE.

The cosine similarity matrix between the world's top and lowest-ranked research universities were calculated based on the data of publications. As shown in Table 6, the resulting adjacency matrix represents the level of similarity between two universities. For example, the cosine similarity between Universidade Estadual do Oeste do Paraná and Columbia University is 0.872 whereas that between Universidade Estadual do Oeste do Paraná and Cornell University is 0.659. It means Universidade Estadual do Oeste do Paraná is more

similar to Columbia University than Cornell University in the distribution of research subjects. Zhejiang University of Technology and Massachusetts Institute of Technology are very similar (0.860).

Based on the data that we can see in Table 6, the top and lowest universities are similar in publication patterns. No surprisingly, these results show us that top research universities are influencing and being followed by the rest of the universities ranked by THE, particularly those universities in the 1000+ band. This constitutes an opportunity to enhance research collaboration, especially with those universities in countries that hold abundant and unique natural resources, not to mention the talent of local researchers and their universities.

The partnerships with other universities and networks open up global opportunities for students and staff. Collaboration among different kind of universities would give them a spectrum of diversity based on different views and backgrounds. This enriches the quality of the research improving the research profile of the university. Top universities could benefit greatly from collaborations with other low ranked universities as they could offer them different approach not before considered making them more prolific and varied. The participation in international networks links the universities to like-minded universities around the world ensuring that they continue at the vanguard of transnational research exchanges.

### Agglomerative hierarchical clustering (AHC)

The next dendrogram shows the hierarchical relationship in publication preferences between twenty research universities. The dendrogram y-axis can be used to evaluate the similarity between research universities (Fig. 2).

The height axis displays the distance (or dissimilarity function) between observations and/or clusters. The horizontal bars indicate the point at which two clusters/observations are merged. For example, ETH Zurich and massachusetts institute of technology (MIT) are merged at a distance of 4, which is the minimum one among all other distances. Also, California Institute of Technology and Princeton University are merged at the value of 6. Finally, ETH Zurich and massachusetts institute of technology (MIT) and California Institute of Technology and Princeton University are merged, and their distance is 10.

University of Cambridge and University of California at Berkeley are more similar to each other than they are to University of Chicago, Stanford University and Imperial College London.

University of Chicago and Stanford University are more similar to each other than they are to University of Cambridge, University of California at Berkeley, and Imperial College London.

Imperial College London is substantially different from all of these research universities.

In the next Fig. 3, we have created a co-occurrence term map based on data used in title and abstract field of the publications of top 20 research universities. We have chosen the binary counting, which means that only the presence or the absence of a term matters.

The terms used in the publications are categorized into four clusters: Cluster 1 or red (terms related to health sciences), Cluster 2 or green (physics sciences), Cluster 3 or blue (social sciences) and Cluster 4 or yellow (terms related to life sciences). The highest number of words are concentrated in the cluster 1- this has a connection to the average research preferences of universities in medicine. The weight given to the color of a certain cluster is determined by the number of items belonging to that cluster in the neighborhood of the point.

**Table 4** Adjacency matrix of top research universities\* cosine similarity in research subject areas

	Cali- for- nia insti- tute of tech- nol- ogy	Colum- bia univer- sity	Cornell univer- sity	Duke univer- sity	Har- vard univer- sity	Impe- rial col- lege Lon- don	Impe- rial col- lege Lon- don	Johns Hop- kins univer- sity	Mas- sachu- setts insti- tute of tech- nol- ogy	Prince- ton univer- sity	Stan- ford univer- sity	Swiss Federal insti- tute of tech- nol- ogy Zurich	Univer- sity of Lon- don	Univer- sity at Berke- ley	Univer- sity of Califor- nia	Univer- sity of Califor- nia at Los Ange- les	Univer- sity of Cam- bridge	Univer- sity of Chi- cago	Univer- sity of Oxford	Univer- sity of Penn- sylva- nia	Univer- sity of Toron- to	Yale Univer- sity
California institute of tech- nology	1	0.738	0.721	0.788	0.805	0.761	0.819	0.842	0.772	0.852	0.815	0.817	0.765	0.847	0.787	0.843	0.784	0.763	0.870			
Columbia univer- sity	0.738	1	0.704	0.876	0.800	0.732	0.790	0.706	0.825	0.731	0.887	0.698	0.836	0.751	0.776	0.798	0.823	0.801	0.811			
Cornell univer- sity	0.721	0.704	1	0.722	0.715	0.839	0.787	0.863	0.792	0.684	0.738	0.809	0.720	0.824	0.742	0.840	0.817	0.789	0.790			
Duke univer- sity	0.788	0.876	0.722	1	0.833	0.773	0.859	0.752	0.726	0.726	0.752	0.844	0.694	0.778	0.765	0.844	0.803	0.747	0.820			
Harvard univer- sity	0.805	0.800	0.715	0.833	1	0.777	0.854	0.715	0.743	0.798	0.688	0.895	0.750	0.872	0.734	0.756	0.787	0.870	0.866			
Imperial college London	0.761	0.732	0.839	0.773	0.777	1	0.797	0.744	0.803	0.675	0.748	0.751	0.845	0.761	0.818	0.812	0.754	0.825	0.842			
Johns Hop- kins univer- sity	0.819	0.790	0.787	0.859	0.854	0.797	1	0.780	0.758	0.770	0.713	0.786	0.793	0.891	0.693	0.821	0.725	0.792	0.879			

Table 4 (continued)

