



Productivity trends and citation impact of different institutional collaboration patterns at the research units' level

Lipeng Fan¹ · Yuefen Wang^{1,2}  · Shengchun Ding¹ · Binbin Qi³

Received: 14 December 2019 / Published online: 13 July 2020
© Akadémiai Kiadó, Budapest, Hungary 2020

Abstract

In order to gain a deeper understanding of how research performance and collaboration patterns of institutions affect productivity trends and citations, this paper classifies institutions into two types: main and normal institutions, and then divides the dataset into six types: M and N as intra-institution collaboration types, and M&M, M&N, N&M, N&N as inter-institution types (M: main institutions, N: normal institutions). After analysing the productivity trends and citation impact at the research units' level, the main results are shown as following: through a large-scale and long-span data, M papers account for the highest percentage, and play an important leading role in the beginning, and the average citation value of M&M papers is significantly higher than other types; although the number of papers with multi-authors is increasing over time, the impact of the number of authors on citations may vary from discipline to discipline, and there is a slightly negative relationship between them in artificial intelligence field in our data; despite the number of institutions and countries has a positive impact on citations in whole dataset, it differs when considering different institutional collaboration patterns and the first author's country; no matter what institutional collaboration pattern is, the papers with USA as first author's country always have a significant greater impact than China as first author's country. After analysing two negative binomial regression models, some results support the above conclusions. Moreover, we find that the number of M institutions has a significant greatest impact on citations, while M institution as first author's affiliation only has a slightly influence; China as first author's country has a negative impact, while USA as first author's country has a moderately positive impact, and slightly lower than that of the number of countries, moderately higher than that of the number of institutions.

Keywords Institutional collaboration pattern · Productivity trends · Citation impact · Negative binomial · Artificial intelligence

✉ Yuefen Wang
yuefen163@163.com

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

² Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing, China

³ School of Information Management, Nanjing University, Nanjing, China

Introduction

Scientific research collaboration refers to the form between individuals and individuals, individuals and groups, groups and groups work together to accomplish the same scientific research task. Scientific research is a complex and arduous group work. The interaction between people in scientific research activities directly affects the completion of scientific research collaboration and programs. Through research collaboration, however, new academic ideas can be generated, knowledge and culture can be promoted to exchange, research productivity can be increased, and research costs can be decreased (Katz and Martin 1997; Beaver 2001; Yuan et al. 2018). Some studies have revealed that collaboration papers are becoming more and more prevalent in scientific research (Abramo et al. 2004; Nguyen et al. 2017).

The existing studies about collaboration papers are mainly from two aspects: collaboration patterns and the relationship between collaboration patterns and citations.

There are several collaboration patterns have been proposed by researchers. According to the attributes of research units, Ni and An (2018) classify the international collaboration papers into seven types by the economic level of countries; Liu et al. (2012) following the classification concept of patentees explore the institutional collaboration papers among six levels: private enterprises, government, university, hospital, research institutes and not-for-profit organizations; Lee et al. (2012) propose four institutional collaboration types by categorizing network analyses into two dimensions: structural positions and the relational characteristics of individual nodes; Wang et al. (2017) investigate collaboration patterns from scholar's local perspectives based on their academic ages. Moreover, according to the number of research units in a publication, several studies classifies the collaboration patterns into domestic, international, intra-institutional and inter-institutional forms (Sooryamoorthy 2009; Han et al. 2014). For the relationship between collaboration patterns and citations, the research of Gazni and Didegah (2011) has shown that there was a positive correlation between the number of citations and the number of research units; Ibáñez et al. (2013) have defined that international collaboration results on average in publications with higher citation rates than national and institutional collaborations; Ni and An (2018) show that the subset of international collaboration with the same economic level has the higher value.

In this paper, we aim to explore collaboration patterns based on the performance of institutions and how the patterns affect productivity trends and citations. As some studies indicate that the relationship between collaboration patterns and citation differ from discipline to discipline (Franceschet 2011; Gazni et al. 2012), we take the field of Artificial Intelligence (AI) as an example, for AI, a branch of Computer Science, draws a great attention of scientists and becomes one of the most popular terms on internet in recent years. In particular, we will address the following research questions in this analysis: (1) How do the productivity trends and citation impact vary among different institutional collaboration patterns? (2) How does the research units affect the productivity trends and citation impact among different institutional collaboration patterns? (3) How does the impact vary among different factors on citations according to different institutional collaboration patterns?

Data and methodology

Data sources and pre-processing

To make the data source more representative, we used the category named “Computer Science, Artificial Intelligence” in Web of science core database to search publications, and limited the retrieved data to articles published from 1997 to 2017 (collection date was May 10, 2018). Eventually, the dataset covers 765,491 academic articles and 72,916 institutions. In order to analyse collaboration patterns and citations, we extracted the author’s name (AU), authors’ address information (C1), publish year (PY), cited frequency (Z9) and WOS number (UT) from the dataset, and excluded the papers without C1 tags and length of AU tags in which is smaller than 2. Finally, a total of 669,569 were obtained for analysis in this paper.

In consideration of the ambiguity of countries and institutions in the analysis process, we used C1 tag to help cleaning the country and institution data.

Due to some institutions with same names are in different countries, we clean the country data firstly. For example, UK has four or more regions, such as England, Scotland, Wales and Northern Ireland. In common sense, England represents UK, so we transform the other three regions into England. Although Hong Kong, Macau and Taiwan are regions of Peoples R China, most institutions belonging to Hong Kong and Macau have written the country name as Peoples R China. So, we take Taiwan as an independent region in this paper, and replace the other two regions with Peoples R China.

