

# Geographical distribution of high-novelty research

Kuniko MATSUMOTO\*, \*\*

*\*k-matsumoto@nistep.go.jp*

Center for S&T Foresight and Indicators, National Institute of Science  
and Technology Policy (NISTEP), Japan

*\*\*kuniko.matsumoto@oecd.org*

Science and Technology Policy Division Directorate for Science, Technology and Innovation, OECD

In this study, trial analyses using bibliometric approaches were performed to investigate the geographical distribution of high-novelty research. Data on approximately 2.55 million academic papers published in 2021 were examined as a pilot to show worldwide statistical data on novelty research. A combinatorial novelty indicator measuring units comprising paired reference papers was adopted in the analyses. This study shows the main three results: the top 20 countries/regions in the top 10% of high-novelty papers, the share of the top 10% high-novelty papers in each country/region, and the share of the top 10% high-novelty papers by field in China and the United States, which contribute globally to the top 10% of high-novelty papers.

## 1. Introduction

For many decades, citation count has been considered the main bibliographic indicator for evaluating the quality of research, relying on the general assumption that it reflects the impact of a scientific publication. However, multifaceted evaluations are necessary because many arguments highlight the limitations of the use of citation counts alone (Baird & Oppenheim, 1994; MacRoberts & MacRoberts, 1996).

Novelty is an important aspect of research evaluation. Research using novel approaches is often said to drive breakthroughs in innovation research. Research that adopts a novel approach has a higher potential for major impact, even though it also faces a higher level of impact uncertainty (Wang et al., 2017). Highest-impact science is primarily grounded in features that introduce novel combinations into familiar knowledge domains (Uzzi et al., 2013). Novelty can be an indicator of potential breakthroughs, and a range of indicators to assess research novelty have been proposed in previous bibliometrics or scientometrics research. Worldwide statistical data that focus on research novelty, which allows comparison by scientific field or country, are useful for both scientists and science and technology policymakers to determine the status of global research activities. However, comprehensive statistical reports on research novelty through analysis using large-scale datasets, such as those consisting of more than one million datasets, are scarce. One of the rare study cases shows the decline in disruptiveness across six decades in academic research as observed through analysis using data from 45 million papers and 3.9 million patents (Park et al., 2023)

In this study, trial analyses using a bibliometric approach were carried out to investigate the geographical distribution of high-novelty research, such as the top 20 countries/regions in the top 10% high-novelty papers and the share of the top 10% high-novelty papers in each country/region, using data on approximately 2.55 million academic papers published in 2021 (the latest year).

## 2. Data and Methods

### 2.1. Measuring novelty of scientific publication

One of the main approaches to assessing research novelty is focusing on new combinations of knowledge sources, that is, the combinatorial novelty literature (Wang et al., 2017). Combinatorial novelty indicators are typically measured using pairs of reference papers, journals, or keywords as units. In this analysis, a combinatorial novelty indicator measuring units comprising paired reference papers was adopted (Matsumoto et al, 2021). The use of paired reference papers is advantageous because it discerns more elaborate and unusual combinations of existing knowledge.

The indicator quantifies how unusual the combinations of knowledge references in the focal publication are among the pre-existing combinations in its knowledge domain, determined by two conditions: at least one of the references of the focal paper and those whose field classification<sup>1</sup> completely matches that of the focal paper. The degree of citation similarity is referred to as the overlap score (OS). The overlap score  $OS_{ij}$  is then defined as the count of documents cited by both  $i$  and  $j$  divided by the sum of the unique citations in  $i$  or  $j$ . The novelty score of a focal paper,  $i$ , was calculated by subtracting the mean overlap score for papers in the same domain from 1. The resulting measure of the indicator may range between 0 and 1, where 0 indicates completely identical citation patterns to same-domain papers and 1 indicates completely dissimilar citation patterns.

In this analysis, the top 10% of high-novelty papers indicate research high in novelty. It refers to papers with a standardised novelty score in the top 10%. As the novelty scores indicators are calculated to be close to 1 (see Table 1) differences in citation patterns across fields may affect novelty scores and standardisation of the score is carried out by field.