California institute of technology	0.842	0.706	0.863	0.752	0.715	0.744	0.780	1	0.787	0.790	0.840	0.698	0.821	0.696	0.853	0.734	0.858	0.759	0.750	0.776
Massachusetts institute of technology	0.828	0.794	0.758	0.747	0.743	0.803	0.758	0.787	1	0.762	0.834	0.727	0.906	0.743	0.902	0.747	0.747	0.741	0.757	0.849
Princeton university	0.772	0.825	0.792	0.726	0.798	0.675	0.770	0.790	0.762	1	0.729	0.851	0.712	0.794	0.730	0.776	0.796	0.809	0.765	0.805
Stanford university	0.852	0.731	0.684	0.752	0.688	0.748	0.713	0.840	0.834	0.729	1	0.699	0.834	0.705	0.824	0.740	0.743	0.739	0.721	0.721
Swiss Federal institute of technology Zurich	0.815	0.887	0.738	0.844	0.895	0.751	0.786	0.698	0.727	0.851	0.699	1	0.697	0.821	0.745	0.810	0.817	0.878	0.831	0.844
University college London																				

**Table 4** (continued)

	Cali- for- nia insti- tute of tech- nol- ogy	Colum- bia univer- sity	Cor- nell univer- sity	Duke univer- sity	Har- vard univer- sity	Impe- rial college Lon- don	Johns Hop- kins univer- sity	Mas- sachu- setts insti- tute of tech- nology	Prince- ton univer- sity	Stan- ford univer- sity	Swiss Federal insti- tute of tech- nology Zurich	Uni- versity col- lege Lon- don	Univer- sity of Cali- for- nia at Berke- ley	Univer- sity of Cali- for- nia at Los Ange- les	Uni- versity of Cam- bridge	Uni- versity of Chi- cago	Univer- sity of Oxford	Univer- sity of Penn- sylva- nia	Univer- sity of Toronto	Yale Univer- sity
Univer- sity of Califor- nia at Berke- ley	0.817	0.698	0.809	0.694	0.750	0.845	0.793	0.821	0.906	0.712	0.834	0.697	1	0.778	0.876	0.739	0.723	0.726	0.819	0.835
Univer- sity of Califor- nia at Los Ange- les	0.765	0.836	0.720	0.778	0.872	0.761	0.891	0.696	0.743	0.794	0.705	0.821	0.778	1	0.724	0.793	0.727	0.719	0.834	0.805
Univer- sity of Cam- bridge	0.847	0.751	0.824	0.765	0.729	0.779	0.785	0.853	0.902	0.730	0.824	0.745	0.876	0.724	1	0.747	0.793	0.773	0.738	0.829
Univer- sity of Chi- cago	0.787	0.776	0.742	0.722	0.734	0.818	0.693	0.734	0.747	0.776	0.740	0.810	0.739	0.793	0.747	1	0.755	0.749	0.755	0.746
Univer- sity of Oxford	0.843	0.798	0.840	0.844	0.756	0.812	0.821	0.858	0.747	0.796	0.743	0.817	0.723	0.727	0.793	0.755	1	0.882	0.778	0.874



**Table 4** (continued)

Cali- for- nia insti- tute of tech- nol- ogy	Colum- bia univer- sity	Cor- nell univer- sity	Duke univer- sity	Har- vard univer- sity	Impe- rial col- lege Lon- don	Johns Hop- kins univer- sity	Mas- sachu- setts insti- tute of tech- nology	Prince- ton univer- sity	Stan- ford univer- sity	Swiss Federal insti- tute of tech- nology Zurich	Uni- versity col- lege Lon- don	Univer- sity of Cali- for- nia at Berke- ley	Uni- versity of Cali- for- nia at Los Ange- les	Uni- versity of Cam- bridge	Uni- versity of Chi- cago	Univer- sity of Oxford	Uni- versity of Penn- sylva- nia	Univer- sity of Toronto	Yale Univer- sity			
Univer- sity of Penn- sylvania	Univer- sity of Toronto	Yale univer- sity	0.784	0.823	0.817	0.803	0.787	0.754	0.725	0.759	0.741	0.809	0.739	0.878	0.726	0.719	0.773	0.749	0.882	1	0.863	0.833
0.763	0.801	0.789	0.747	0.870	0.825	0.792	0.750	0.757	0.765	0.721	0.831	0.819	0.834	0.738	0.755	0.778	0.863	1	0.830			
0.870	0.811	0.790	0.820	0.866	0.842	0.879	0.776	0.849	0.805	0.721	0.844	0.835	0.805	0.829	0.746	0.874	0.833	0.830	1			

**Table 5** Adjacency matrix of top research universities' cosine similarity in sub-subject areas

	California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins university	Masachusetts institute of technology	Princeton university	Stanford university	Swiss Federal institute of technology Zurich	University of London	University of California Berkeley	University of California Los Angeles	University of Cambridge	University of Chicago	University of Oxford	University of Pennsylvania	University of Toronto	Yale university
California institute of technology	1	0.749	0.788	0.818	0.729	0.853	0.789	0.885	0.861	0.794	0.844	0.747	0.825	0.823	0.785	0.830	0.764	0.789	0.805	0.719
Columbia university	0.749	1	0.727	0.731	0.747	0.779	0.723	0.788	0.725	0.781	0.836	0.718	0.711	0.713	0.697	0.752	0.766	0.706	0.724	0.696
Cornell university	0.788	0.727	1	0.751	0.824	0.758	0.746	0.757	0.753	0.799	0.763	0.791	0.771	0.803	0.777	0.749	0.756	0.799	0.772	0.771
Duke university	0.818	0.731	0.751	1	0.767	0.836	0.820	0.825	0.786	0.748	0.821	0.695	0.792	0.779	0.732	0.743	0.749	0.818	0.819	0.773
Harvard university	0.729	0.747	0.824	0.767	1	0.771	0.742	0.785	0.774	0.769	0.794	0.800	0.746	0.790	0.801	0.761	0.706	0.785	0.737	0.785
Imperial college London	0.853	0.779	0.758	0.836	0.771	1	0.808	0.847	0.818	0.780	0.872	0.776	0.854	0.834	0.748	0.844	0.763	0.726	0.780	0.771
Johns Hopkins university	0.789	0.723	0.746	0.820	0.742	0.808	1	0.819	0.796	0.786	0.870	0.691	0.838	0.842	0.768	0.803	0.737	0.779	0.785	0.843