Compared with country information extraction, identifying the unique name of institutions is a complicated task (van Raan 2005). After analysing the institution information, we clean the institutions with the following errors: (1) *Name has been changed*, such as “Beihang Univ” and “Beijing Univ Aeronaut and Astronaut”. (2) *Different abbreviations*, such as “Swiss Fed Inst Technol” and “ETH”. (3) *Different positions of “Univ”*, such as “Univ Washington” and “Washington Univ”. (4) *Institutional departments*, such as both “INRIA Rhone Alpes” and “INRIA Sophia Antipolis” all belong to “INRIA”. (5) *Different translation*, such as “Free Univ Brussels” and “Univ Libre Bruxelles” actually are the same institution. Finally, we merge the institutions with the above problems, and set the right name with the higher number of publications.

Definition of institutional collaboration patterns

Before defining institutional collaboration patterns, we need to rank the institutions in our dataset.

H-index, proposed by Hirsch, is a method to assess the influence of scientists by publications and citations simultaneously, and can also be extended to evaluate the performance of institutions, countries and journals. When calculating publications and citations of institutions, it is difficult to distribute the credits among ordered co-institutions. Since the order of institutions in a paper represents a certain order of importance, we chose Arithmetic counting, one of the most prominent counting methods, to calculate the h-index of institution, and named this method as AH-index. In this counting method, the credits are linearly distributed in decreasing order among several co-institutions according to the following formula (Trenchard 1992; Van Hooydonk 1997):

$$c_i^n = \frac{n-i+1}{\sum_{k=1}^n k} = \frac{2(n+1-i)}{n(n+1)}, \quad (i \in n) \quad (1)$$

In formula (1), i is the position of the institution, n is the number of institutions in a paper. We use the same formula to obtain the citations of each institutions. If there are some same institutions in a paper, the publications and citations of this institution will be calculated multiply times.

Although corresponding author and first author are equally important in many cases, they are the same in almost 96% papers of our dataset. So, we only take the order of institutions into consideration to reduce the workload in our study.

After sorting in descending order according to the AH-index of institutions, we classify institutional collaboration patterns as follows:

- *Main institutions and normal institutions* Before classifying institutional collaboration patterns in our data based on the performance of institutions, we firstly divide all institutions into two types according to the institutional rank. In order to avoid the impact of institutional type' productivity on institutional collaboration patterns, the number of publications of each institutional type should be consistent. Therefore, we select the institutions that the sum publications of which are not less than half of our total dataset as *Main institutions*, while the remain as *Normal institutions*. In this paper, the number of *Main institutions* is 561 with 50.64% of total papers. Table 1 lists the top 20 of *Main institutions*.

Table 1 Top 20 of Main institutions

Rank	Institution	Country	Publications	Citations	AH-index
1	Nanyang Technol Univ	Singapore	4112.32	56,775.6	82.69
2	Carnegie Mellon Univ	USA	3451.47	58,065.23	79.03
3	MIT	USA	2235.01	41,417.67	73.66
4	Chinese Acad Sci	P. R. China	6880.77	47,583.21	68.45
5	Stanford Univ	USA	1601.14	31,656.08	63.41
6	Univ Calif Berkeley	USA	1466.07	64,332.18	63.37
7	Hong Kong Polytech Univ	P. R. China	1814.32	29,688.51	61.61
8	Univ So Calif	USA	2085.81	27,150.27	61.55
9	Univ Illinois	USA	2347.08	42,383.32	61.06
10	Swiss Fed Inst Technol	Switzerland	1541.89	31,233.73	60.7
11	Natl Univ Singapore	Singapore	2571.57	33,375.42	59.09
12	Univ Oxford	England	967.01	26,182.3	55.1
13	Tsinghua Univ	P. R. China	5262.58	30,471.49	54.43
14	Ecole Polytech Fed Lausanne	Switzerland	1191.3	21,716.58	53.92
15	Chinese Univ Hong Kong	P. R. China	1967.11	23,341.94	52.14
16	Univ Maryland	USA	1695.92	24,391.16	51.8
17	Univ Granada	Spain	1524.34	21,332.6	51.22
18	Southeast Univ	P. R. China	1826.21	18,886.26	50.12
19	Washington Univ	USA	1427.22	17,108.22	49.51
20	INRIA	France	1003.76	19,706.27	48.07

- *M, M&M, M&N, N, N&M, N&N* In order to analysis how the number of main institutions and main institution as first author’s affiliation affect the productivity trends and citation impact, we divide the collaboration papers into six parts, as shown in Table 2. For example, M represents an intra-institution collaboration type, papers of which published by only one main institution; M&M represents an inter-institution collaboration type, papers of which published by one main institution as first author’s affiliation and at least one main institution in the remaining institutions; M&N also represents an inter-institution collaboration type, papers of which published by one main institution as first author’s affiliation and one or more normal institutions in the remaining.
- *Dataset of a specific country* When considering how the productivity trends and citation impact of different institutional collaboration patterns vary among different countries, papers of a country need to be obtained. Since the first author is the main contributor of a paper, we select the papers with the country where the first author’s affiliation belongs to as the dataset of that country.

Statistical tests

We run several statistical tests to compare the citation impact of different collaboration types statistically. Since the citation data is not normal distribution, we use two non-parametric tests in this research: Kruskal–Wallis test and Mann–Whitney test. The Kruskal–Wallis test is to analyse whether there is a significant difference of the distributions among three or more samples, but it cannot identify the differences between two independent samples. So, we use the Mann–Whitney test to rank the value between two samples. In all cases, the significant level of these tests is 0.05.

Results and analysis

Productivity trends and citation impact of different institutional collaboration patterns

This part is to describe what the main productivity trend is and how the citation impact varies among institutional collaboration types.

On the whole, both Figs. 1 and 2 show that intra-institution papers are the main trends, and the percentage of M papers (35.6%) is higher than that of N papers (29.5%).

