An Observation of Research Complexity in Top Universities Based on Research Publications

Ivan Lee
University of South Australia
Mawson Lakes, SA, Australia
Dalian University of Technology
Dalian, Liaoning, China
ivan.lee@unisa.edu.au

Feng Xia
Dalian University of Technology
Dalian, Liaoning, China

Göran Roos
Nanyang Business School, Nanyang Technological University
Singapore

ABSTRACT

This paper investigates research specialisation of top ranked universities around the world. The revealed comparative advantage in different research fields are determined according to the number of research articles published. Subsequently, measures of research ubiquity and diversity, and research complexity index of each university, are obtained and discussed. The study is conducted on top-ranked universities according to Shanghai Jiao Tong Academic Ranking of World Universities, with bibliographical details extracted Microsoft Academic Graph data set and research fields of journals labelled with SCImago Journal Classification. Diversity-ubiquity distributions, relevance of RCI and university ranks, and geographical RCI distributions are examined in this paper.

Keywords

scientometrics; complexity modelling

1. INTRODUCTION

Research output analysis has been an active research area, and it is one major factor assessing the performance of a university, subsequently influence the allocation of education fund and research grants, the attractiveness to organisations interested in procuring applied research services, as well as the attention from prospective students. Most popular university rankings, such as Shanghai Jiao Tong Academic Ranking of World Universities (ARWU), Quacquarelli Symonds World University Ranking (QS), and Times Higher Education World University Rankings (THE), consider research output as one of the crucial factors towards ranking universities around the world. The rankings, from a global prospective, may be considered the scientific competitiveness of nations [Cimini et al. 2014].

While universities usually comprise of a number of faculties or divisions and pursue research in different fields, most university rankings (include the aforementioned examples) do provide additional ranking break-downs associated to research fields. Analysis of disciplinary specialisation has also attracted attentions in the research community, to explore its relevance and differences to university ranks [López-Illescas et al. 2011] [Robinson-García and Calero-Medina 2014]. Mapping of research disciplines has also been investigated, such as the study based on ISI Web of Science (WoS) [Leydesdorff and Rafols 2009] and the study based on combined WoS and Scopus [Börner et al. 2012].

Further to analysing and visualising the distribution of different research fields, clarification of universities according to research specialty has also been investigated in the past. Ortega et al. has applied a clustering algorithm to identify specialisation of Spanish institutes based on their research outputs [Ortega et al. 2011], and classify these institutes as either Humanistic, Scientific, or Technological. Studies on the most frequently publishing universities in Spain suggest that university rankings should account for disciplinary specialisation [López-Illescas et al. 2011]. Li et al. has conducted another study of research specialisation at country-level in China, and the change of disciplinary structure over time has been investigated [Li et al. 2015]. Beyond the country-level, studies of research specialisation has been reported for European universities [Daraio et al. 2015] [Pastor and Serrano 2016], and for universities around the world [Harzing and Giroud 2014]. While bias in research grant allocation to small universities are observed according to the study by Murray et al. [Murray et al. 2016], analysis of research specialisation may help improving the grant allocation model.

While the research outcome of a research institute comes in different forms, including research publications, copyrights, patents [Wong and Singh 2010] [Payumo et al. 2014]. Among these, research publication is possibly the most popular benchmark to assess the research productivity of a university. This paper attempts a different approach to examine research publications produced by top universities around the world. Economic Complexity Index (ECI) [Ricardo et al. 2008], a modern technique for modelling the economic complexity according to the export record of countries in the world [Hidalgo and Hausmann 2009], has been adopted for analysing research complexity of universities. The original model measures a country's industrial composition by analysing its export record, and combines the diversity of ubiquity to form the economic complexity index, which can be used to pre-

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3053054

dict the economic growth. In this paper, instead of using the export record, research publications are utilised to measure the diversity and ubiquity of an institute's research output. The complexity model for academic output analysis has been recently investigated [Guevara et al. 2016], and the work by Guevara et al. has focused on grouping the publications according to research fields, and analyse the proximity across different research fields. This paper compares top universities in terms of their research ubiquity and diversity, as well as their complexity indices. Correlation between opportunity value and research complexity index is also investigated in this paper.

