On the Correlation Between Research Complexity and Academic Competitiveness

Chapter · November 2020

DOI: 10.1007/978-3-030-64452-9_39

5 authors, including:

Lei Wang
BMW Group
141 PUBLICATIONS **4,363** CITATIONS
SEE PROFILE

Feng Xia
Federation University Australia
464 PUBLICATIONS **11,811** CITATIONS
SEE PROFILE

On the Correlation between Research Complexity and Academic Competitiveness

Jing Ren[1], Ivan Lee[2], Lei Wang[1], Xiangtai Chen[1], and Feng Xia[3]

[1] School of Software, Dalian University of Technology, Dalian 116620, China
[2] STEM, University of South Australia, Adelaide, SA 5001, Australia
[3] School of Engineering, IT and Physical Sciences, Federation University Australia, Ballarat, VIC 3353, Australia
ch.yum@outlook.com; ivan.lee@unisa.edu.au; leonard_wl@outlook.com; chenxiangtai@outlook.com; f.xia@ieee.org

Abstract. Academic capacity is a common way to reflect the educational level of a country or district. The aim of this study is to explore the difference between the scientific research level of institutions and countries. By proposing an indicator named Citation-weighted Research Complexity Index (CRCI), we profile the academic capacity of universities and countries with respect to research complexity. The relationships between CRCI of universities and other relevant academic evaluation indicators are examined. To explore the correlation between academic capacity and economic level, the relationship between research complexity and GDP per capita is analysed. With experiments on the Microsoft Academic Graph data set, we investigate publications across 183 countries and universities from the Academic Ranking of World Universities in 19 research fields. Experimental results reveal that universities with higher research complexity have higher fitness. In addition, for developed countries, the development of economics has a positive correlation with scientific research. Furthermore, we visualize the current level of scientific research across all disciplines from a global perspective.

Keywords: Big Scholarly Data · Correlation Analysis · Research Complexity · Academic Competitiveness · Scientometrics

1 Introduction

Research output assessment and analysis have always been the essential component in most academic ranking systems [6, 10]. Nowadays, no matter for universities or countries, academic research is playing a significant role in improving international competitiveness. Specifically, knowledge innovation is regarded as the core element among national competitiveness [3]. High achievement in research activity can help universities apply for research funds and projects, promote academic and industrial collaborations, as well as attract talents [5]. Thus, evaluating current scientific research from diverse perspectives is crucial for both universities and countries.

Over the last couple of decades, many institutions have applied bibliometrics to evaluate the academic performance of universities [14, 15]. A common approach of evaluating academic impact is university ranking. Common world university ranking systems, such as Quacquarelli Symonds World University Ranking (QS), Shanghai Jiao Tong Academic Ranking of World Universities (ARWU), and Times Higher Education World University Rankings (THE), all regard research achievement as a major impact factor in their ranking algorithms [12, 11]. Therefore, the rankings of these algorithms are similar with each other every year.

In order to measure the comprehensive capacity of academic research, Research Complexity Index (RCI) was propsed to profile universities [9]. However, this indicator only takes the number of publications into consideration, while regardless of their impacts. Therefore, we propose Citation-weighted Research Complexity Index (CRCI), which imposes citation as weight to denote publications instead of the simple count of publications. Comparing with the previous studies in academic capacity analysis, this paper proposes a new indicator, from the perspective of research complexity instead of research productions, to profile universities and countries. Major contributions of this paper include:

1. Proposing an indicator Citation-weighted Reserch Complexity Index (CRCI) to profile the current scientific research of universities and countries from a new perspective.
2. Exporing the relationship between research complexity with fitness and opportunity value of universities, and GDP per capita of countries.
3. Comparing the research complexity difference of countries in different fields.
4. Visualizing the geographical distribution of research complexity of 64 top-ranked universities and 183 countries.