Table 2 Institutional collaboration patterns

	Institution types		
	Collaboration types		
	Intra-institution	Inter-institution	
Main institutions		Normal institutions	
Main institutions	M	M&M	M&N
Normal institutions	N	N&M	N&N

Fig. 1 Percentage of collaboration papers according to institutional collaboration patterns

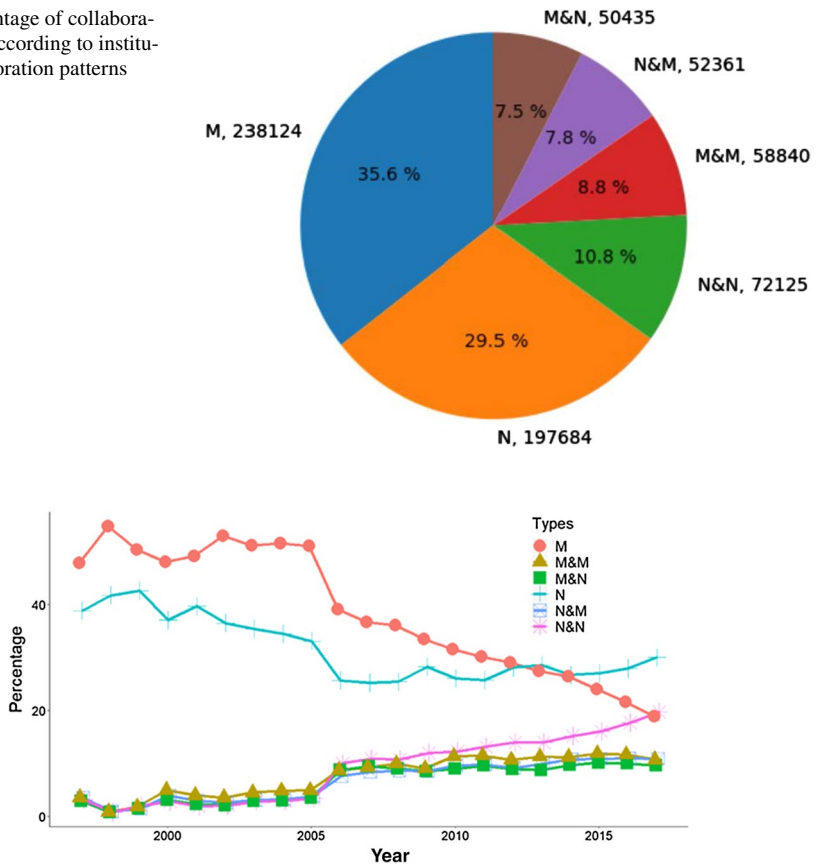


Fig. 2 Productivity trends of different institutional collaboration patterns

However, the trends of intra-institution types both keep descending, while that of inter-institution papers keep increasing over time in Fig. 2. Figure 2 also shows that in intra-institution types, M type is always the main type before 2012, and after that, the percentage of N type surpass M papers to become the primary pattern until now, which indicates that main institutions play a leading role at the beginning, and after a certain time, normal institutions follow them and publish a lot of papers. In inter-institution types, the percentages of N&N (10.8%) and M&M (8.8%) are higher than that of the N&M (7.8%) and M&N (7.5%) type in Fig. 1. Besides, in Fig. 2, M&M type is the main pattern before 2005, and after that, N&N type rise quickly and surpass the other types to become the main trend until now. The results indicate that collaboration is more often among institutions of similar status in the field of AI, and N and N&N papers will be the main trend in the recent years.

Table 3 shows the average citations of different institutional collaboration types. From Table 3, we can see that M&M type has the highest average citations (11.65 ± 23.83), followed by N&M type (7.05 ± 17.29), M&N type (6.73 ± 17.05), M type (5.34 ± 15.12) and N&N type (3.38 ± 10.39), while the lowest average value corresponds to N type (2.37 ± 8.31).

Table 3 Mean \pm standard deviation of citations for different collaboration types

Types	Citation count
M	5.34 \pm 15.12
M&M	11.65 \pm 23.83
M&N	6.73 \pm 17.05
N	2.37 \pm 8.31
N&M	7.05 \pm 17.29
N&N	3.38 \pm 10.39

The value in bold represents the highest mean citation in different collaboration types

The results of the Kruskal–Wallis test show that there are significant differences among the six collaboration types based on the average value of citations. Then, we run Mann–Whitney tests to rank the six subsets. The results show that, in general, the average value of inter-institution papers is statistically higher than that of intra-institution collaboration papers except N&N type, the value of which is statistically lower than that of M type, which indicate that the number of institutions has a positive relationship with citations, and the number of main institutions has a greater impact on citations than that. To exclude the influence on the number of main institutions, we calculate the average citation value of M&N and N&M with the same number of main institutions, and find that the value of M&N (6.73 \pm 17.05) is statistically a slightly higher than that of N&M* (6.51 \pm 16.43), where N&M* represents the N&M papers published with only one M type institution. This phenomenon also exists in M&M and N&M papers when considering the citations at the same number of main institutions, which indicate that main institution as first author’s affiliation has a slightly positive relationship with citations. Moreover, the value of N&M is statistically higher than that of M&N type from Table 3, and the average number of main institutions in N&M type is more than that in M&N type. We think that, the number of main institutions has a greater impact on citations than main institution as first author’s affiliation.

Based on the above results, we find that M type was guiding AI field at the beginning, while the N type follows them and tends to be the main productivity pattern in the recent years. Besides, both main institution as the first author’s affiliation and the number of main institutions have a statistically positive relationship with citations, and the latter is higher than the former.

Productivity trends and citation impact of different institutional collaboration patterns at the author level

In this part, we wonder whether the number of authors has an impact on productivity trends and citations in different institutional collaboration patterns.

The average citation value and standard deviation of different number of authors are: two-author papers (5.05 \pm 15.01), three-author papers (5.05 \pm 14.67), four-author papers (5.0 \pm 14.45), and five or more author papers (5.05 \pm 14.1). We can see that the number of authors has little impact on citations in this analysis.