The rest of this article is organised as follows: we start from describing how revealed comparative advantage are obtained through academic publications in section 2, followed by the definition of ubiquity and complexity of a university, and the concepts of research complexity index, capability distance, and opportunity value in section 3. Sections 4 and 5 covers the methods for data processing and the experimental results. Finally, our concluding remarks are drawn in section 6.

2. REVEALED COMPARATIVE ADVANTAGE

In economic analysis[Hidalgo and Hausmann 2009], export record across different industrial sectors are utilised for complexity modelling. To evaluate the performance of research institutions, publication outputs of different research disciplines can be used to replace the export values. In our study, we use the SCImago Journal Classification which labels journal in 27 research areas. Although, there are a number of multi-disciplinary publications, such as prestigious journals *Nature* and *Science*, accept articles from more than one research field. Let i denote an academic institute (university), and f denotes a research field. For a paper $p(i,f)$ published in journals labelled with m disciplines where $m \in \mathbb{Z}$ and $m \geq 1$, a normalisation factor $m_{p(i,f)}$ needs to be applied. We define the publication matrix $\boldsymbol{P}_{if}$ as the summation over all papers $p(i,f)$ in research field f and has at least one author from university (or institute) i, normalised by the number of research fields covered by the journal. Mathematically, $\boldsymbol{P}_{if}$ is expressed as

$$\boldsymbol{P}_{if} = \sum_{p(i,f)} \frac{1}{m_{p(i,f)}}. \tag{1}$$

Revealed comparative advantage (RCA) is defined as the percentage of academic publications of a particular discipline in an academic institute, divided by the percentage of all papers by the academic institute over all academic publications. Thus, if the academic institute has a strong focus in a particular research discipline, then the nominator term increases hence increase the overall RCA value. While the academic institute may be big or small in terms of the volume of research outputs, the normalising term in the denominator helps reduce the impact due to the institute's size. Mathematically, the revealed comparative advantage $\boldsymbol{RCA}_{ij}$ for research institute i and research field f can be expressed as

$$\boldsymbol{RCA}_{if} = \frac{\boldsymbol{P}_{if} / \sum_f \boldsymbol{P}_{if}}{\sum_i \boldsymbol{P}_{if} / \sum_{i,f} \boldsymbol{P}_{if}}. \tag{2}$$

It should be noted that RCA represents the level of specialisation, and not a measure of competitiveness of a research field.

The RCA values across different disciplines are used to construct the $\boldsymbol{M}_{if}$ matrix, and each element in $\boldsymbol{M}_{if}$ is a boolean value which indicate whether an academic institute has revealed comparative advantage in a particular field of research.

$$\boldsymbol{M}_{if} = \begin{cases} 1 & \text{if } \boldsymbol{RCA}_{if} \geq t, \\ 0 & \text{if } \boldsymbol{RCA}_{if} < t. \end{cases} \tag{3}$$

where t is a manually selected threshold value. In this paper, the threshold value follows the setup in Economic Complexity analysis [Ricardo et al. 2008] and t value is chosen as 1, expect for investigating the correlation between the complexity indices and university ranks, where multiple t values are used. Further details are covered in sections 3 and 5.

3. DIVERSITY, UBIQUITY, AND RESEARCH COMPLEXITY

With RCA values summarised in M_{if}, the measure of an institute's research fields with reveal comparative advantage, or diversity $\boldsymbol{D}_i$, can be formulated as:

$$\boldsymbol{D}_i = k(i,0) = \sum_f \boldsymbol{M}_{if}. \tag{4}$$

Similarly, the measure of the degree of how many research institutes has revealed comparative advantage in a research field can be obtained. This measure, known as the ubiquity $\boldsymbol{U}_f$, can be formulated as

$$\boldsymbol{U}_f = k(f,0) = \sum_i \boldsymbol{M}_{if}. \tag{5}$$

Using the *Method of Reflections* [Hidalgo and Hausmann 2009], the value of $k(i,n)$ can be formulated in terms of $k(f,n-1)$:

$$k(i,n) = \frac{1}{k(i,0)} \sum_f \boldsymbol{M}_{if} k(f,n-1). \tag{6}$$

Likewise, the value $k(f,n)$ can be obtained in terms of $k(i,n-1)$:

$$k(f,n) = \frac{1}{k(f,0)} \sum_i \boldsymbol{M}_{if} k(i,n-1). \tag{7}$$

Subsequently, each institute may be characterised by the vector $\boldsymbol{k}_i = \{k(i,n) \mid n = 1 \ldots N\}$ and $\boldsymbol{k}_f = \{k(f,n) \mid n = 1 \ldots N\}$ where N is a pre-determined number of iterations and $N \in \mathbb{Z}$ and $N > 1$. Equations (6) and (7) can be written as

$$k(i,n) = \frac{1}{k(i,0)} \sum_f \boldsymbol{M}_{if} \frac{1}{k(f,0)} \sum_i \boldsymbol{M}_{if} k(i,n-2), \tag{8}$$

which can be rewritten as

$$k(i,n) = \sum_{i'} \widetilde{M_{ii'}} k(i',n-2). \tag{9}$$

The research complexity index of an institute, $RCI(i)$, is calculated according to

$$RCI(i) = \frac{\boldsymbol{K}_i - \overline{\boldsymbol{K}_i}}{\sigma(\boldsymbol{K})}, \tag{10}$$

where $\boldsymbol{K}_i$ denotes the eigenvector of $\widetilde{M_{ii'}}$ according to [Ricardo et al. 2008], and $\sigma(\boldsymbol{K}_i)$ denotes the standard deviation of the $\boldsymbol{K}$ vector.

Similarly, the complexity of research fields can be modelled by field complexity index, FCI, where

$$FCI(f) = \frac{\boldsymbol{Q}_f - \overline{\boldsymbol{Q}_f}}{\sigma(\boldsymbol{Q})}. \quad (11)$$

where $\boldsymbol{Q}_f$ denotes the eigenvector of $\widetilde{M_{ff'}}$, following the same process of obtaining $\widetilde{M_{ii'}}$ for $k(i,n)$.

The measure of proximity is based on the conditional probability that an institutes has both reveal comparative advantage in both research fields f and f'. Proximity $\phi(f,f')$ for a pair of research fields f and f' can be expressed as:

$$\phi(f,f') = \frac{\sum_i \boldsymbol{M}_{if}\boldsymbol{M}_{if'}}{max\,(k(f,0) - k(f',0))}. \quad (12)$$

Subsequently, capability distance can be obtained. The concept of capability distance show how far away for an institute to pursue a research field it's not currently specialised (i.e. without revealed comparative advantage). The capability distance d_{if} is defined as:

$$d(i,f) = \frac{\sum_{f'}(1-\boldsymbol{M}_{if'})\phi(f,f')}{\sum_{f'}\phi(f,f')}. \quad (13)$$

The opportunity value OV of an institute i is defined as:

$$OV(i) = \sum_{f'}(1-d(i,f'))(1-M(i,f'))FCI(f') \quad (14)$$

and a higher opportunity value indicates that the institute has more research fields in close proximity or its research fields are more complex.

4. DATA PROCESSING

In this paper, we've chosen the Microsoft Academic Service (MAG)[Sinha et al. 2015] in our study. In comparison with other popular collections such as DataBase systems and Logic Programming (DBLP) and American Physical Society's Data Sets for Research (APS), with focuses on computer science and physics bibliography respectively, MAS covers most research disciplines. In comparison with popular multidisciplinary datasets from ISI Web of Science (WoS) and Google Scholar Citations, the full collection of the MAS dataset is open to the public to access and download through Microsoft Academic Graph (MAG). The MAG dataset was the official dataset recommended for the 2016 KDD CUP competition on predicting the publication acceptance and measuring the impact of research institutions [Sandulescu and Chiru 2016]. Although, it was also observed that MAG's popularity in the bibliometric research community has been declining [Orduña-Malea et al. 2014], and this paper has chosen the collection of an earlier year (2014) for analysis.

This paper focuses on the comparison of research complexity between top ranked universities in the world. While there are multiple world university rankings, each ranking system distinguishes itself from one another due to the choice of indicators and the ranking scores formulae. At the same time, these ranking systems exhibits a considerable overlap in the top 100 listed universities according to the study by Moed [Moed 2016].

Top 100 universities listed on the ARWU 2016 ranking outcome has been selected in our study. It should be noted that the selection of the ranking year doesn't need to match the publication year. One potential extension of the study could be the investigation of how research complexity influence subsequent university ranking, which is beyond the scope of this paper and may be considered for future work.