Table 5 (continued)

California institute of technology	0.885	0.788	0.757	0.825	0.785	0.847	0.819	1	0.829	0.773	0.901	0.757	0.855	0.763	0.781	0.811	0.805	0.728	0.769	0.755
Massachusetts institute of technology	0.861	0.725	0.753	0.786	0.774	0.818	0.796	0.829	1	0.783	0.833	0.739	0.820	0.823	0.705	0.854	0.727	0.802	0.777	0.746
Princeton university	0.794	0.781	0.799	0.748	0.769	0.780	0.786	0.773	0.783	1	0.837	0.719	0.762	0.782	0.804	0.771	0.723	0.727	0.681	0.732
Stanford university	0.844	0.836	0.763	0.821	0.794	0.872	0.870	0.901	0.833	0.837	1	0.764	0.859	0.796	0.833	0.838	0.803	0.798	0.813	0.812
Swiss Federal institute of technology Zurich	0.747	0.718	0.791	0.695	0.800	0.767	0.691	0.757	0.739	0.719	0.764	1	0.739	0.739	0.727	0.782	0.707	0.796	0.711	0.622
University college London																				

Table 5 (continued)

California institute of technology	0.825	0.711	0.771	0.792	0.746	0.854	0.838	0.855	0.820	0.762	0.859	0.739	1	0.827	0.737	0.783	0.802	0.757	0.788	0.812
University of California at Berkeley	0.823	0.713	0.803	0.779	0.790	0.834	0.842	0.763	0.823	0.782	0.796	0.739	0.827	1	0.745	0.795	0.765	0.787	0.821	0.840
University of California at Los Angeles	0.785	0.697	0.777	0.732	0.801	0.748	0.768	0.781	0.705	0.804	0.833	0.727	0.737	0.745	1	0.715	0.705	0.820	0.676	0.768
University of Cambridge	0.830	0.752	0.749	0.743	0.761	0.844	0.803	0.811	0.854	0.771	0.838	0.782	0.783	0.795	0.715	1	0.735	0.775	0.764	0.727
University of Chicago	0.764	0.766	0.756	0.749	0.706	0.763	0.737	0.805	0.727	0.723	0.803	0.707	0.802	0.705	0.735	0.735	1	0.701	0.786	0.680

**Table 5** (continued)

California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins university	Massachusetts institute of technology	Princeton university	Stanford university	Swiss Federal institute of technology Zurich	University of London	University of California Berkeley	University of California Los Angeles	University of Cambridge	University of Chicago	University of Oxford	University of Pennsylvania	University of Toronto	Yale university
0.789	0.706	0.799	0.818	0.785	0.726	0.779	0.728	0.802	0.727	0.798	0.796	0.757	0.787	0.820	0.775	0.701	1	0.808	0.737
University of Pennsylvania	0.805	0.724	0.772	0.819	0.737	0.785	0.769	0.777	0.681	0.813	0.711	0.788	0.821	0.676	0.764	0.786	0.808	1	0.774
University of Toronto	0.719	0.696	0.771	0.773	0.785	0.771	0.843	0.746	0.732	0.812	0.622	0.812	0.840	0.768	0.727	0.680	0.737	0.774	1
Yale university																			

**Table 6** Adjacency matrix of cosine similarity between the world's top and lowest-ranked research universities

California institute of technology	0.737	0.872	0.659	0.809	0.671	0.671	0.696	0.740	0.792	0.787	0.807	0.755	0.707	0.772	0.759	0.773	0.721	0.723	0.729	0.717
Universidade Estadual do Oeste do Paraná	0.806	0.766	0.753	0.773	0.736	0.736	0.796	0.753	0.852	0.794	0.794	0.749	0.783	0.801	0.800	0.737	0.760	0.769	0.736	0.773
University of West Bohemia	0.767	0.825	0.812	0.777	0.764	0.764	0.816	0.760	0.749	0.862	0.642	0.840	0.773	0.829	0.770	0.776	0.823	0.766	0.808	0.867
University of Yamaguchi	0.774	0.775	0.706	0.756	0.703	0.703	0.683	0.712	0.764	0.771	0.747	0.777	0.738	0.802	0.719	0.846	0.778	0.799	0.773	0.739
University of Zagreb	0.766	0.709	0.784	0.802	0.845	0.845	0.808	0.770	0.748	0.632	0.722	0.709	0.809	0.777	0.823	0.772	0.763	0.694	0.780	0.770
Warsaw university of technology	0.721	0.848	0.766	0.787	0.809	0.809	0.747	0.688	0.756	0.774	0.697	0.771	0.696	0.814	0.687	0.786	0.791	0.787	0.779	0.755
Wenzhou medical university																				

Table 6 (continued)