Figure 3 shows the productivity trends of different institutional collaboration patterns at the number of authors level. In intra-institution collaboration papers, two-author papers are both the main trends in M type and N type, but there is a subtle difference between

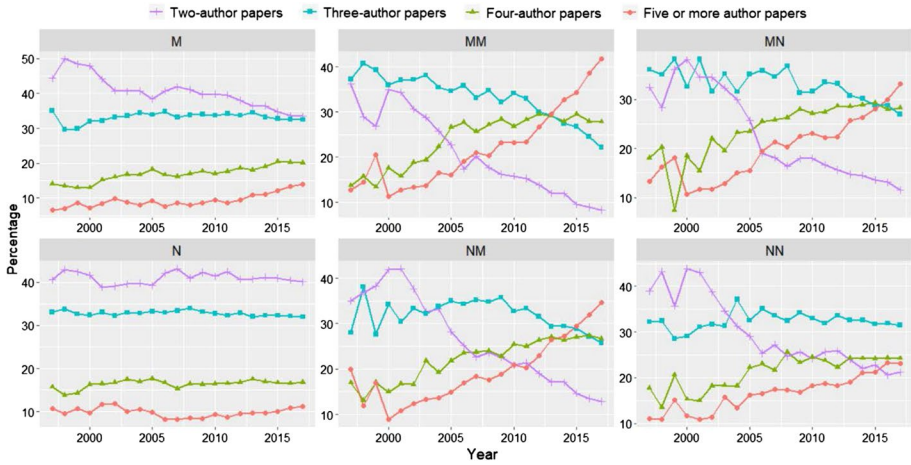


Fig. 3 Productivity trends of different institutional collaboration patterns at the author level

them. In M type, the trend of two-author papers inclines to downtrend, and other lines keep on the rise, whereas, all the lines in N type keep stable over time, which indicates that the main trend of M type may change to three or more author papers in the future, while N type does not. In inter-institution papers, three-author papers are the main trend at the beginning, whereas, in the recent few years, five or more author papers become the main trend and keep rising dramatically to become the main trend in different types except N&N type. In N&N type, although the percentage of four or more author papers presents an ascending trend, the percentage of three-author papers keeps stable and higher than that of four or more author papers over time. Therefore, the main trend of N&N type will still be the three or more author papers in the future. The results show that in inter-institution papers, three or more authors will be the main trends, while it will still be two or three authors in intra-institution papers.

Table 4 shows the citation impact of different institutional collaboration types according to the number of authors. We find that no matter how many authors in collaboration papers, the M&M institutional collaboration type always has the highest average citation. Moreover, with increasing number of authors, the citation value has a slightly downtrend in

Table 4 Mean ± standard deviation of citations for different collaboration types according to the number of authors

Types	Number of authors			
	2	3	4	≥ 5
M	5.92 ± 16.39	5.22 ± 14.81	4.59 ± 13.44	4.67 ± 13.25
M&M	13.58 ± 26.83	12.47 ± 24.99	11.59 ± 23.63	9.67 ± 20.45
M&N	8.56 ± 20.39	7 ± 17.38	6.39 ± 16.56	5.45 ± 14.12
N	2.56 ± 8.89	2.3 ± 7.91	2.17 ± 7.99	2.16 ± 7.61
N&M	8.19 ± 19.53	7.45 ± 17.99	6.41 ± 15.64	6.2 ± 15.77
N&N	3.75 ± 11.74	3.38 ± 10.1	3.19 ± 9.46	3.17 ± 10.13

The value in bold represents the highest mean citation in the same number of authors

all six types. The results of Kruskal–Wallis test show that there are significant differences among different number of authors in different types except N&N type. The results of Mann–Whitney show that the citation of two-author papers is statistically higher than others. Based on the findings, we think that, the number of authors has a statistically slightly negative relationship with citations among different institutional collaboration patterns.

According to Fig. 3 and Table 4, we can find that although the number of authors in a paper keeps rising over time, it is not the more the better when considering the effects on citation counts. Actually, there are different results in different disciplines about the relationship between the number of authors and citation counts according to previous studies. For example, the results in Peng & Zhu (2012) show that the number of authors is one of the significant predictors of citations in social science articles about Internet studies, while So et al. (2014) found that the number of co-authors had a significant negative effect on the number of citations in science and technology field. Moreover, after analysing 6 subject fields to examine the factors affecting the number of citations of articles, Onodera and Yoshikane (2014) found that the number of authors was shown to be a significant predictor in only two of the six fields. Hence, the results in our study a slightly negative relationship between the two variables maybe due to the discipline of AI field our paper selected, in which we guess the research of basic algorithms may involves fewer researchers and receives more citations than that of specific applications with AI techniques, and this reason needs to be confirmed further.

Productivity trends and citation impact of different institutional collaboration patterns at the institution level

Since the productivity trends of intra-institutional collaboration papers have been shown in Fig. 2, here, we only explore the productivity trends of inter-institution papers, and also take whether it is an international collaboration paper into consideration.

Figure 4 shows the productivity trends of domestic and international papers according to the number of institutions. We find that the trends in inter-institution papers almost keep stable during the past over twenty years in AI field, and two-institution papers are the main pattern not only in domestic papers but also in international papers. Obviously, M&M and N&M papers are more inclined to be international papers than the other datasets, while M&N and N&N papers prefer to be domestic papers, which indicate that collaboration among main institutions tends to be international, while it tends to be domestic among normal institutions. Moreover, the percentage of three or more institution papers are almost the same trends in M&M and N&M domestic and international papers, and much more than that in M&N and N&N papers, which indicate that the main institution tends to collaborate with more number of institutions than the normal institution.