The format of academic publication may be considered semi-structured data, and MAG applies algorithms to automatically parse details such as author and institutions [Sinha et al. 2015]. In our study, institutional information is required, and according to our observation there are multiple affiliation IDs that may be associated with the same institution. It is observed that punctuation, such as the comma in "University of California, Los Angeles", is removed in the MAG data set. Thus, adjustment needs to be made for institution names extracted from ARWU. Apart from the challenge of automated information extraction, some institutes might have multiple well perceived (not necessarily official) names. For example, "University of Sydney", "The University of Sydney", "Sydney University" appear as separated entries and they are combined in our study. In addition, well-perceived abbreviations such as "MIT" for "Massachusetts Institute of Technology" are considered. Although, in this particular example, "MIT" can be a sub-string of another affiliation, therefore only the abbreviations that fit within word boundaries are taken. Lastly, we observed that some automatically parsed entries appear to be short biographies, and we remove affiliation entries with 300 characters or more. After the affiliation name adjustment, we still observe that some of these top ranked universities yield very low number of publications. Universities with less than 1000 publications in 2014 according to the MAG data set are excluded from our study.

Commercial bibliography tools such as WoS and SCImago provide classification codes for identifying the research fields of different journals. Unfortunately, classification codes are not available in the raw MAG dataset, which represents a constraint from adopting the MAG dataset for research complexity analysis, and it requires significant effort to classify the entire database. Instead of designing a new set of classification codes for the MAG dataset, sciMAG2015 adopts the codes from SCImago Journal Classification and labels all MAG journals [De Domenico et al. 2016]. In this paper, we use the 27 distinct area classification from sciMAG2015 for research complexity analysis.

5. INSTITUTIONAL RCI

The ubiquity-diversity scatter plot of the selected institutes is shown in Figure 1. The mean diversity $\overline{k_{i,0}}$ partition institutes into two groups with the number of fields with revealed comparative advantage higher or lower than the average, whereas the mean ubiquity $\overline{k_{i,1}}$ divides institutes specialised in niche or popular/standard research fields compares to the mean value. Combination of the two mean dividers partitions the scatter plots into four quadrants, and these quadrants represent:

- upper-left quadrant: non-diversified institutes publishing in standard areas;
- upper-right quadrant: diversified institutes publishing in standard areas;
- lower-left quadrant: non-diversified institutes publishing in exclusive areas;

- lower-right quadrant: diversified institutes publishing in exclusive areas.

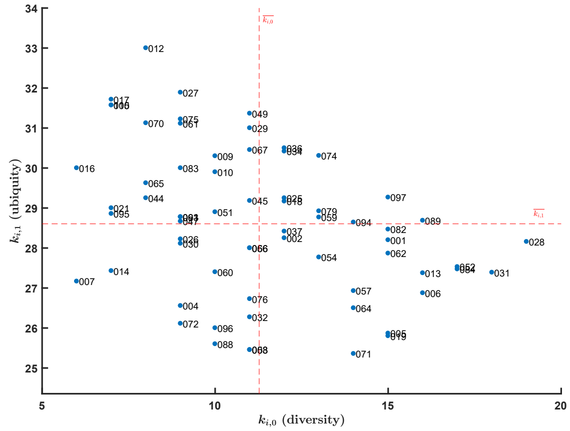


Figure 1: Ubiquity-diversity scatter plot

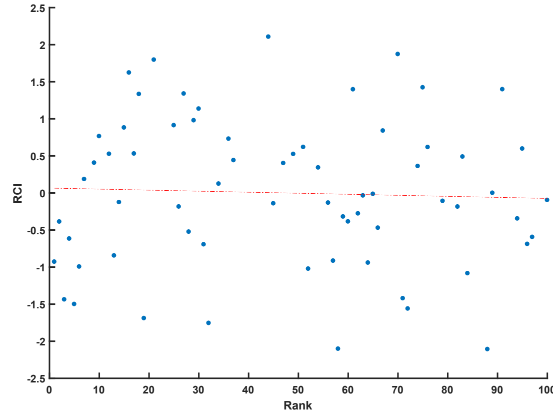
Figure 1 shows that most universities are scattered over the upper-left, lower-left, and lower-right quadrants, with 24 (or 36%), 15 (or 23%), and 17 (or 26%) universities, respectively. The upper-right quadrant has the least number of universities, with 10 (or 15%) universities found in this quadrant and most of them positioned close to the mean ubiquity or the man diversity line. According to the definition of diversity and ubiquity, the sparse upper-right quadrant shows that only a small number of universities have diversified research disciplines with high RCA values, and these research specialisation span across research fields with more universities also with high RCA values (i.e. standard areas that many universities claim specialisation.)