California institute of technology	0.784	0.857	0.687	0.798	0.684	0.684	0.730	0.773	0.832	0.778	0.768	0.755	0.827	0.795	0.821	0.730	0.753	0.793	0.730	
West university of Timisoara																				
Wroclaw university of science and technology	0.734	0.783	0.812	0.748	0.763	0.763	0.801	0.751	0.714	0.747	0.676	0.769	0.781	0.733	0.744	0.804	0.806	0.841	0.778	
Yamagata university	0.695	0.787	0.755	0.721	0.777	0.777	0.718	0.708	0.786	0.765	0.746	0.771	0.802	0.804	0.786	0.744	0.762	0.803	0.732	
Yamaguchi university	0.777	0.757	0.773	0.726	0.797	0.797	0.745	0.740	0.806	0.827	0.771	0.810	0.740	0.733	0.686	0.706	0.746	0.821	0.802	
Yanshan university	0.809	0.724	0.667	0.744	0.711	0.711	0.799	0.764	0.828	0.738	0.826	0.818	0.761	0.765	0.642	0.726	0.701	0.759	0.787	
Yarmouk university	0.782	0.825	0.736	0.724	0.750	0.750	0.750	0.769	0.763	0.794	0.738	0.800	0.806	0.731	0.818	0.786	0.733	0.805	0.743	

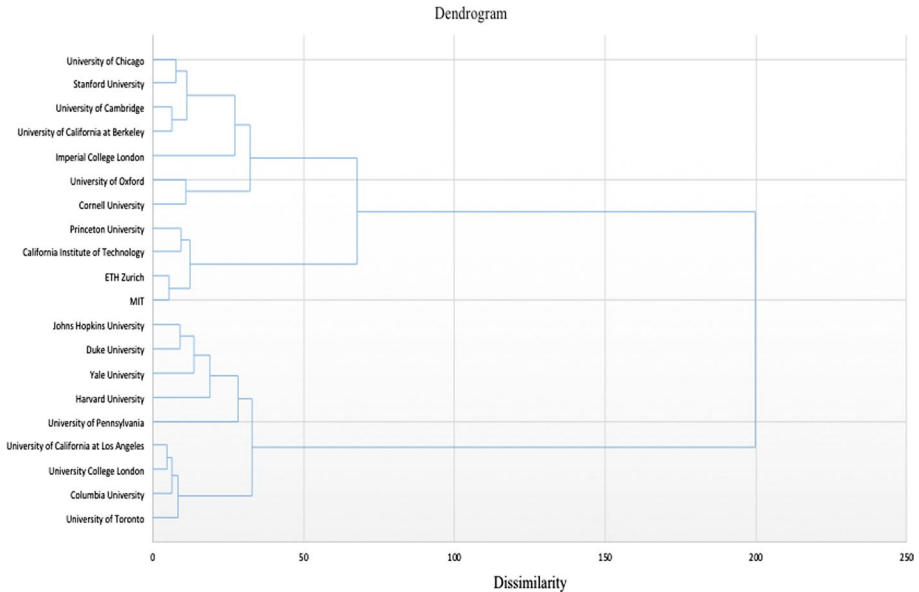
Table 6 (continued)

	California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins university	Massachusetts institute of technology	Princeton university	Stanford university	Swiss Federal institute of technology Zurich	University of California Berkeley	University of California Los Angeles	University of Cambridge	University of Chicago	University of Oxford	University of Pennsylvania	University of Toronto	Yale university
Yazd university	0.726	0.763	0.777	0.684	0.783	0.783	0.717	0.805	0.752	0.825	0.751	0.791	0.789	0.734	0.842	0.786	0.772	0.842	0.761
Yeditepe university	0.790	0.825	0.715	0.817	0.705	0.705	0.779	0.759	0.744	0.840	0.746	0.781	0.855	0.706	0.808	0.770	0.687	0.687	0.745
Yildiz technical university	0.720	0.776	0.799	0.709	0.708	0.708	0.634	0.773	0.777	0.822	0.711	0.764	0.651	0.776	0.844	0.743	0.774	0.699	0.718
Yokohama national university	0.805	0.804	0.655	0.766	0.662	0.662	0.651	0.803	0.794	0.782	0.863	0.784	0.708	0.806	0.774	0.711	0.754	0.774	0.689
Yuan Ze university	0.777	0.864	0.719	0.752	0.724	0.724	0.698	0.750	0.741	0.818	0.736	0.853	0.788	0.692	0.786	0.814	0.817	0.794	0.754
Zagazig university	0.783	0.770	0.734	0.804	0.830	0.830	0.742	0.755	0.809	0.729	0.803	0.692	0.754	0.756	0.811	0.766	0.717	0.704	0.740
Zanjan university	0.819	0.801	0.749	0.856	0.833	0.833	0.846	0.729	0.768	0.709	0.781	0.825	0.765	0.787	0.697	0.838	0.811	0.808	0.851



**Table 6** (continued)

California institute of technology	Columbia university	Cornell university	Duke university	Harvard university	Imperial college London	Johns Hopkins university	Massachusetts institute of technology	Princeton university	Stanford university	Swiss Federal institute of technology Zurich	University of London	University of California Berkeley	University of California Los Angeles	University of Cambridge	University of Chicago	University of Oxford	University of Pennsylvania	University of Toronto	Yale university
0.829	0.745	0.820	0.825	0.812	0.812	0.772	0.860	0.814	0.760	0.844	0.780	0.794	0.696	0.794	0.768	0.811	0.760	0.778	0.778
Zhejiang university of technology																			



**Fig. 2** Dendrogram with classification of top research universities in clusters

A high number of repetitions of terms means that they are continually being chosen by universities in their publications. Therefore, these occurrences used by the leading universities might provide other researchers an overview either to avoid these research topics which are related with these particular words or it might show that it is a potential research topic that they can explore in future to become a new research topic.

The term Patient is most used in the cluster 1/red with a number of 4896 occurrences and a relevance of 0.81; Research is the word most used in the cluster 2/green with 4490 with a relevance of 0.67; the third is Structure used 4197 times in the cluster 3/blue with a relevance of 0.29; and Cell with 3967 occurrences and 0.82 relevance in the cluster 4/ yellow.