Table 5 shows the average citation value of domestic and international papers with different number of institutions. We can see that papers published by four or more institutions in both domestic and international papers get the highest citation value than the other subsets. For example, papers published by domestic four or more institutions have the highest value (6.76 ± 17.17), followed by the ones published by domestic three institutions (6.53 ± 17.17), the ones published by domestic two institutions have the lowest value (5.76 ± 15.91). Moreover, the lowest value in international papers is higher than the highest value in domestic papers, that is, papers published by international two institutions (8.1 ± 19.31) have the higher average citation value than the ones published by domestic four or more institutions (6.76 ± 17.17). The results of Kruskal–Wallis test show that there

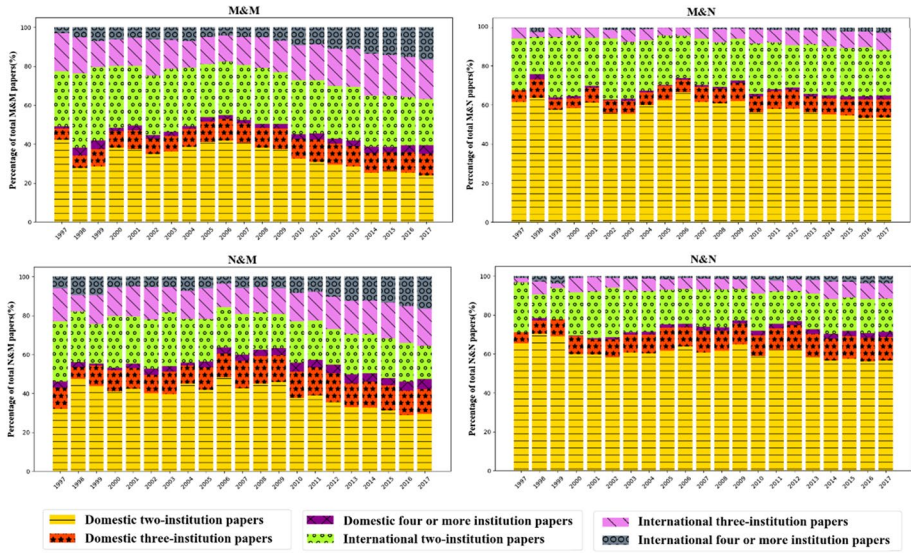


Fig. 4 Productivity trends of inter-institution collaboration types at the institution level

Table 5 Mean and standard deviation of citations according to the number of institutions

Types	Number of institutions	Citation count
Domestic papers	2	5.76 ± 15.91
	3	6.53 ± 17.17
	≥ 4	6.76 ± 17.17
International papers	2	8.1 ± 19.31
	3	9.02 ± 19.91
	≥ 4	9.91 ± 21.26

The value in bold represents the highest mean citation at the same type of papers

are significant differences among the six subsets. Then we run the Mann–Whitney test to rank any two subsets in the six subsets, the results show that the citation rank is statistically consistent with the findings in Table 5, which means that there is a statistically positive relationship between the number of institutions and citations, the citation value of international papers is statistically higher than that in domestic papers, and the number of countries has a greater impact on citations than the number of institutions.

Table 6 shows the average citation value of different types according to the number of institutions. The results show that, in domestic papers, papers published by three institutions always have the highest value except N&M papers, which is four or more institutions. While in international papers, it is four or more institutions only in M&M and N&M papers. After running the Kruskal–Wallis test at the same institutional collaboration patterns, we find that there are significant differences in different types except M&M domestic papers and M&N papers, which means the citation value rank in M&M domestic and M&N papers according to different number of institutions is not statistically valid. The

Table 6 Mean and standard deviation of citations of inter-institution collaboration patterns according to the number of institutions

Types	Number of institutions	Institutional collaboration patterns			
		M&M	M&N	N&M	N&N
Domestic papers	2	10.93 ± 23.19	6.13 ± 16.05	6.37 ± 16.83	2.98 ± 9.62
	3	11.14 ± 23.21	6.26 ± 16.77	6.67 ± 16.95	3.14 ± 9.96
	≥ 4	10.66 ± 22.47	6.07 ± 15.55	7.01 ± 17.31	2.9 ± 8.52
International papers	2	11.72 ± 24.07	8.14 ± 19.23	7.19 ± 17.48	4.35 ± 12.02
	3	12.48 ± 24.26	7.39 ± 17.28	7.95 ± 17.51	4.65 ± 12.62
	≥ 4	13.17 ± 25.31	7.66 ± 18.12	8.47 ± 18.5	4.35 ± 11.38

The value in bold represents the highest mean citation at the same type and institutional collaboration pattern

results of Mann–Whitney test show that the citation value of three-institution papers is statistically higher than two-institution papers in both N&M domestic papers and N&N papers, while it is four or more institutions statistically corresponding to the highest citation value in both M&M and N&M international papers, which indicate that there is only a positive relationship between the number of institutions and citations in M&M and N&M international papers.

From Tables 5 and 6, we can find that no matter in the whole dataset or any institutional collaboration pattern data, the citation of international papers is always higher than that of domestic papers, which is in line with the results in most previous studies (Sooryamoorthy 2009; Barbara et al. 2012; Puuska et al. 2014). However, despite the conclusion that there is a positive relationship between the number of institutions and citations in the whole dataset, it differs when considering in different institutional patterns (see Table 6), which can be confirmed by the results in Table 3, where the citation value in M papers (5.34 ± 15.12) is higher than that in N&N papers (3.38 ± 10.39). The reason for the results in Table 6 may be that in M&N and N&N papers, arising the number of institutions means increasing the number of N type institutions, which is not the more the better for citations, while in M&M and N&M papers, the arising citation value may due to the average number of the M type institutions, which is increasing with the number of institutions (the citation rank of M&M domestic papers in our data is not statistically valid according to the Kruskal–Wallis test). Note that, the number of M institutions has a significant positive impact on citations, which has been proved in Table 3.

Productivity trends and citation impact of different institutional collaboration patterns at the country level

In this part, we firstly explore the productivity trends and citation impact of different institutional collaboration patterns according to the number of countries, and then take P. R. China (papers:153,946; citations: 668,360) and USA (papers: 92,109; citations: 1,1200,336) as two examples to show how the productivity trends and citation impact of institutional collaboration patterns vary in different countries, as the two countries have the highest quantity or quality papers in our data.