Compares Figure 1 with the matching plot for economic complexity by Hidalgo & Hausmann [Hidalgo and Hausmann 2009], there's a noticeable difference in the lower-left quadrant, with few economies scatter over this quadrant. While the economic complexity analysis covers most economies in the world, whereas the study in this paper focuses on top universities which is a small fraction of all universities in the world, the difference may be associated to the difference of the sample coverage. Investigations on universities with all ranking levels could be considered as a potential extension of this work.

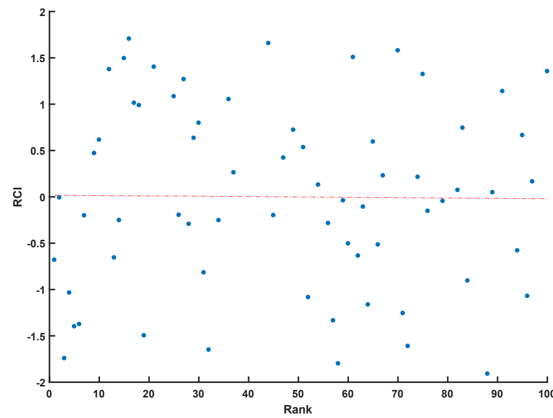
Table 1 shows the RCI and OV values of selected universities. The order of the list follows the ARWU 2016 ranks, with 66 universities due to incomplete data set for some universities, as explained in section 4.

Scatter plots of RCI against ARWU ranking are shown in Figure 2. In the experiment, different RCA threshold values are used (i.e. the threshold for determining whether an institute has revealed comparative advantage in a particular research field, according to Equation (2).) According to the first-order linear regression (dotted red line) plots, there's no clear evidence of correlation between RCI and the university rank. While the selected universities are mostly well-funded, extension of the study to cover smaller universities could be considered future work.

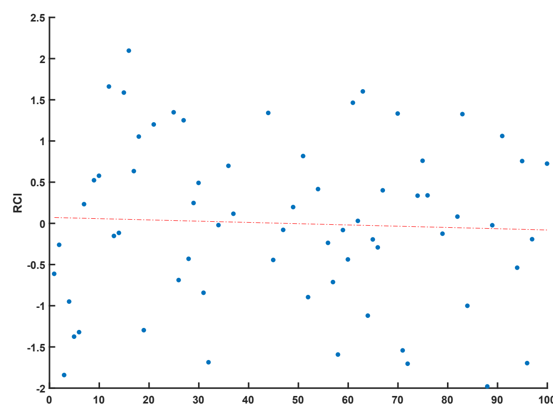
A scatter plot of opportunity value versus RCI can be found in Figure 3. Note, while part of the text labels in Fig-



