An Analysis of the Relationship Between Author' s Academic Age and Scientific Collaboration Using Multivariate Mathematical Regression

Jingran Sheng[1], Lin Wang[2] and Xiaofan Wang[3*]

sheng_jingran@sjtu.edu.cn[1], wanglin@sjtu.edu.cn[2] and xfwang@sjtu.edu.cn[3]

Shanghai Jiao Tong University, Department of Automation, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China[1]

Shanghai University, School of Mechatronic Engineering and Automation, Shanghai, China[2]

School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai, China[3]

Abstract. This article is based on the Microsoft Academic Graph (MAG) database and selects publication and citation data in the disciplines of economics and chemistry from 1970 to 2020. The academic age of each scholar is dynamically defined, and the cluster analysis method is used to classify the academic age of the author, namely: young scholars, middle-aged scholars, and senior scholars. The authors' cooperation is divided using the maximum and minimum method. Based on existing data, a negative exponential function model is constructed to investigate the relationship between the number of publications and the type of author collaboration. The fitted parameters are statistically tested using mathematical regression. On this basis, the relationship between academic age and scientific research output is analyzed. It is concluded that collaborating with young scholars contributes more to scientific research output in terms of publication count, but there are disciplinary differences in citation count: for economics, authors of other ages groups contribute more when collaborating with middle-aged scholars; in chemistry, authors of other age groups contribute more when collaborating with young scholars. An average academic age at peak contribution under optimal collaboration scale is also given.The abstract needs to summarize the content of the paper.

Keywords: academic age; cluster analysis; skewed distribution; negative exponential function model; scientific research output

1 Introduction

With the continuous expansion of the scale of scientific research teams and the increase in interdisciplinary collaboration, many studies have shown that age structure distribution has a significant impact on scientific research output and author collaboration[1]. For example, some studies found that teams with a more balanced age structure tend to achieve higher levels of scientific research results[2]. At the same time, in scientific collaboration, authors of different ages have differences in knowledge background and research approaches, and these differences may affect scientific output[3-5]. Therefore, in-depth exploration of the relationship between scientific output, age structure distribution, and author collaboration types is of great significance for promoting scientific innovation and improving research quality.

ICBBEM 2024, March 29-31, Wuhan, People's Republic of China
Copyright © 2024 EAI
DOI 10.4108/eai.29-3-2024.2347459

In recent years, the collaboration between authors has become increasingly close and collaboration networks have become more complex[6]. These collaborative relationships have an important impact on both individual scholars and teams' research outputs. Meanwhile, the age structure distribution of authors is also an important factor affecting their research outputs. However, there is a lack of systematic study on the impact of academic age structure distribution among collaborating authors on research outputs; further discussion is needed.

2 Literature Review and Research Questions

2.1 Literature Review

Lee, J and Bozeman, B pointed out in their article "The impact of research collaboration on scientific productivity" published in "Social Studies of Science" that the literature studies the impact of collaboration between authors on scientific research output. The study found that collaborative relationships between scholars can significantly increase their research output[7–10]. Kim, H., & Jeong, D published an article in "a case study of Korean researchers" titled "Collaboration patterns, productivity, and citation impacts of aging scientists". By analyzing the collaboration patterns and research outputs of Korean scholars, they explored the impact of scholars' age on their collaboration patterns and research outputs. The study found that older Korean scholars are more inclined to collaborate with peers of the same age and compared to younger scholars, their research results are more likely to be cited [11].

Miao Yajun published a study on the academic age characteristics of scholars in the "Science Research" journal, focusing on highly cited scholars in the fields of economics and chemistry listed in the American Science Information Research Institute database. The study, based on academic productivity and influence, concluded that both these factors approximately follow the "Golden Ratio" and have a cubic polynomial distribution [12–14] with academic age.

Liu Junwan analyzed the physiological age distribution of outstanding scholars' paper influence in biology and genetics, mathematics and computer science from the perspective of citation counts, and found the peak citation counts for molecular biology and genetics scholars at ages 41-55 and for mathematics and computing science [15] at ages 31-45. Jin Bihui et al. discovered that about 50% of highly cited papers by around 70-year-old molecular biology and genetics scholars were published between ages 55-70, with another 30% published after age 60 [16].