2 Revealed Symmetric Comparative Advantage

Except for research production, it is also important to consider research specialisation when profiling an institution or country. Revealed Comparative Advantage (RCA) [1] in this paper can be understood as a relative indicator to judge whether the academic capacity is above average. In this paper, multi-disciplinary papers in the data set are labeled with different fields and subfields. As for papers labeled with more than one research field, the weight is allocated to each research field equally. Let u denote university, and f denote research field. Let $Z(u, f)$ denote a set including all papers $p(u, f)$ published by university u in research field f. Considering that a paper may be cited by different number of publications, a weighting factor $n_{p(u,f)}$ is applied, denoting that paper $p(u, f)$ is cited by n publications. Therefore, the academic achievement $\boldsymbol{P}_{uf}$ of university u in research field f can be calculated as:

$$\boldsymbol{P}_{uf} = \sum_{p(u,f) \in Z(u,f)} n_{p(u,f)}. \tag{1}$$

According to $\boldsymbol{P}_{uf}$, the revealed comparative advantage $\boldsymbol{RSCA}_{uf}$ [7] for university u in research field f can be calculated as

$$\boldsymbol{RSCA}_{uf} = \frac{\boldsymbol{P}_{uf}\sum_{u,f}\boldsymbol{P}_{uf} - \sum_{u}\boldsymbol{P}_{uf}\sum_{f}\boldsymbol{P}_{uf}}{\boldsymbol{P}_{uf}\sum_{u,f}\boldsymbol{P}_{uf} + \sum_{u}\boldsymbol{P}_{uf}\sum_{f}\boldsymbol{P}_{uf}}. \tag{2}$$

When we calculate the corresponding variables for countries, u denoting universities in equations can be replaced by c denoting countries. All equations corresponding to universities are suitable for countries as well.

3 Citation-weighted Research Complexity Index

In order to study the complexity and diversity of academic research, research complexity index and opportunity value [9] are adopted to analyze the research specialisation of universities in different disciplines. According to the definition in Section 2, the bool value of M_{uf} indicates whether university u has revealed comparative advantage in field f. Diversity of university D_u is initialized to the number of research fields having revealed comparative advantage:

$$D_u = k(u,0) = \sum_{f}\boldsymbol{M}_{uf}. \tag{3}$$

Similarly, ubiquity of field U_f can be regarded as the number of universities who have revealed comparative advantage in field f, which is formulated as Eq. 4:

$$U_f = k(f,0) = \sum_{u}\boldsymbol{M}_{uf}. \tag{4}$$

Considering the symmetry of the university-publication bipartite network, *Method of Reflections* [4] can be adopted to calculate $k(u,n)$ in terms of $k(f,n-1)$:

$$k(u,n) = \frac{1}{k(u,0)}\sum_{f}\boldsymbol{M}_{uf}k(f,n-1). \tag{5}$$

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

$$k(f,n) = \frac{1}{k(f,0)}\sum_{u}\boldsymbol{M}_{uf}k(u,n-1). \tag{6}$$

The citation-weighted research complexity index of a university $CRCI(u)$, is calculated as:

$$CRCI(u) = \frac{\boldsymbol{K}_u - \overline{\boldsymbol{K}_u}}{\sigma(\boldsymbol{K}_u)}, \tag{7}$$

where $\boldsymbol{K}_u$ denotes the eigenvector of $\widetilde{M}_{uu'}$ associated with the second largest eigenvalue [2], and $\sigma(\boldsymbol{K}_u)$ denotes the standard deviation of the $\boldsymbol{K_u}$ vector. The opportunity value OV of a university u is defined as:

$$OV(u) = \sum_{f'}(1-d(u,f'))(1-M(u,f'))CFCI(f'). \tag{8}$$

According to the definition of opportunity value, universities with high opportunity value denotes having more research fields in close proximity, or their research fields can be complex easier than universities with low opportunity value.

4 Fitness and Complexity

The raw fitness value of an university and the raw complexity value of a field can be obtained as follows:

$$\widetilde{F}_u^{(n+1)} = \sum_f M_{uf} Q_f^{(n)}, \tag{9}$$

$$\widetilde{Q}_f^{(n+1)} = \left[\sum_u M_{uf} (F_u^{(n)})^{-1} \right]^{-1}, \tag{10}$$

where $\widetilde{F}_u^{(n+1)}$ denotes the raw fitness value of a university u, and $\widetilde{Q}_f^{(n+1)}$ denotes the raw complexity value of a field at the n+1-th iteration. $F_u^{(n)}$ and $Q_f^{(n)}$ are the normalised values of fitness and complexity respectively. They are defined as:

$$F_u^{(n)} = \frac{\widetilde{F}_u^{(n)}}{N_u^{-1} \sum_u \widetilde{F}_u^{(n)}}, \tag{11}$$

$$Q_f^{(n)} = \frac{\widetilde{Q}_f^{(n)}}{N_f^{-1} \sum_f \widetilde{Q}_f^{(n)}}, \tag{12}$$

where N_u and N_f are the total number of universities and research fields, respectively. The initial values of fitness $F_u^{(0)}$ and complexity $Q_f^{(0)}$ are set to 1.