(a) RCI with threshold $t = 0.75$



(b) RCI with threshold $t = 1.00$



(c) RCI with threshold $t = 1.25$

Figure 2: RCI-rank scatter plots with different threshold values

Table 1: RCI of top ranked universities

Rank	Institute Name	RCI	OV	Rank	Institute Name	RCI	OV
001	Harvard University	-0.68	-2.20	051	University of Munich	0.53	1.78
002	Stanford University	-0.01	1.00	052	University of Maryland, College Park	-1.08	-5.77
003	University of California, Berkeley	-1.74	-4.05	054	University of Zurich	0.13	1.33
004	University of Cambridge	-1.03	-1.43	056	University of Helsinki	-0.28	-0.21
005	Massachusetts Institute of Technology	-1.40	-6.25	057	University of Bristol	-1.33	-5.26
006	Princeton University	-1.37	-7.04	058	Tsinghua University	-1.80	-4.54
007	University of Oxford	-0.20	-0.03	059	University of California, Irvine	-0.04	0.68
009	Columbia University	0.47	1.66	060	Uppsala University	-0.50	-0.48
010	University of Chicago	0.61	1.88	061	Vanderbilt University	1.51	2.80
012	University of California, Los Angeles	1.38	2.29	062	Ghent University	-0.64	-1.81
013	Cornell University	-0.66	-2.92	063	McGill University	-0.11	0.45
014	University of California, San Diego	-0.25	0.06	064	Purdue University - West Lafayette	-1.16	-4.52
015	University of Washington	1.49	1.64	065	Aarhus University	0.59	1.22
016	Johns Hopkins University	1.70	1.26	066	Utrecht University	-0.52	-0.79
017	University College London	1.01	1.37	067	University of Oslo	0.23	1.54
018	University of Pennsylvania	0.99	3.72	070	University of Pittsburgh	1.58	2.21
019	ETH Zurich	-1.50	-6.90	071	Peking University	-1.26	-4.50
021	University of California, San Francisco	1.40	1.48	072	Nagoya University	-1.61	-2.51
025	Duke University	1.08	4.04	074	University of Groningen	0.21	1.99
026	Northwestern University	-0.20	0.36	075	Boston University	1.32	2.77
027	University of Toronto	1.27	2.71	076	University of California, Davis	-0.15	0.11
028	University of Wisconsin - Madison	-0.29	-0.28	079	Monash University	-0.05	1.14
029	New York University	0.63	2.65	082	University of Sydney	0.07	1.67
030	University of Copenhagen	0.80	1.65	083	McMaster University	0.74	1.74
031	UIUC	-0.82	-4.56	084	National University of Singapore	-0.91	-4.87
032	Kyoto University	-1.65	-4.04	088	Moscow State University	-1.91	-3.92
034	University of British Columbia	-0.25	0.08	089	The Hebrew University of Jerusalem	0.05	1.76
036	UNC	1.05	4.06	091	University of Florida	1.14	2.20
037	Rockefeller University	0.26	1.62	094	KU Leuven	-0.58	-1.01
044	Karolinska Institute	1.66	2.21	095	Leiden University	0.66	0.92
045	The University of Texas at Austin	-0.20	0.31	096	Osaka University	-1.07	-1.91
047	Heidelberg University	0.42	1.26	097	Rutgers	0.16	1.88
049	University of Southern California	0.72	2.91	100	University of Utah	1.35	1.70

ure 3 appear cluttered (especially for the points with RCI values close to zero), the values of ranking-RCI pairs can also be found in Table 1. Figure 3 illustrates a high degree of correlation between opportunity value and RCI. As shown in the first and second order regression, opportunity value well fit as a quadratic function of RCI. The third order regression is an over-fit and it is not included in 3 to increase the readability. According the definition in Equation (14), institutes with high RCI values has more research fields within close proximity to pursue, or its current research fields are considered complex. The plot does not exhibit associations between university rank to the OV-RCI pairs (i.e. there's no cluster of closely ranked universities), similar to the observation in Figure 2 that no evidence of associate between the university rank has been found.

Figure 4 shows the geographical distribution of RCI values. The colours represent the "band" of the RCI values, with markers in green for $RCI > 0.5$, red for $RCI < -0.5$, and yellow otherwise. The size of the marker represents the magnitude (or value) of RCI. The geographical longitude and latitude details of the selected universities are extracted from Google search as well as from the Wikipedia. It should be noted that most yellow markers are too small to display due to their small magnitude/value (for example,

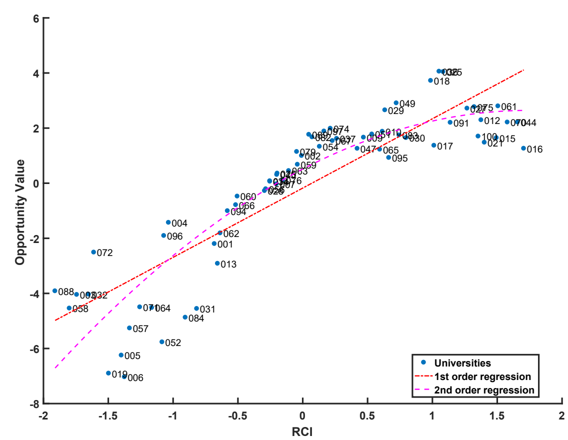


Figure 3: Opportunity Value vs RCI scatter plot

the two Australian universities in Table 1), despite their positive/negative signs. It is observed that top universities in North America in general have high positive RCI values, Asian universities have the other extreme with very low negative RCI values, whereas European universities exhibit to have a good mixture of both. The difference could be due to a number of factors, such as diversification as a proxy of the university age, the composition of the economy which leads to bias in discipline-based research funds. These factors are beyond the scope of this paper, and further investigations may be considered in the future work.