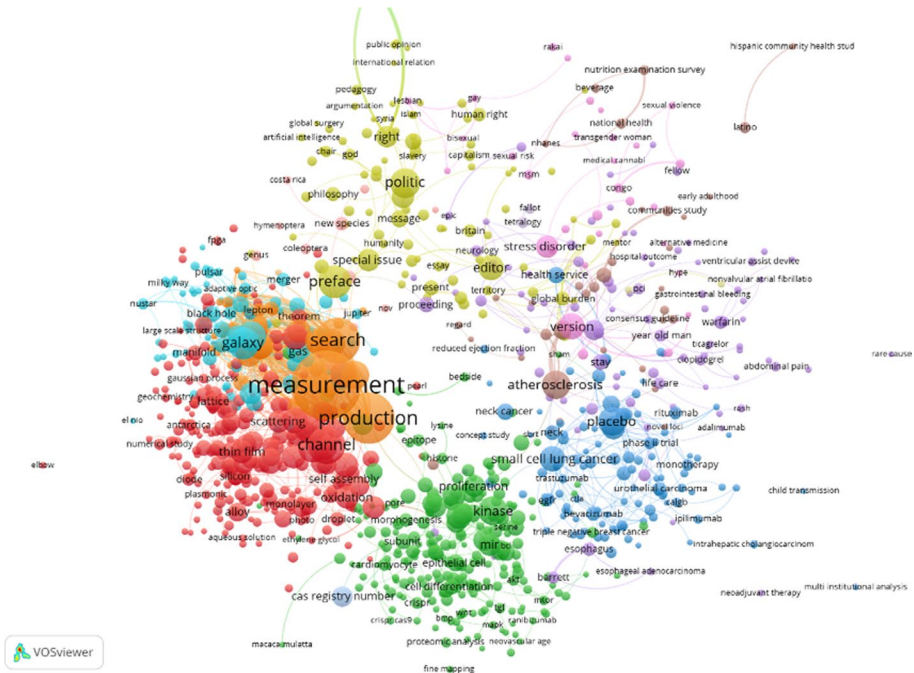
As a mode of reflection we have noticed that words resulting from university publications are fairly general therefore we are considering that perhaps a study by keywords would be relevant as future work and could give us more information.

In Fig. 4 we have represented the most commonly terms used only in the titles. The links represent the words relations.

The interrelationships of terms in the map above are based on distance, i.e. the distance of circles shows the relationship between terms. Bigness or smallness of each circle reveal the number of occurrences of each concept. Then, as shown in Fig. 4, concentration of the words in the titles would mostly be upon the subjects such as “measurement”, “production”, “search”, “channel”, “preface”, “galaxy”. The distribution of the words in the map reflects that scholars in research universities have used words from different subjects in their titles.

For the purpose of finding out how research universities collaborate with each other we have done a search on Scopus for the affiliation field with the same publishing window.





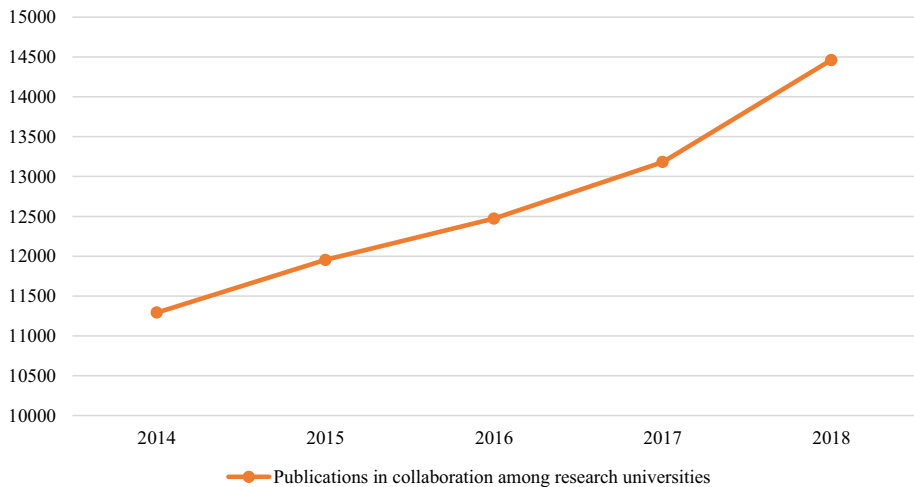
**Fig. 4** Co-occurrence map based on text data from titles

The five areas where research universities collaborated the most are: Medicine (23,136), Physics and Astronomy (12,711) and Biochemistry, Genetics and Molecular Biology (11,614). On the contrary, Business, Management and Accounting (704), Decision Sciences (685), Health Professions (466) are the subject areas with fewer publications in common (Fig. 6).

The three most collaborating universities are University of Oxford (11,220), UCL (10,168), University of Cambridge (9647), all of them from the United Kingdom (Fig. 7).

The topics involved in the collaborations can be outlined in the keywords assigned to each publication. Keywords provide quick access to scientific publications and are highly effective in terms of bibliometric analysis when examining the knowledge structure of scientific fields (Zhang et al. 2016; Vargas-Quesada et al. 2017). Keywords offer a practical description of research hotspots and topic prominence (the attention by researchers to set of related research problems and concepts). In the present study, VOSviewer was used to create a knowledge map of keyword co-occurrence with 1500 terms in 11 clusters, 16,119 Links (Fig. 8) and to identify the top 5 keywords.

The top 5 keywords were “Measurement” “Search” “Observation” “Event” and “Galaxy”.