Table 1. Descriptive statistics of novelty score by fields

#### a) Fields with top 3 average

Field	n	average	sd	min	max
Chemistry	257,952	0.976	0.022	0.000	0.999
Biochemistry, Genetics, and Molecular Biology	299,806	0.976	0.024	0.000	0.999
Chemical Engineering	159,451	0.975	0.024	0.000	0.998

#### b) Fields with bottom 3 average

Field	n	average	sd	min	max
Decision Sciences	49,788	0.953	0.044	0.000	0.997
Mathematics	186,414	0.937	0.069	0.000	0.999
Arts and Humanities	66,741	0.931	0.089	0.000	0.999

### 2.2. Bibliometric data

This analysis retrieved the data used in the novelty score calculation from the Scopus Custom Data, extracted in December 2022. It focused on all academic papers published in 2021 (the latest year in the dataset), which is approximately 2.55 million. Academic papers refer to articles and conference papers in journals and conference papers in conference proceedings.

<sup>1</sup> All Science Journal Classification (ASJC), the smallest science field unit in Scopus Custom Data used in this analysis, was adopted to identify papers sharing the same domain as the focal paper.

In the analysis by country/region, the number of papers published for each country/region was calculated using fractional counting. In the field analysis, subject areas assigned to Scopus Custom Data were used as field data; there were 27 fields.

### 3. Results

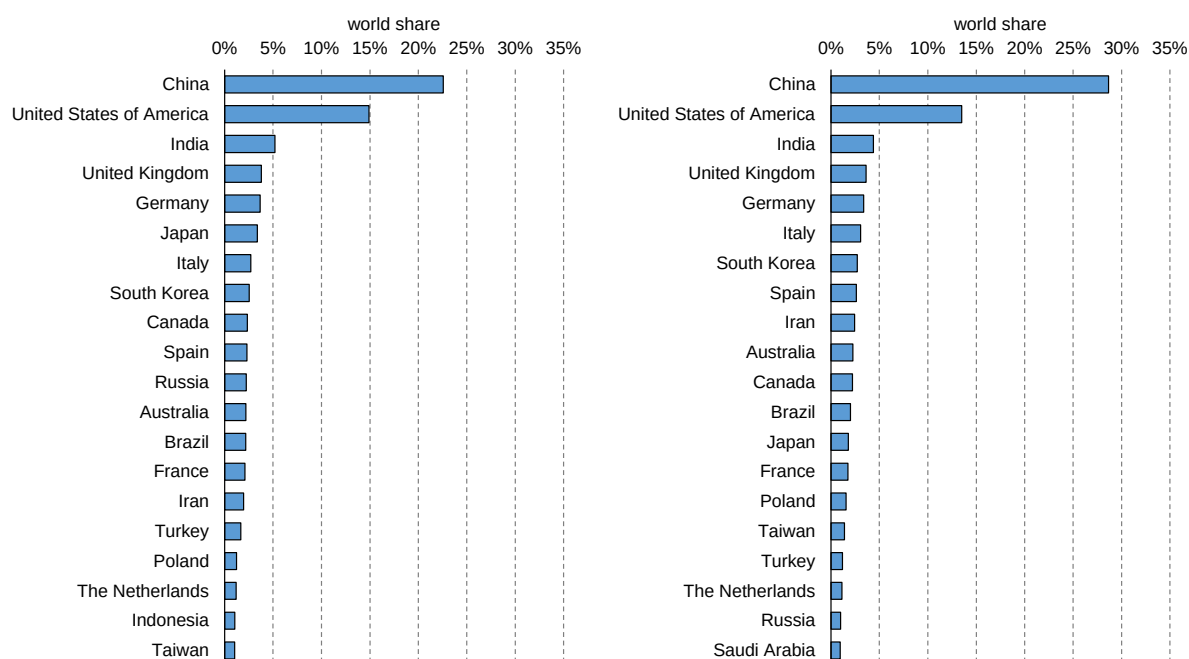
#### 3.1. Top 20 countries/regions in the number of papers, top 10% high-novelty papers in 2021

China has the highest share of papers in 2021 worldwide, which is approximately 23% of the total papers in 2021 (see Fig. 1). The United States has a 14.9% share followed by India with 5.2 %. These countries/regions also record a high share of the top 10% high-novelty papers.