Further analysis, such as other factors associate to the RCI, as well as the relevance of RCI to the university ranking over time, will be considered the future work of this paper.

6. CONCLUSIONS

This paper examined research complexity based on the research outputs of top ranked university around the world, From the study of diversity and ubiquity, most of the selected universities are either non-diversified institutes, or diversified institute publishing in exclusive areas. From the study of research complexity analysis, there was no association observed between RCI values and the institution's ranking. On the other hand, opportunity value highly correlates with RCI according the scatter plot, suggesting that high RCI institutes has more research fields within close proximity, or its current research fields are considered complex. According to geographical distribution, it was observed that most North America institutes possesses high positive RCI values, Asian institutes possess low negative RCI values, and European institutes have a good mixture of both. The observation suggested further investigations are required of the underlying factors.

7. REFERENCES

[Börner et al. 2012] Katy Börner, Richard Klavans, Michael Patek, Angela M. Zoss, Joseph R. Biberstine, Robert P. Light, Vincent Larivière, and Kevin W. Boyack. 2012. Design and Update of a Classification System: The UCSD Map of Science. *PLOS ONE* 7, 7 (July 2012), e39464.

[Cimini et al. 2014] Giulio Cimini, Andrea Gabrielli, and Francesco Sylos Labini. 2014. The Scientific Competitiveness of Nations. *PLOS ONE* 9, 12 (Dec. 2014), e113470.

[Daraio et al. 2015] Cinzia Daraio, Andrea Bonaccorsi, and Léopold Simar. 2015. Efficiency and economies of scale and specialization in European universities: A directional distance approach. *Journal of Informetrics* 9, 3 (July 2015), 430–448.

[De Domenico et al. 2016] Manlio De Domenico, Elisa Omodei, and Alex Arenas. 2016. Quantifying the Diaspora of Knowledge in the Last Century. *Applied Network Science* 1, 15 (2016).

[Guevara et al. 2016] Miguel R. Guevara, Dominik Hartmann, Manuel Aristarán, Marcelo Mendoza, and César A. Hidalgo. 2016. The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* 109, 3 (Dec. 2016), 1695–1709.

[Harzing and Giroud 2014] Anne-Wil Harzing and Axèle Giroud. 2014. The competitive advantage of nations: An application to academia. *Journal of Informetrics* 8, 1 (Jan. 2014), 29–42.

[Hidalgo and Hausmann 2009] César A. Hidalgo and Ricardo Hausmann. 2009. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences* 106, 26 (June 2009), 10570–10575.

[Leydesdorff and Rafols 2009] Loet Leydesdorff and Ismael Rafols. 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology* 60, 2 (Feb. 2009), 348–362.

[Li et al. 2015] Feng Li, Yajun Miao, and Jing Ding. 2015. Tracking the development of disciplinary structure in China's top research universities (1998–2013). *Research Evaluation* 24, 3 (July 2015), 312–324.

[López-Illescas et al. 2011] Carmen López-Illescas, Félix de Moya-Anegón, and Henk F. Moed. 2011. A ranking of universities should account for differences in their disciplinary specialization. *Scientometrics* 88, 2 (Aug. 2011), 563–574.

[Moed 2016] Henk F. Moed. 2016. A critical comparative analysis of five world university rankings. *Scientometrics* (Dec. 2016), 1–24.

[Murray et al. 2016] Dennis L. Murray, Douglas Morris, Claude Lavoie, Peter R. Leavitt, Hugh MacIsaac, Michael E. J. Masson, and Marc-Andre Villard. 2016. Bias in Research Grant Evaluation Has Dire Consequences for Small Universities. *PLOS ONE* 11, 6 (June 2016), e0155876.

[Orduña-Malea et al. 2014] Enrique Orduña-Malea, Alberto Martín-Martín, Juan M. Ayllon, and Emilio Delgado Lopez-Cozar. 2014. The silent fading of an academic search engine: the case of Microsoft Academic Search. *Online Information Review* 38, 7 (2014), 936–953.