According to the calculation formula mentioned above, fitness can be understood as how many complex research fields a university has revealed comparative advantage [8]. Universities possessing more complex research fields have higher fitness value than univerisities with less. The fitness value is also influenced by the complexity degree of research fields. In turn, Complexity of a research field depends on how many universities have revealed comparative advantage. For example, the complexity of a research field with only one university possessing revealed comparative advantage would be higher than that of two.

5 Experiments

In this section, Microsoft Academic Graph (MAG) [13] is used to analyze research output. Besides, we compare the difference of CRCI with other methods of academic capacity evaluation, and carry out several analysis and visualizations on the ranking results.

5.1 Data of universities and countries

During the period of 2003-2017, 64 universities that are always located in the top 100 list of (ARWU) are chosen as the experimental targets in this paper. As for the countries, all countries involved in MAG (183 totally) are used to observe the difference of their research complexity. The year of publications selected in this paper range from 1970 to 2017. The gross domestic product per capita of the country is obtained from THE WORLD BANK[4].

5.2 Experimental Results

In Fig. 1a, a strong correlation between fitness and CRCI can be observed. By fitting a curve of the second order regression, the value of fitness can well fit as a quadratic function of CRCI with R^2 equal to 0.46 approximately. The result shows that universities with higher research complexity have higher value of fitness, meaning that a university with strong comprehensive strength is more competitive. In Fig. 1b, the correlation between opportunity value and CRCI well fit as a quadratic function with the openning down, where R^2 equal to 0.58 approximately. This result shows that, as for the universities with lower research complexity, they have more opportunies to improve their research complexity by means of breaking through new research fields than others. The scatter plots of CRCI versus GDP_{pc} is shown in Fig. 1c. With R^2 equal to 0.21, GDP_{pc} fit as a first order regression of CRCI for only developed countries, indicating that the scientific research of university can impact a country's economic development.

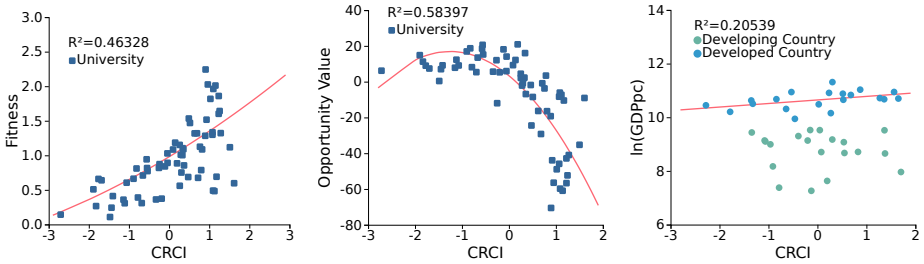

Fig. 1. (**a**) Scatter plot of Fitness against CRCI of universities. (**b**) Scatter plot of opportunity value against CRCI of universities. (**c**) Scatter plot of $ln(GDP_{pc})$ against CRCI of countries.

Visualization The geographical distribution of CRCI values from 64 top universities and 183 countries is shown in Fig. 2. The colour of countries represent

[4] https://www.worldbank.org

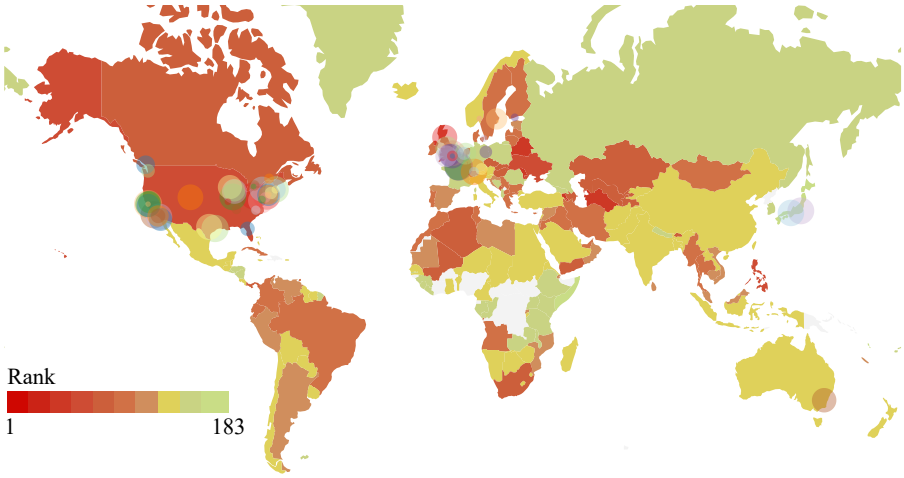

Fig. 2. The geographical distribution of 64 top-ranked universities and research complexity of 183 countries in 2015.