**Fig. 5** Number of publications in collaboration between the top 20 research universities

## Conclusions

Our study has analyzed the research preferences of twenty leading research universities according to the last edition of THE rankings 2020. We have used the Scopus database to extract data and SciVal to apply indicators with a 5-year publication window from 2014 to 2018.

Most of the universities in our sample belong to the United States and all of them are classified as "R1: Doctoral Universities—Very high research activity" in the Carnegie Classification.

The top twenty research universities have in common high levels of scientific collaborations and high number of publications in top journals. Their publications have been cited twice more than would be expected based on the world average for similar publications as stated in their FWCI.

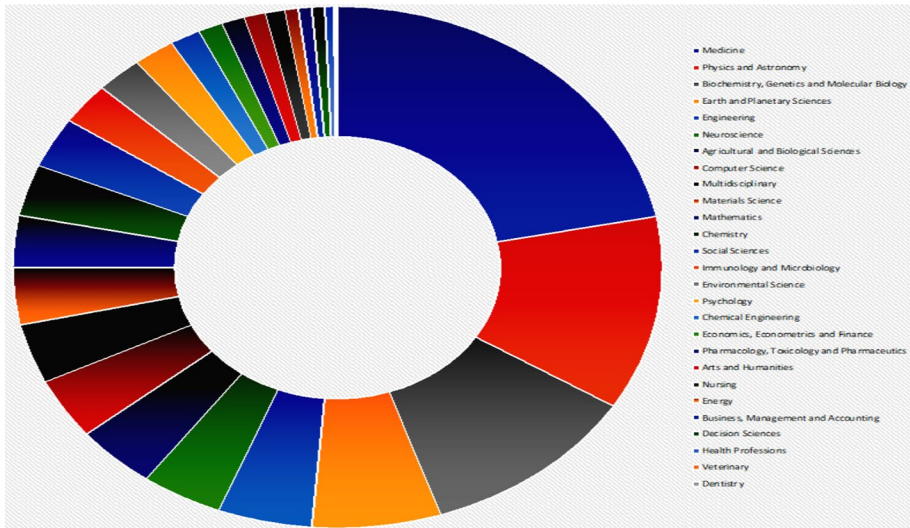
According to Spearman the indicators Research from THE, citations per publication, FWCI and the output in top 10 citation percentiles from SciVal have an impact on the Overall score of universities in the last edition.

Through cosine similarity we have been able to find out that research universities are quite similar to each other in research preferences, however there are some more related to each other than others, such as Princeton University and University of Cambridge or John Hopkins University with University of California at Los Angeles. Regarding to the analysis of sub subject areas we detected that University of Oxford is very similar to MIT and ETH Zurich.

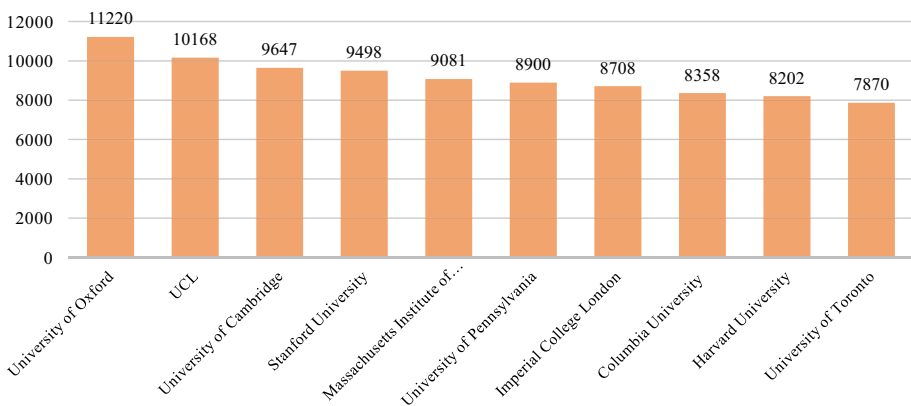
With agglomerative hierarchical clustering (AHC) we were able to classify research universities in groups according their dissimilarity applying Euclidean distance.

Looking the data in this study, we can affirm that the lowest-ranked research universities are similar in research preferences to the twenty top.

The keywords used in the publications of research universities are categorized into four clusters: Cluster 1 or red (keywords related to health sciences), Cluster 2 or green (keywords related to physics sciences), Cluster 3 or blue (keywords related to social



**Fig. 6** Research preference in collaboration among universities

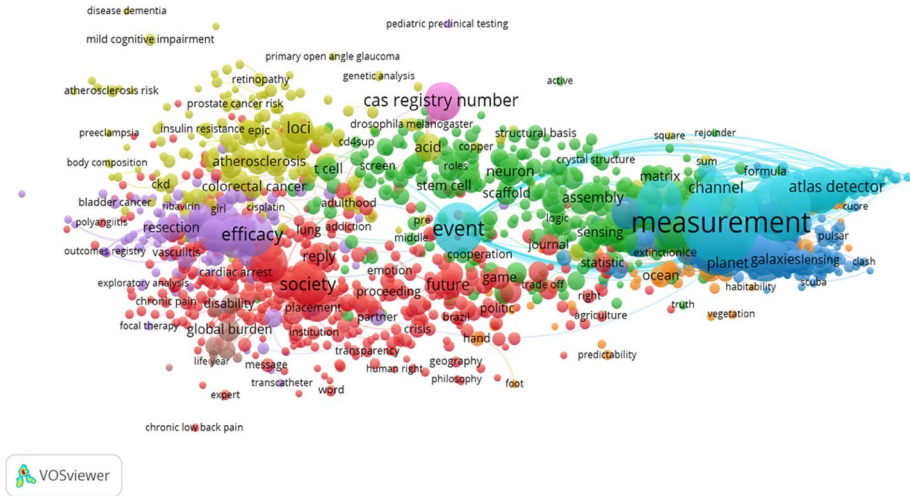


**Fig. 7** Most collaborating research universities

sciences)- Cluster 4 or yellow (keywords related to life sciences). Most of the words are concentrated in the first clusters of health sciences.