According to the number of countries, we firstly calculate the average citation value and standard deviation in different number of countries, that is, single-country

papers (4.47 ± 13.58), two-country papers (8.41 ± 19.59), three or more country papers (9.93 ± 20.83). After running Kruskal–Wallis test, the results show that there are significant differences in the three types. Moreover, the results of Mann–Whitney test show that the value in three or more country papers is statistically higher than that in two-country and single-country papers, which indicate that there is a positive relationship between the number of counties and the citation value. Then, from the institutional collaboration patterns, we find that the productivity trends of inter-institution collaboration patterns keep stable over time, that is, single-country papers are the main trend in M&N, N&M, and N&N papers, while two-country papers is the main trend in M&M papers in recent several years. After calculating the citation impact of inter-institution patterns in different number of countries, we find the result is consistent with the result in Table 6, which is that the positive relationship between the number of counties and citations only exists in M&M and N&M papers.

According to the datasets of P. R. China and USA, we explore how the productivity trends and citations of different institutional collaboration patterns vary on specific countries.

Figure 5 compare the productivity trends of different institutional collaboration types between P. R. China and USA dataset. From Fig. 5, we can see that the trends of different types in both P. R. China and USA dataset are almost the same, that is, the percentage of intra-institution papers has a downtrend, and that of inter-institution papers has an uptrend. In intra-institution papers, M papers in USA dataset has a slower downtrend than that in P. R. China, and N papers has a faster downtrend than that in P. R. China. In inter-institution papers, the percentage of M&M papers in USA dataset is higher than that in P. R. China over time, while the percentage of N&N papers in USA dataset is lower than that in P. R. China. Meanwhile, the percentage of M&N and N&M in USA and P. R. China keep shoulder and shoulder over time. The results indicate that although the productivity of P. R. China dataset (153,946) is significantly higher than that of USA (92,109), the productivity trend of M and M&M papers in P. R. China dataset is obviously lower than that in USA.

Table 7 shows that whether in P. R. China or USA dataset, M&M papers always have the highest citation value, while N and N&N papers are corresponding to the lowest value.

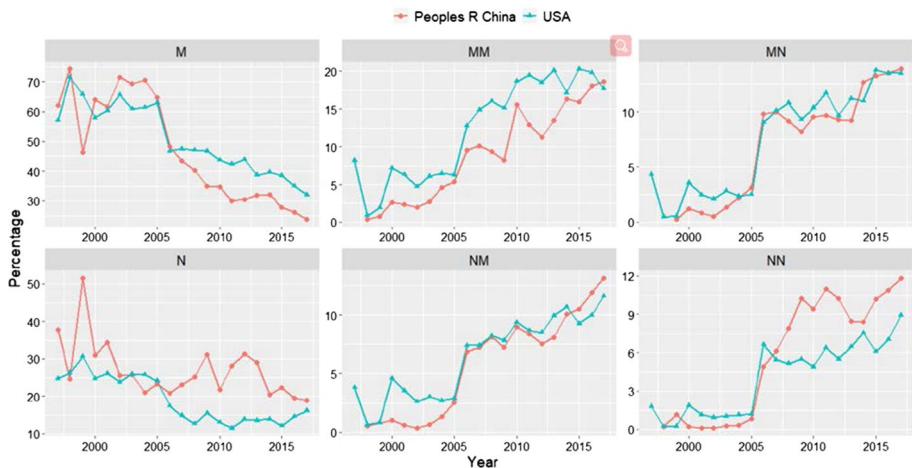


Fig. 5 Productivity trends of institutional collaboration types in P. R. China and USA

Table 7 Mean and standard deviation of citations of institutional collaboration types in P. R. China and USA

Types	Country of first author	
	Peoples R China	USA
M	3.24 ± 11.6	7.4 ± 18.75
MM	11.62 ± 24.45	14.22 ± 26.94
MN	4.94 ± 14.58	9.22 ± 22.36
N	0.99 ± 4.75	3.48 ± 11.74
NM	4.92 ± 13.83	11.77 ± 24.87
NN	1.84 ± 7.08	5.27 ± 15.47

The value in bold represents the highest mean citation at the same country of first author

The results of Kruskal–Wallis test show that there are significant differences among different types both in China and USA dataset. The results of Mann–Whitney test show that the value of M&M papers is statistically higher than the other subsets both in China and USA. Additionally, the value of all types in China dataset is statistically lower than the corresponding value in USA.

Table 8 shows that in the dataset of China, the more countries in a paper, the higher the citation value, while there is no such relationship in the dataset of USA. After running the Kruskal–Wallis test, the results show that there are significant differences among the different types only in the dataset of China, which means the citation value rank in the dataset of USA is not statistically valid. After running the Mann–Whitney test, the results show that there is actually a statistically positive relationship between the citation value and the number of countries in the dataset of China.

Based on the above results, we can see that no matter what institutional collaboration pattern is, the papers with USA as first author’s country always have a significant greater impact on citations than China as first author’s country. Moreover, although there is a positive relationship between the number of countries and citations in the whole dataset, it differs when considering the country that first author’s affiliation belongs to in different institutional collaboration patterns. In our data, when China as first author’s country, it will be better to collaborate with more other countries, especially in N&M and N&N papers.

Table 8 Mean and standard deviation of citations of inter-institution collaboration types in P. R. China and USA according to the number of countries

Country of first author	Number of countries	Inter-institution collaboration patterns			
		M&M	M&N	N&M	N&N
Peoples R China	1	10.58 ± 23.54	3.9 ± 11.97	3.9 ± 11.95	1.39 ± 5.83
	2	12.01 ± 24.46	8.82 ± 21.34	7.04 ± 17.24	4.51 ± 11.47
	≥ 3	17.12 ± 29.68	9.87 ± 20.9	11.04 ± 19.27	10.69 ± 19.18
USA	1	14.34 ± 27.09	9.19 ± 22.59	11.89 ± 25.31	5.14 ± 15.66
	2	14.26 ± 27.02	9.4 ± 22.22	11.28 ± 23.42	5.59 ± 15.05
	≥ 3	12.87 ± 24.85	6.77 ± 13.65	12.71 ± 26.33	5.76 ± 13.97

The value in bold represents the highest mean citation at the same country of first author and collaboration pattern

While when the country is USA as first author's country, there is little influence on citations with the increasing of the number of countries.