When comparing the top 20 countries/regions in the number of papers with those of the top 10% high-novelty papers, most of the top 20 countries/regions in the number of papers fall into the top 20 countries/regions in the top 10% high-novelty papers. However, a slight change in ranking was observed. Japan and Russia have fallen in the world share ranking of the top 10% of high-novelty papers by more than five positions, whereas Iran has risen. Indonesia, which ranks 19th in the world in the number of papers published, is out of the top 20 countries/regions in the top 10% of high-novelty papers. Saudi Arabia has replaced Indonesia on the list.

Figure 1: Top 20 countries/regions in 2021

a) by the number of papers                      b) by the number of top 10% of high-novelty papers



Source: Author calculations based on Scopus Custom Data extracted in December 2022.

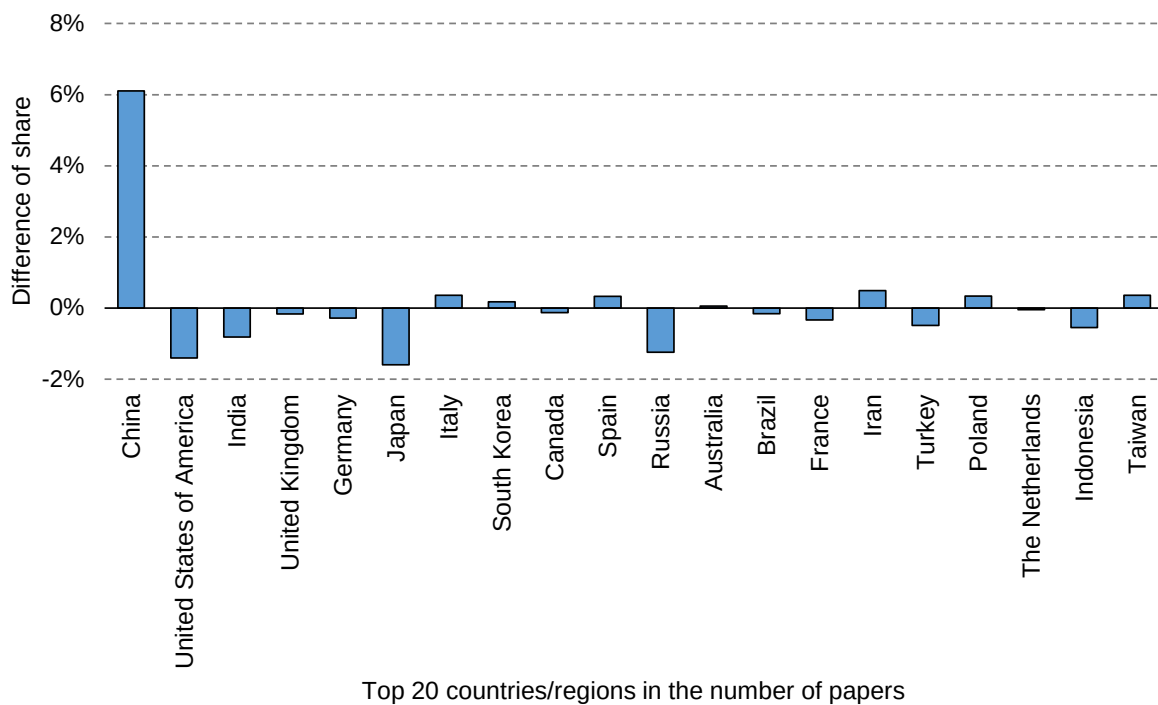
Note 1) Papers refer to articles, conference papers in journals, and conference papers in conference proceedings.

2) Focused on papers published in 2021.

3) The number of papers for each country/region is calculated by fractional counting.

The share in the number of papers of eight of the top 20 countries/regions in the world have a higher world share in the top 10% high-novelty papers than the world share in the number of papers (see Fig 2). Specifically, the world share of the number of top 10% high-novelty papers in China was approximately 6% higher than the world share of the number of papers. In contrast, Japan, the United States, and Russia are countries/regions where the world share of the number of papers is more than 1% higher than the world share in the top 10%.

Figure 2: Difference of world share between the number of papers and the number of top10% high-novelty papers in 2021



Source: Author calculations based on Scopus Custom Data extracted in December 2022.

Note 1) Papers refer to articles, conference papers in journals, and conference papers in conference proceedings.

2) Focused on papers published in 2021.