[Ortega et al. 2011] José Luis Ortega, Elena López-Romero, and Inés Fernández. 2011. Multivariate approach to classify research institutes according to their outputs: The case of the CSIC's institutes. *Journal of Informetrics* 5, 3 (July 2011), 323–332.

[Pastor and Serrano 2016] José Manuel Pastor and Lorenzo Serrano. 2016. The determinants of the research output of universities: specialization, quality and inefficiencies. *Scientometrics* 109, 2 (Nov. 2016), 1255–1281.

[Payumo et al. 2014] Jane G. Payumo, Prema Arasu, Anas Miftah Fauzi, Iskandar Zulkarnaen Siregar, and Deni Noviana. 2014. An entrepreneurial, research-based university model focused on intellectual property management for economic development in emerging economies: The case of Bogor Agricultural University, Indonesia. *World Patent Information* 36 (March 2014), 22–31.

[Ricardo et al. 2008] Hausmann Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jimenez, Alexander Simoes, and Muhammed A. Yıldırım. 2008. The atlas of economic complexity: Mapping paths to prosperity. *The Observatory of Economic Complexity. Harvard HKS/CDIMIT Media Lab* (2008).

[Robinson-García and Calero-Medina 2014] Nicolás Robinson-García and Clara Calero-Medina. 2014.

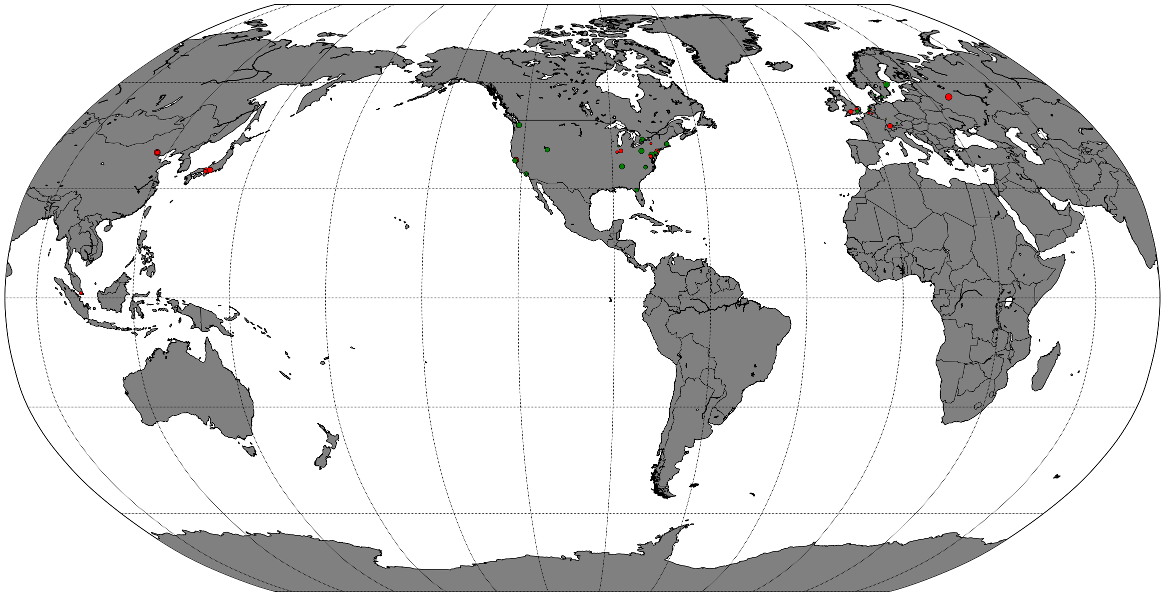


Figure 4: Geographical RCI distribution

What do university rankings by fields rank? Exploring discrepancies between the organizational structure of universities and bibliometric classifications. *Scientometrics* 98, 3 (March 2014), 1955–1970.

[Sandulescu and Chiru 2016] Vlad Sandulescu and Mihai Chiru. 2016. Predicting the future relevance of research institutions - The winning solution of the KDD Cup 2016. (Sept. 2016).

[Sinha et al. 2015] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. 2015. An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 243–246.

[Wong and Singh 2010] Poh Kam Wong and Annette Singh. 2010. University patenting activities and their link to the quantity and quality of scientific publications. *Scientometrics* 83, 1 (April 2010), 271–294.