the value of CRCI, with red denoting high research complexity of countries. Besides, the colorful bubbles denotes different universities, with larger size representing high research complexity. It should be noted that the color of bubble just for better differentiation, without any practical meanings. The geographical longitude and latitude of the selected universities are extracted from Google. From observation, top-ranked universities mainly located in countries with high levels of research complexity.

6 Conclusion

This paper has proposed an indicator to profile the academic capacity of universities and countries from the perspective of research complexity. By profiling the scientific research of acadmic entities from various perspectives, this method can empower explorers with comprehensive and valuable information that will encourage them to take on the challenge, thereby speeding up the process of academic and economic development. Through the analysis of the competitive advantage of countries in different fields, we can see clearly which field need to be improved and extended, and which field has been studied throughly by many countries.

In future work, this indicator could be extended to profile other groups of universities, such as the intra-national universities. Besides, the research complexity can also be applied to offer insights to fund allocation, policy decision and brain gain.

References

1. Balassa, B.: Trade liberalisation and revealed comparative advantage 1. The manchester school **33**(2), 99–123 (1965)
2. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., Pietronero, L.: Measuring the intangibles: A metrics for the economic complexity of countries and products. PloS one **8**(8), e70726 (2013)
3. Delgado, M., Ketels, C., Porter, M.E., Stern, S.: The determinants of national competitiveness. Tech. rep., National Bureau of Economic Research (2012)
4. Hidalgo, C.A., Hausmann, R.: The building blocks of economic complexity. Proceedings of the national academy of sciences **106**(26), 10570–10575 (2009)
5. Kong, X., Shi, Y., Wang, W., Ma, K., Wan, L., Xia, F.: The evolution of turing award collaboration network: Bibliometric-level and network-level metrics. IEEE Transactions on Computational Social Systems **6**(6), 1318–1328 (2019)
6. Kong, X., Shi, Y., Yu, S., Liu, J., Xia, F.: Academic social networks: Modeling, analysis, mining and applications. Journal of Network and Computer Applications **132**, 86–103 (2019)
7. Laursen, K.: Revealed comparative advantage and the alternatives as measures of international specialization. Eurasian Business Review **5**(1), 99–115 (2015)
8. Lee, I., Tie, Y.: Fitness and research complexity among research-active universities in the world. IEEE Transactions on Emerging Topics in Computing (2018)
9. Lee, I., Xia, F., Roos, G.: An observation of research complexity in top universities based on research publications. In: Proceedings of the 26th International Conference on World Wide Web Companion. pp. 1259–1265. International World Wide Web Conferences Steering Committee (2017)
10. Ren, J., Wang, L., Wang, K., Yu, S., Hou, M., Lee, I., Kong, X., Xia, F.: API: An index for quantifying a scholar's academic potential. IEEE Access **7**, 178675–178684 (2019)
11. Robinson-Garcia, N., Torres-Salinas, D., Herrera-Viedma, E., Docampo, D.: Mining university rankings: Publication output and citation impact as their basis. arXiv preprint arXiv:1905.07141 (2019)
12. Selten, F., Neylon, C., Huang, C.K., Groth, P.: A longitudinal analysis of university rankings. arXiv preprint arXiv:1908.10632 (2019)
13. Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.J., Wang, K.: An overview of microsoft academic service (mas) and applications. In: Proceedings of the 24th international conference on world wide web. pp. 243–246 (2015)
14. Xia, F., Wang, W., Bekele, T.M., Liu, H.: Big scholarly data: A survey. IEEE Transactions on Big Data **3**(1), 18–35 (2017)
15. Yu, S., Liu, Y., Ren, J., Bedru, H.D., Bekele, T.M., Wan, L., Xia, F.: Mining key scholars via collapsed core and truss. In: 2019 IEEE Intl Conf on Cloud and Big Data Computing. pp. 305–308. IEEE (2019)