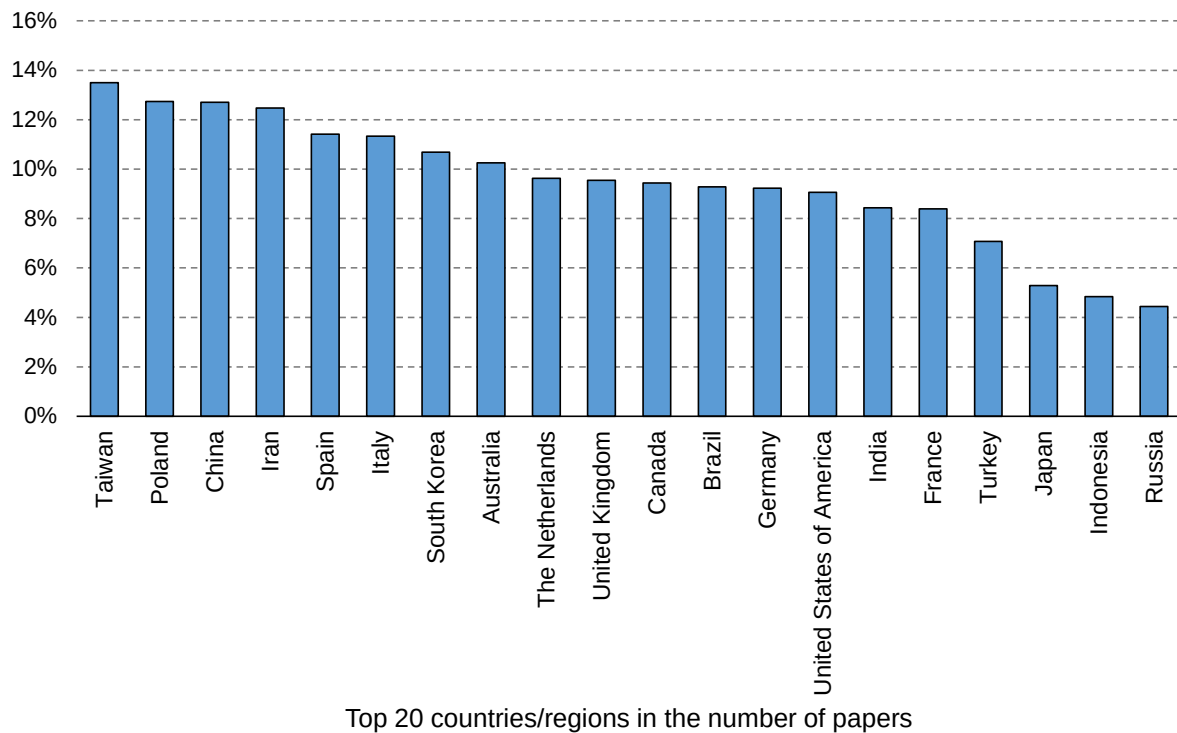
3) The number of papers for each country/region is calculated by fractional counting.

4) Difference of world share is calculated by subtracting world share of papers from world share of top10% high-novelty papers.

### 3.2. Share of top 10% high-novelty papers within each country/region

Among the top 20 countries/regions in terms of the number of papers, the percentage of the top 10% of high-novelty papers within each country/region varied between 4% and 14% (see Fig. 3). This indicator measures the degree of novelty of publications in a given country and year. It is calculated as the ratio of the number of papers in the top 10% of high-novelty papers worldwide versus the total number of papers in the country that year. Taiwan leads globally, with 13.5% of its papers among the top 10% high-novelty papers. Poland is second (12.74%), followed by China (12.71%) and Iran (12.47%). In contrast, Japan, Indonesia, and Russia account for less than 6% of their papers in the top 10% of high-novelty papers. The United States, the second largest contributor to the top 10% of high-novelty papers, is in the 14th place (9.06%).

Figure 3: Share of the top 10% high-novelty papers in each country/region in 2021



Source: Author calculations based on Scopus Custom Data extracted in December 2022.

Note 1) Papers refer to articles, conference papers in journals, and conference papers in conference proceedings.

2) Focused on papers published in 2021.

3) Share of papers in the top 10% of high-novelty papers worldwide by the total number of papers in the country/region. The number of papers published in each country/region was calculated by fractional counting.