Influence intensity of different factors on citations

From the above results, we can see that the citation could be influenced by the number of authors (N_auths), institutions (N_instis), countries ($N_countries$), main institutions ($N_mainInstis$), the country of first author is P. R. China ($China_first$) or USA (USA_first) and the institution of first author is main institution ($mainInsti_first$) or not. We take all these factors in a regression model to verify the results above and compare the influence intensity on citations of these factors, as shown in Table 9.

Since the number of citations to an article is a count data, a discrete count data model should be used such as poisson or negative binomial (NB) regression (Didegah and Thelwall 2013; Sud and Thelwall 2016). One important feature in poisson regression is the equal mean and variance, which does not satisfy in our data, so we choose NB model. In the NB model, the citation counts are divided by average citation count in the same year to control for variation across time (Reingewertz and Lutmar 2018), and taken natural logarithms accounts for stable data without changing the correlativity, as well as other continuous variables. The value of dummy variables is 1 if yes 0 otherwise.

Table 9 presents the results of the NB model. Data in Model 1 is intra-institution papers and in Model 2 is inter-institution papers.

In Model 1, the coefficients for N_auths and $China_first$ are both negative and significant, while USA_first and $mainInsti_first$ are both positive and significant. The results show that in single institution papers, the number of authors and China as the first author's country both have a negative effect on research performance, while both M type and USA as the first author's country has a positive effect, and the former is much higher than the latter. In Model 2, the coefficients of $N_mainInsti$ is the greatest one, followed by $N_countries$,

Table 9 Results of negative binomial regression

Variables	Research performance	
	Model 1	Model 2
Constant	-1.492*** (0.01)	-1.201*** (0.012)
Continuous variables		
log(number of authors)	-0.046*** (0.008)	-0.147*** (0.008)
log(number of institutions)	-	0.105*** (0.013)
log(number of countries)	-	0.254*** (0.008)
log(number of main institutions)	-	0.627*** (0.01)
Dummy variables		
P. R. China as the country of first author	-0.511*** (0.008)	-0.021*** (0.007)
USA as the country of first author	0.147*** (0.007)	0.21*** (0.008)
main institution as first author's affiliation	0.576*** (0.006)	0.068*** (0.008)
Observations	435,808	233,761
Wald \times 2	16,421.25***	17,100.31***
Log pseudolikelihood	-276,480.16	-203,503.23

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

USA_first, N_instis, and mainInsti_first corresponding to the lowest one, while the coefficients of N_auths and China_first are still negative. The results show that the number of M type institutions has a significant higher impact than the other factors, and the number of counties is slightly higher than that of USA as the first author's country and moderately higher than that of number of institutions, while mainInsti_first only has a slightly impact on citations.

Comparing Model 1 and Model 2, we can see that N_auths always has a negative relationship with the research performance, which confirms the results in Table 4. Moreover, China as the first author's country also always has a negative relationship, and that is lower impact in multi-institutions than that in single institution papers, which confirms the results in Table 8, where the more counties to collaborate with the better when China as the first author's country.

Based on these analyses, we find that both the number of authors and China as the first author's country have a negative relationship with the citations; the number of main institutions has the positively greatest impact on citations; main institutions as first author's organization has the highest impact in single institution papers, while slightly impact in multi-institution papers; the number of countries has a slightly higher impact than that of USA as first author's country and a moderately greater impact that the number of institutions.

Discussion and conclusions

This study aims to explore how the institutional research performance affect the productivity trends and citation impact. Using the AH-index method, we rank all the institutions extracted from a large-scale data and divide the institutions into two types: main institutions (M type) and normal institutions (N type). After that, we divide all the collaboration papers into six parts as six institutional collaboration patterns (M, M&M, M&N, N, N&M and N&N). Based on the six patterns, we propose three specific questions as shown in Introduction part.

The results of the first question show that M type (35.6%) accounts for a huge percentage much higher than other types, and it is more common to collaborate among institutions at the same type (M&M and N&N papers: 10.8% and 8.8%, M&N and N&M papers: 7.5% and 7.8%). According to the productivity trends, M and M&M papers are the main trends in the beginning, and after a certain time, N and N&N papers become the main collaboration types. According to the citation impact, the results show that the citation value of M&M type is significantly higher than other types, and the number of main institutions has a greater impact than the main institution as first author's affiliation and the number of institutions, which is confirmed by the results in the regression model, as shown in Table 9.

The second question has three parts according to the number of authors, institutions, and countries respectively.

From the number of authors, the results show that in the future, the main pattern in M type is three or more author papers, and that is two-author papers in N type, five or more author papers in M&M, M&N and N&M type, three author papers in N&N type, which indicate that the number of authors in inter-institution types has rising dramatically, while in intra-institutions, two or three authors will still be the main trend in recent years. According to the citation impact, we can see that there is a slightly negative relationship between the number of authors and citations. As we investigate several previous studies,

the relationship between number of authors and citations may differ from discipline to discipline. Therefore, the results in our research may due to the selected AI field.

From the number of institutions, we find that domestic two-institution and international two-institution papers are the main trends in all institutional collaboration types. According to the number of institutions, the results show that there is a positive relationship between the citation value and the number of institutions both in domestic and international papers, and the citation value of international papers is statistically higher than that in domestic papers. These results are consistent with some existing studies (Ibáñez et al. 2013). According to institutional collaboration patterns, we find that increasing the number of M type institutions has a positive relationship with citations, while it has no significant relationship when increasing the number of N type institutions, which indicate that the number of M institutions has a greater impact than that of the number of institutions. This is also confirmed by the results in regression models (see Table 9).