The twenty research universities have a total of 63,361 publications in collaboration. There is a slight increase in collaborative publications over the years. They have collaborated the most in the field of Medicine, Physics and Astronomy, Biochemistry, Genetics and Molecular Biology. University of Oxford, UCL and University of Cambridge are the most collaborating universities. University of Washington is the university that collaborates most with the 20 research universities.

Also, we presented results and an analysis of similarity among the leading universities and the ones in the 1000+ band of THE WUR and indication of the opportunity this represents for enhancing international collaboration and partnership in a win-to-win relationship.



**Fig. 8** Co-occurrence map based in collaborations

### Next steps and future works

In future research, we would like to conduct a deeper analysis of the research patterns of the lowest- ranked research universities with the top-ranked and explore further opportunities in addition to the ones we have presented in this study, and possible extend the band to universities ranked in the 800+ positions. At the same time, we could analyse the scientific collaboration between the runners-up universities and top universities to know the benefits obtained from this diversity. Also, we would like to analyse the differences between comprehensive universities and specialized universities. This makes it possible to compare more universities by research fields and specializations. To do this we would use bibliographic databases to extract publications. We would like to find out how beneficial are these collaborations in terms of visibility and impact. The international co-authorship also increases the citations of the institution, yet there are untapped potential to enhance the collaboration among institutions.

Another future research idea is to analyze the scientific contribution of these twenty universities to the scientific production in the 17 goals UN SDGs (No Poverty; Zero Hunger; Good Health and Well-being; Quality Education; Gender Equality; Clean Water and Sanitation; Affordable and Clean Energy; Decent Work and Economic Growth; Industry, Innovation, and Infrastructure; Reducing Inequality; Sustainable Cities and Communities; Responsible Consumption and Production; Climate Action; Life Below Water; Life On Land; Peace, Justice, and Strong Institutions; Partnerships for the Goals. For this we would extract the production of these universities in these 17 goals of UN SDGs how the production of the universities is in each of the challenges and we would apply cosine similarity to see if they are similar to each other on these issues as well.

## References

- Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*, 1129–1144. <https://doi.org/10.1007/s11192-014-1269-8>.
- Altbach, P. G. (2015). What counts for academic productivity in research universities? *International Higher Education*, *79*, 6–7. <https://doi.org/10.6017/ihe.2015.79.5837>.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256.
- Atkinson, R. C., & Blanpied, W. A. (2008). Research universities: Core of the US science and technology system. *Technology in Society*, *30*(1), 30–48.
- Barirani, A., Agard, B., & Beaudry, C. (2013). Competence maps using agglomerative hierarchical clustering. *Journal of Intelligent Manufacturing*, *24*(2), 373–384.
- Bolli, T., Olivares, M., Bonaccorsi, A., Daraio, C., Aracil, A. G., & Lepori, B. (2016). The differential effects of competitive funding on the production frontier and the efficiency of universities. *Economics of Education Review*, *52*, 91–104. <https://doi.org/10.1016/j.econedurev.2016.01.007>.
- Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Application*, *42*(5), 2785–2797.
- Budd, J. (2017). Faculty publications and Citations: A longitudinal examination. *College & Research Libraries*, *78*(1), 80. <https://doi.org/10.5860/crl.78.1.80>.
- Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy*, *34*(8), 565–574. <https://doi.org/10.3152/030234207X254404>.
- Carnegie Foundation. (2001). *The carnegie classification of institutions of higher education*. California: Carnegie Foundation.
- Chen, Y., Nixon, M. R., Gupta, A., & Hoshower, L. (2010). Research productivity of accounting faculty: An exploratory study. *American Journal of Business Education*, *3*(2), 101–115.
- Carnegie classification of institutions of higher education (2019). Retrieved from 13 May 2020 <https://carnegieclassifications.iu.edu/>
- Cloete, N., Bailey, T., & Pillay, P. (2011). *Universities and economic development in Africa*. African Minds: Cape.
- CWTS Leiden Ranking. (2020). Retrieved November 24, from <https://www.leidenranking.com/>.
- European commission (2019)
- Fernandez, F., & Baker, D. (2017). Science production in the United States: An unexpected synergy between mass higher education and the super research University. In J. J. W. Powell, D. P. Baker, & F. Fernandez (Eds.), *The century of science (international perspectives on education and society)* (Vol. 33, pp. 85–111). Bingley: Emerald Publishing Limited.
- [https://eacea.ec.europa.eu/national-policies/eurydice/content/types-higher-education-institutions-91\\_en](https://eacea.ec.europa.eu/national-policies/eurydice/content/types-higher-education-institutions-91_en)  
Retrieved from 13 May 2020.
- Hu, Z., Guo, F., & Hou, H. (2017). Mapping research spotlights for different regions in China. *Scientometrics*, *110*, 779–790.
- Huang, M. (2012). Exploring the h-index at the institutional level: A practical application in world university rankings. *Online Information Review*, *36*(4), 534–547. <https://doi.org/10.1108/14684521211254059>.
- Jung, J. (2012). International research collaboration among academics in china and south korea. *International Journal of Chinese Education*, *1*(2), 235–254. <https://doi.org/10.1163/22125868-12340005>.
- Lancho-Barrantes, B. S., & Cantu-Ortiz, F. J. (2019). Quantifying the research preferences of top research universities: Why they make a difference? *Paper presented at the 17th international conference on scientometrics and informetrics, ISSI 2019 - Proceedings* (vol. 2, pp. 1488–1499).
- League of European Research Universities (2019). [www.leru.org](http://www.leru.org).
- Leydesdorff, L., Bornmann, L., & Mingers, J. (2019). Statistical significance and effect sizes of differences among research universities at the level of nations and worldwide based on the leiden rankings. *Journal of the Association for Information Science and Technology*, *70*(5), 509–525. <https://doi.org/10.1002/asi.24130>.
- Lin, G., Hu, Z., & Hou, H. (2018). Research preferences of the G20 countries: A bibliometrics and visualization analysis. *Current Science*, *115*(8), 1477–1485.
- Mammadov, R., & Aypay, A. (2020). Efficiency analysis of research universities in Turkey. *International Journal of Educational Development*. <https://doi.org/10.1016/j.ijedudev.2020.102176>.
- McGill, M. M., & Settle, A. (2012). Identifying effects of institutional resources and support on computing faculty research productivity, tenure, and promotion. *International Journal of Doctoral Studies*, *7*, 167.
- Mohrman, K., Ma, W., & Baker, D. P. (2008). The research university in transition: The emerging global model. *Higher Education Policy*, *21*(1), 5–27.



- Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., et al. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78.
- Musiige, G., & Maassen, P. (2015). Faculty perceptions of the factors that influence research productivity at Makerere University. Knowledge production and contradictory functions in African higher education, pp. 109–127. <https://zenodo.org/record/824663#.X72HyhP7R0s>.
- Powell, J. J., & Dusdal, J. (2017a). Science production in Germany, France, Belgium, and Luxembourg: Comparing the contributions of research universities and institutes to science, technology, engineering, mathematics, and health. *Minerva*, 55(4), 413–434.
- Powell, J. J. W., & Dusdal, J. (2017b). The European Center of science productivity: Research universities and institutes in France, Germany, and the United Kingdom. *International Perspectives on Education and Society*, 33, 55–83.
- QS World University Rankings. (2020). Retrieved November 24, from <https://www.topuniversities.com/qs-world-university-rankings>.
- Sammour, M. (2019). An agglomerative hierarchical clustering with association rules for discovering climate change patterns. *International Journal of Advanced Computer Science and Applications*, 10(3), 233–240. <https://doi.org/10.14569/IJACSA.2019.0100330>.
- SciVal (2019) Retrieved from 28 Jan 2019 <https://www.scival.com/>.
- Scopus (2019a) Retrieved from 28 Jan 2019 <https://www.scopus.com/home.uri>.
- Scopus (2019b). ASJC classification. Retrieved from 13 Dec 2019. [https://service.elsevier.com/app/answers/detail/a\\_id/14834/supporthub/scopus/](https://service.elsevier.com/app/answers/detail/a_id/14834/supporthub/scopus/).
- Singhal, A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24, 35–43.
- Taylor, J. (2006). Managing the unmanageable: The management of research in research-intensive universities. *Higher Education Management and Policy*, 18(2), 3–4.
- The Russell Group. Retrieved from 13 May 2020 <https://russellgroup.ac.uk/about/>.
- Toutkoushian, R. K., Porter, S. R., Danielson, C., & Hollis, P. R. (2003). Using publications counts to measure an institution's research productivity. *Research in Higher Education*, 44(2), 121–148.
- University World News (2013). The role of research universities in developing countries. Retrieved from 28 Jan 2019 <https://www.universityworldnews.com/post.php?story=20130811091502202>.
- van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053–1070.
- Van Raan, A. F. J. (2008). Bibliometric statistical properties of the 100 largest European research universities: Prevalent scaling rules in the science system. *Journal of the American Society for Information Science and Technology*, 59(3), 461–475. <https://doi.org/10.1002/asi.20761>.
- Vargas-Quesada, B., Chinchilla-Rodríguez, Z., & Rodríguez, N. (2017). Identification and visualization of the intellectual structure in graphene research. *Frontiers in Research Metrics and Analytics*, 2, 7. <https://doi.org/10.3389/frma.2017.00007>.
- VOSviewer - Visualizing scientific landscapes. (2020). Retrieved November 24, from <https://www.vosviewer.com/>.
- Waltman, L., & van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B*. <https://doi.org/10.1140/epjb/e2013-40829-0>.
- Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635.
- Webber, K. L. (2011). Measuring Faculty Productivity. In J. Shin, R. Toutkoushian, & U. Teichler (Eds.), *University Rankings. The changing academy – the changing academic profession in international comparative perspective*. Dordrecht: Springer.
- World University Rankings | Times Higher Education (THE). (2020). Retrieved November 24, from <https://www.timeshighereducation.com/world-university-rankings>.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. *SIGMOD*, 25(2), 103–114.
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science Technology*, 67, 967–972. <https://doi.org/10.1002/asi.23437>.
- Zhao, K., & Suzuki, E. (2015). Clustering classifiers learnt from local datasets based on cosine similarity, pp. 150–159. [https://doi.org/10.1007/978-3-319-25252-0\\_16](https://doi.org/10.1007/978-3-319-25252-0_16).
- Zhigang, H., Gege, L., Haiyan, H. (2017). Research preferences of the G20 countries: A bibliometrics and visualization analysis. Paper presented at the ISSI 2017 - 16th International Conference on Scientometrics and Informetrics, Conference Proceedings, pp. 709–720.

## Affiliations

Barbara S. Lancho-Barrantes<sup>1</sup>  · Francisco J. Cantu-Ortiz<sup>2</sup>

Francisco J. Cantu-Ortiz  
fcantu@tec.mx

<sup>1</sup> University of Leeds, Leeds LS2 9JT, UK

<sup>2</sup> Tecnológico de Monterrey, Eugenio Garza Sada 2501, 64849 Monterrey, Mexico