However, it is very necessary to analyze the academic productivity and influence of scientific researchers from the perspective of academic age. For example, J.E. Hirsch proposed dividing the negative H index by academic age [17]. Du Jian et al., using variance and correlation analysis methods, focused on academic age as a primary factor to analyze researchers' performance indicators like total number of papers, total citations count, citation count per paper, h negative index, g negative index, and A negative index etc., across different combinations of academic ages [18]. Falagas et al.'s study revealed that biomedical field scholars' academic output decreases with increasing academic age [19].

Bao Wei, Jin Honghao and Tian Mingzhou, in an article published in the magazine "Higher Education Research", used a two-way fixed-effects model to analyze the relationship between the age structure of the teaching staff of research institutions and their scientific outputs. They found optimal research output when young scholars made up 51.2%, middle-aged scholars 43%, and senior scholars 5.8%. There was varying deviation from this optimal value among different types of research units and teacher aging issue will continue to worsen [20–25].

Roberta Sinatra et al., in their paper "Quantifying the evolution of individual scientific impact" published in "Science", quantified factors influencing scholars' impact throughout their scientific career by constructing a random model linking productivity, personal ability and luck revealing universal patterns of scientific success. This model assigns each scholar a unique individual parameter Q that remains stable throughout career, and it accurately predicts the evolution of a scholar's influence from a negative h-index, to cumulative citations, to independent recognition [26].

In summary, previous studies indicate significant impacts on scientific research due to collaborations, team science, and interpersonal networks. However, there is currently a lack of research on the specific age structure distribution, which is the focus of this article.

The main text should be written using Times New Roman, 10pt, fully justified. Italics can be used for emphasis and bold typeset should be avoided.

2.2 Problem Research

Based on the following hypothesis, we studied the impact of academic age structure distribution on collaboration networks and research output, which varies across different disciplines or research themes. This study attempts to study the following issues through empirical evidence and data:

First, the impact of academic age structure distribution on scientific output and excellence of scientific achievements;

Secondly, the impact of different age structure distributions on collaboration quantity, types, and their changing trends;

Third, the differences in author collaborations between different academic ages across various disciplines.

3 Research Design

3.1 Data Sources and Processing

Source of data. Our data comes from Microsoft Academic Graph (MAG), which is a database composed of papers covering multiple disciplines and fields such as biology, computer science, engineering, medicine, economics, physics, and psychology. Based on the Digital Object Identifier (DOI) and author information of the papers, we integrate raw data into a scientific dataset with unique author/paper identifiers. The entire dataset contains 150 million papers with multiple auxiliary information like publication date, title, author, keywords, references, research field and affiliated institutions. The multiple attributes of each paper facilitate our comprehensive analysis from different angles.

We first perform preliminary data cleaning. Papers with incomplete author information and unknown publication dates are deleted. Then we use the method mentioned in Sinatra's paper

[26] for name disambiguation of authors. We obtain a dataset consisting of 22 million papers and 7.9 million authors. Finally, we select articles published between 1970 to 2020 for study and categorize them into 19 different disciplines according to their research field attribute (Field of Study), focusing on economics and chemistry disciplines for research. We end up with a dataset consisting of 122,676 authors, 75,916 articles and 562,747 citations in economics; and 2,299,810 authors, 1,541,566 articles, 5,299,784 citations in chemistry.

Data processing. In data processing, relevant data is extracted from the data source including author names, publication time, citation status, journal information etc. to build networks of collaboration among scholars identifying each scholar's collaborators and frequency of collaborations. Academic age is calculated based on publication times and scholars are categorized into different academic age groups; collaboration patterns and quantities within these networks are analyzed comparing collaborations among different age groups; each scholar's research output including number of articles and citation frequencies etc. is tallied analyzing outputs among various age groups. Based on the scatter plot of the existing data, the least squares method is applied to fit the curves. The rationality of the fitting parameters is explained based on statistics by using regression analysis method. The relationships between age structure distribution, cooperation relations and research outputs are explored to validate the research hypotheses.

3.2 Research Methods

This article selects 50 years of paper data from economics and chemistry disciplines in the Microsoft Academic Graph (MAG) database. The changes in scientific research output of the two disciplines over natural years are shown in Figures 1 and 2.