At the country level, we analyse the second question from the number of countries and some specific countries. According to the number of countries, the results show that single-country papers are the main productivity trend except M&M, the main trend of which is two-country papers, and there is a positive relationship between the number of counties and citations in whole dataset but only in M&M and N&N papers. According to the specific countries, we take P. R. China and USA as an example. The results show that although the productivity of P. R. China dataset (153,946) is significantly higher than that of USA (92,109), the productivity of M and M&M papers in P. R. China dataset is obviously lower than that in USA, which indicate that M institutions in USA published more papers than that in P. R. China. Moreover, the citation value in USA is statistically corresponding higher than that in P. R. China in all institutional collaboration patterns, and there is a positive relationship between the number of counties and the citation value in inter-institution papers only in China. We can see that it is benefit to collaborate with other countries when China as first author's country, which can be confirmed by the regression model in Table 9. While when USA as first author's country, the more counties the better in N&N papers, but the more the worse in M&M papers, which is not statistically, and need to be confirmed by other datasets. All in all, although the number of countries has a positive impact on citations in whole dataset, it differs when considering the institutional collaboration patterns and specific countries.

The last question introduces two negative binomial regression models to analysis the strength of different factors for the impact on citations, one for intra-institution papers, the other for inter-institution papers. The results show that both the number of authors and China as first author's country have no positive relationship with the citations; the number of M institutions has the greatest impact when comparing the other factors, while M type as first author's affiliation only has a slightly impact in multi-institution papers; USA as first author's country has a moderately impact on citations but slightly lower than the number of countries and moderately higher than the number of institutions.

In this paper, the division method of institution type is different with other studies (Wang et al. 2014), which is only related to the selected filed data. Maybe we could compare various division methods to explore whether the methods have an influence on the results or not in the future. Besides, our analysis is limited to collaboration papers in Artificial Intelligence, therefore, the single-author papers are not taken into account, and the results may not be generally applicable. So, in the future work, further research is required to access the above questions, and it is need to verify the results in other fields. Moreover, we are also interested in capturing the topic evolution in different institutional collaboration patterns, and explore how the main institutions affect the topic development.

Acknowledgements The authors are grateful to anonymous referees and editors for their invaluable and insightful comments, and thank for the support by the National Social Science of China (16ZDA224).

References

- Abramo, G., D'Angelo, C. A., & Solazzi, M. (2004). The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics*, *86*(3), 629–643.
- Barbara, S., Barrantes, B. S. L., Bote, V. P. G., Rodriguez, Z. C., & Anegon, F. D. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, *63*(3), 481–489.
- Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, *52*(3), 365–377.
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*(4), 861–873.
- Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology*, *62*(10), 1992–2012.
- Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: A case study of Harvard University's publications. *Scientometrics*, *87*(2), 251–265.
- Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, *63*(2), 323–335.
- Han, P., Shi, J., Li, X., Wang, D., Shen, S., & Su, X. (2014). International collaboration in LIS: Global trends and networks at the country and institution level. *Scientometrics*, *98*(1), 53–72.
- Ibáñez, A., Bielza, C., & Larrañaga, P. (2013). Relationship among research collaboration, number of documents and number of citations: A case study in Spanish computer science production in 2000–2009. *Scientometrics*, *95*(2), 689–716.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.
- Lee, D. H., Seo, I. W., Choe, H. C., & Kim, H. D. (2012). Collaboration network patterns and research performance: The case of Korean public research institutions. *Scientometrics*, *91*(3), 925–942.
- Liu, H. I., Chang, B. C., & Chen, K. C. (2012). Collaboration patterns of Taiwanese scientific publications in various research areas. *Scientometrics*, *92*(1), 145–155.
- Nguyen, T. V., Ho-Le, T. P., & Ut, V. L. (2017). International collaboration in scientific research in Vietnam: An analysis of patterns and impact. *Scientometrics*, *110*(2), 1035–1051.
- Ni, P., & An, X. (2018). Relationship between international collaboration papers and their citations from an economic perspective. *Scientometrics*, *116*(2), 863–877.
- Onodera, N., & Yoshikane, F. (2014). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology*, *66*(4), 739–764.
- Peng, T. Q., & Zhu, J. J. H. (2012). Where you publish matters most: A multilevel analysis of factors affecting citations of internet studies. *Journal of the American Society for Information Science and Technology*, *63*(9), 1789–1803.
- Puska, H. M., Muhonen, R., & Leino, Y. (2014). International and domestic co-publishing and their citation impact in different OECD fields. *Scientometrics*, *98*(2), 823–839.
- Reingewertz, Y., & Lutmar, C. (2018). Academic in-group bias: An empirical examination of the link between author and journal affiliation. *Journal of Informetrics*, *12*(1), 74–86.
- So, M., Kim, J., Choi, S., & Park, H. W. (2014). Factors affecting citation networks in science and technology: Focused on non-quality factors. *Quality and Quantity*, *49*(4), 1513–1530.
- Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, *81*(1), 177–193.
- Sud, P., & Thelwall, M. (2016). Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology*, *67*(8), 1849–1857.
- Trenchard, P. M. (1992). Hierarchical bibliometry: A new objective measure of individual scientific performance to replace publication counts and to complement citation measures. *Journal of Information Science*, *18*(1), 69–75.
- Van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, *48*(10), 944–945.
- Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, *62*(1), 133–143.

- Wang, W., Wu, Y., & Pan, Y. (2014). An investigation of collaborations between top Chinese universities: A new quantitative approach. *Scientometrics*, *98*, 1535–1545.
- Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). Scientific collaboration patterns vary with scholars' academic ages. *Scientometrics*, *112*(1), 329–343.
- Yuan, L., Hao, Y., Li, M., et al. (2018). Who are the international research collaboration partners for China? A novel data perspective based on NSFC grants. *Scientometrics*, *116*(1), 401–422.