### 3.3. Share of the top 10% high-novelty papers in each country/region by field

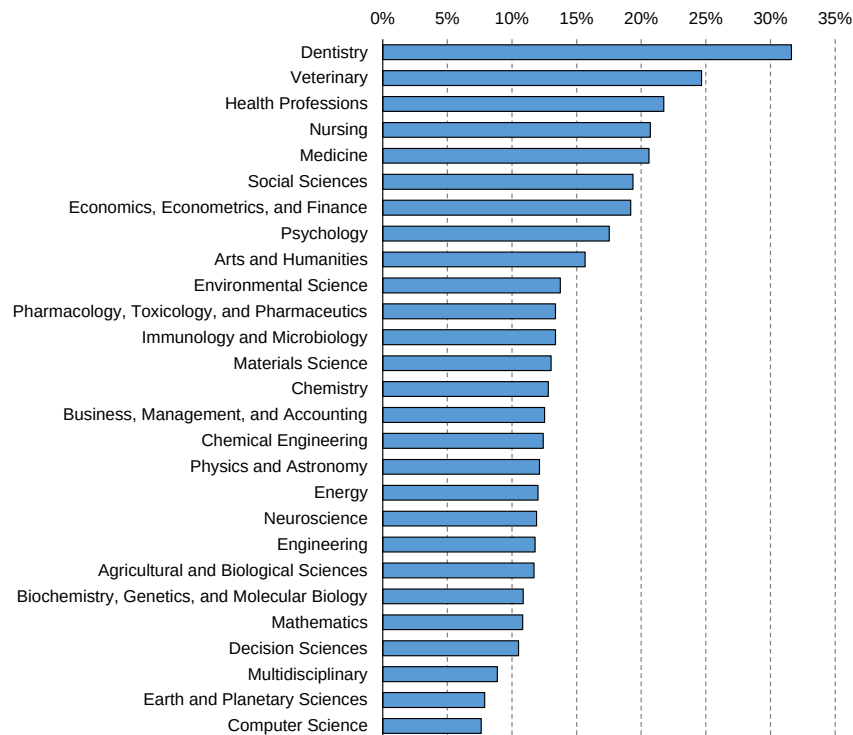
Fig. 4 shows the share of the top 10% high-novelty papers in China and the United States, which contribute to the top 10% of high-novelty papers globally by field. The fields in which both countries score high and low are different.

In China, the fields mainly related to health sciences (i.e., dentistry, veterinary science, health profession, nursing, and medicine) had high ratios of over 20% of their papers among the top 10% high-novelty ones. The fields with low scores (<10%) were in multidiscipline, earth and planetary science, and computer science.

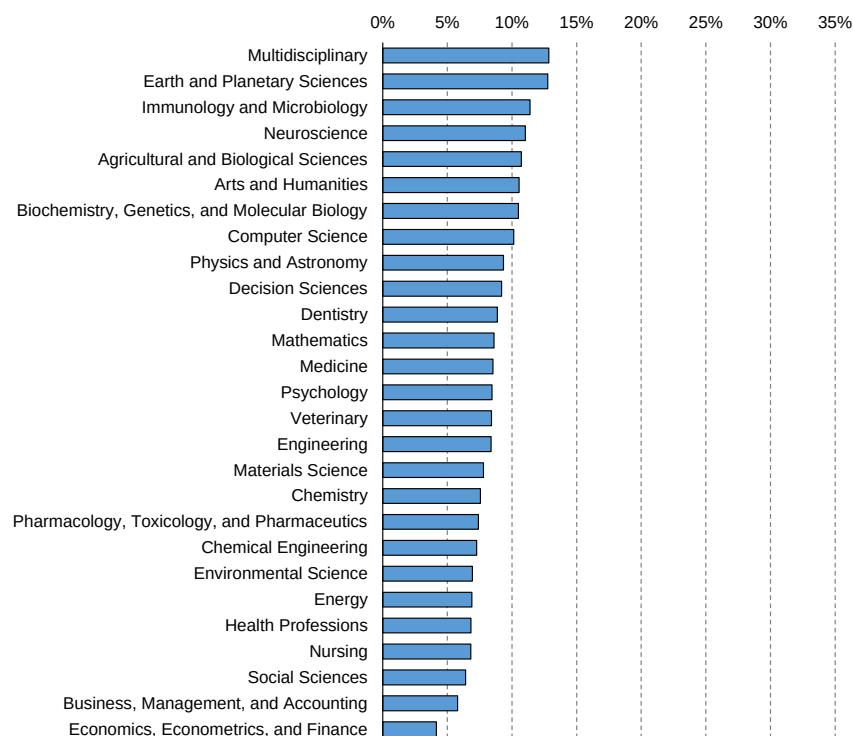
In the United States, the fields that scored low in China (i.e., multidiscipline, earth, and planetary science) had high ratios (>12%) of papers among the top 10% high-novelty papers, followed by immunology and microbiology, neuroscience, and agricultural and biological sciences. In contrast, the fields related mainly to humanities and social sciences had low scores in their papers among the top 10% high-novelty papers.