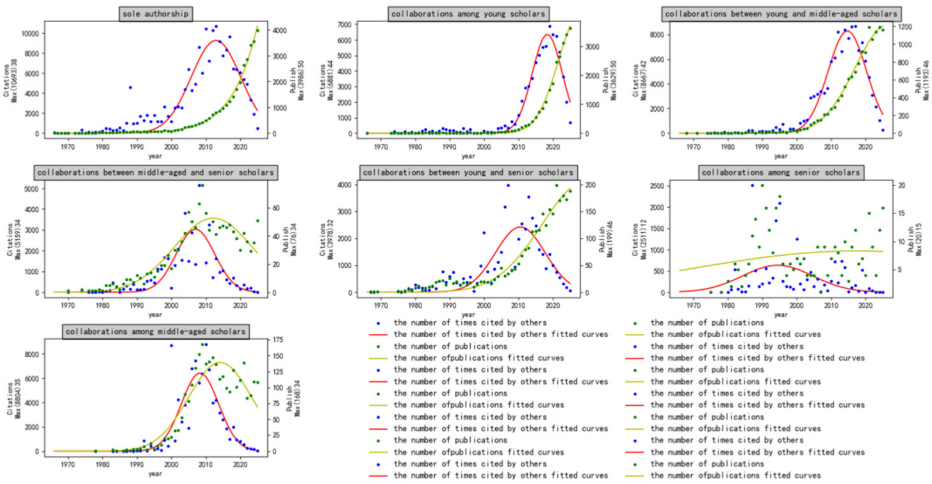

Fig. 1. Statistical chart of the number of publications and citations produced by collaborative papers (Economic Discipline).

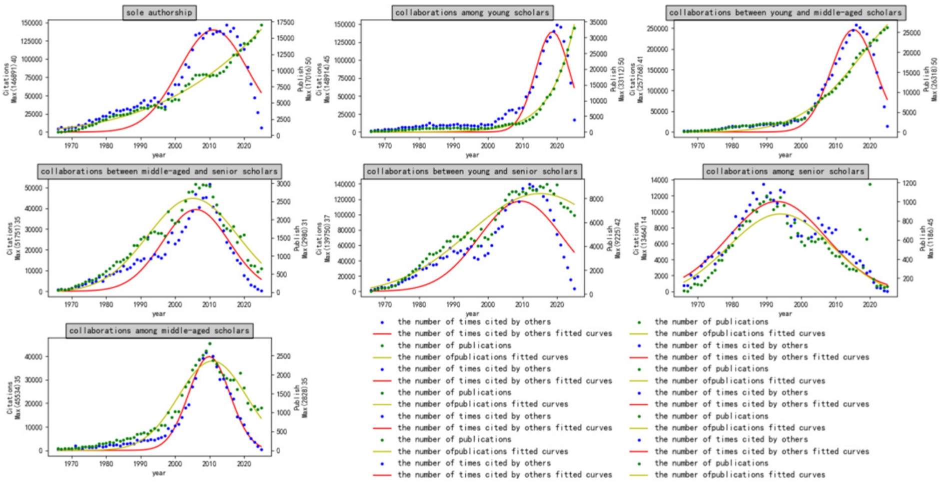

Fig. 2. Statistical chart of the number of publications and citations produced by collaborative papers (chemistry discipline).

On this basis, the academic age of scholars in each article is defined. Authors are classified using cluster analysis method into young scholars, middle-aged scholars, and senior scholars. The authors' cooperation is divided using the maximum and minimum value method. Based on existing data, the change trend of scatter plots conforms to a skewed distribution. Then we construct a negative exponential function model of the relationship between the number of publications and the type of author collaboration, invoke the curve fitting toolbox function of Python software to calculate specific fitting parameter values. Statistical analysis is conducted on the fitted parameters, and T-tests are performed on them. All P-values are less than 0.1, which verifies the correctness of the model.

Definition and calculation method of academic age structure. In research process, it's common to use the year when a scholar published the first paper as the initial year and use current time or end time of study as termination year while calculating scholar's age within this period. In order to dynamically grasp author's publication status, we define author's academic age based on the publication time of the article. The specific definition and calculation method are as follows:

Academic Age (Sa): The time interval formed from an author's first paper publication year to current paper's publication year is defined as an author's academic age in that paper.

$$Sa_j = PY_j - PY_1 + 1 \tag{1}$$

Where: PY_1 represents an author's first article publication year; PY_j represents an author's jth article publication year; j=1,2,3..., N.