Figure 4: Share of top 10% high-novelty papers in each country/region by field in 2021

a) China



b) United States of America



Source: Author calculations based on Scopus Custom Data extracted in December 2022.

Note 1) Papers refer to articles, conference papers in journals, and conference papers in conference proceedings.

2) Focused on papers published in 2021.

3) Share of papers in the top 10% of high-novelty papers worldwide by the total number of papers in the country/region. The number of papers published in each country/region was calculated by fractional counting.

#### 4. Discussion and Conclusion

Novelty, focused on new combinations of knowledge sources, can be an indicator for measuring potential breakthroughs, as supported by previous innovation studies. This study analysed academic papers published in 2021 using bibliometric approaches to determine the geographical distribution of high-novelty research. A comparison of the top 20 countries/regions in the number of papers with those in the top 10% high-novelty papers show no differences in ranking in many of the top 20 countries/regions (except for Japan, Russia, and Iran). Within the top 20 countries/regions in the number of papers, Taiwan, Poland, and China had the highest percentage of top 10% high-novelty papers worldwide. Russia, Indonesia, and Japan had the lowest numbers of countries/regions. China, ranking first among the countries with the top 10% high-novelty papers, had a high share of the top 10% high-novelty papers compared with the total number of papers in the country. By contrast, the United States, the second-largest contributor to the top 10% of high-novelty papers, does not have a high share. Furthermore, compared to the share of the top 10% of high-novelty papers by field, China and the United States, leading global contributors to the top 10% high-novelty papers, the fields scored high and low in different fields. Our indicator measures the degree of unusual knowledge recombination, which is a novel research topic. Given the structure of novelty indicators, several factors influence the degree of unusual knowledge recombination. For example, unusual knowledge recombination can be attributed to the field dissimilarity of references, the oldness/newness of references, and regional differences in references.

This study is a pilot to present worldwide statistical data on research novelty and has room for improvement. Finally, I present three directions for this study. The first is an analysis in the scientific field. In this analysis, field data are shown for China and the United States only, which are the main contributors to the top 10% of high-novelty papers. A future analysis could clarify the results shown in the top 20 countries/regions in terms of the number of papers. The second is to expand the scope of the years analysed. This analysis focuses on 2021, the latest year in the data. How trends in high-novelty research have shifted, especially when comparing the pre-and post-COVID19 period should be one of the most interesting topics. The final direction is a robustness check. This analysis adopted a combinatorial novelty indicator measured through units comprising paired reference papers. Combinatorial novelty indicators with pairs of other knowledge source units such as journals and keywords have already been proposed. Therefore, whether results similar to this analysis can be obtained using other novelty indicators needs to be verified.

## **Disclaimer**

This paper should not be interpreted as representing the official view of my affiliations at the time of writing, that is the OECD and NISTEP. The contents and opinions expressed are those of the author.

## **Open science practices**

Restrictions apply to the availability of raw bibliometric data used under Elsevier's licence.

## **Acknowledgments**

I would like to thank Masatsura Igami for supporting the extraction of Scopus Custom data for the analysis and for providing useful comments.

## **Funding information**

This study was supported by JSPS KAKENHI 22H00871. The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

## **Competing interests**

The authors declare no conflicts of interest relevant to the content of this paper.

## **References**

- Baird, L. M., & Oppenheim, C. (1994). Do citations matter? *Journal of Information Science*, 20(1), 2-15.
- MacRoberts, M., & MacRoberts, B. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.
- Matsumoto, K., Shibayama, S., Kang, B., & Igami, M. (2021). Introducing a novelty indicator for scientific research: validating the knowledge-based combinatorial approach. *Scientometrics*, 126(8), 6891-6915.
- Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, 613(7942), 138-144.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468-472.
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416-1436.