Average Academic Age (ESa): arithmetic mean of all authors' academic ages in a paper. For example, an article has n authors, and the academic age of the i-th author in the article is Sai, (i=1, 2, ...n), then the average academic age of the article is ESa = $\sum_{i=1}^{n} Sa_i / n$. In order to ensure the accuracy of the data, we strive to filter effective paper data and correct scholar information as much as possible, avoiding potential errors in calculations related to age structure and research output effects. We categorize authors' academic age and collaboration status based on their annual productivity. Using cluster analysis, scholars are divided into three categories: young scholars, middle-aged scholars, and senior scholars [14]. The academic age data indicators for the three types of scholars are: Young Scholars (y): (0, 12]; Middle-aged Scholar (m): (12, 24]; Senior Scholar(s): (24, —). We use the max-min method to divide papers in the database into seven groups according to the authors' collaboration relationships:

y0: represents solo-authored papers;

y1: represents papers co-authored by young scholars with young scholars (y+y), Max(Sa)≤2;

y2: represents papers co-authored by young scholars with middle-aged scholars (y+m), Min (Sa)≤12, 12<Max (Sa)≤24;

y3: represents papers co-authored by young scholars with senior scholars (y+s), Min(Sa)≤12, Max(Sa)>24;

y4: represents papers co-authored by middle-aged scholars and middle-aged scholars (m+m), 12<Min(Sa)≤24, 12<Max(Sa)≤24;

y5: represents papers co-authored by middle-aged scholars with senior scholars (m+s), 12<Min(Sa)≤24, Max(Sa)>24;

y6: represents papers co-authored by senior scholars with senior scholars (s+s), Min(Sa)>24.

Variables and their operational definitions. Table 1 provides the operational definitions of the relevant variables and the symbolic representations of the dependent and independent variables.

Table 1. Relevant variables and their operational definitions.

Variables/Indicators	Operational definition	Corresponding group	Symbolic representation
Research output (dependent variable)	Total number of papers	-	Y
Output from sole authorship	Number of single-authored papers	Single-authored	y0
Output from Collaboration	Collaboration among young scholars	Number of papers in group y+y	y1
	Collaboration between young and middle-aged scholars	Number of papers in group y+m	y2
	Collaboration between young and senior scholars	Number of papers in group y+s	y3

	Collaboration among middle-aged scholars	Number of papers in group m+m	y4
	Collaboration between middle-aged and senior scholars	Number of papers in group m+s	y5
	Collaboration among senior scholars	Number of papers in group s+s	y6

It is assumed that the dependent variable in this study is scientific research output (number of papers/number of citations), and the specific observation indicator is the number of papers published in each database. The impact of author collaboration on scientific research output is assessed from the perspective of the number of papers. In addition, this study also uses the number of other citations as another observation indicator for scientific research output to further validate the analysis conclusions.

Model methodology. The data analyzed in this study is panel data, offering more sufficient information from both time development and individual cross-sectional dimensions. The sample data spans nearly 50 years; thus, effects resulting from time evolution should be considered along with differences across disciplines. Therefore, a two-way fixed-effects model that simultaneously controls for both annual differences and disciplinary variations is deemed most appropriate. This study constructed the following model (Model 1) to explore the relationship between academic age structure of teaching staff in research institutions and their research outputs:

$$Y_i = y_{i0} + y_{i1} + y_{i2} + y_{i3} + y_{i4} + y_{i5} + y_{i6} \quad (2)$$

Here Y_i, y_{i0}, y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, and y_{i6} represent total output at an average academic age of i years and outputs for each scholar combination.

4 Empirical Analysis

4.1 Statistics on Research Outputs by Scholars At Different Academic Ages

Based on different academic ages, we conduct a detailed statistical analysis of the number of publications and citations the fields of economics and chemistry over 50 years. The details can be seen in Table 2.

Table 2. Statistics of papers published under different academic age collaborations in economics andchemistry from 1970 to 2020.

Category of discipline	Research output	Sole Au thorship	Collaboration among young scholars	Collaboration between young and middle aged scholars	Collaboration between young and senior scholars	Collaboration among middle aged scholars	Collaboration between middle aged and senior scholars	Collaboration between senior scholars
Economics	Publication count	31,617	23,586	13,486	2,781	2,808	1,336	365
	Citation count	177,277	70,504	117,175	47,058	88,361	46,268	16,104

Chemistry	Publication count	371,293	274,787	426,491	299,473	57,309	80,422	31,791
	Citation count	3,624,700	1,972,539	4,198,963	3,456,158	658,235	1,006,101	383,088

To better study the relationship between collaborative paper output and author collaboration, we use processed data from our database on economics (as shown in Figure 3) and chemistry (as shown in Figure 4) to create scatter plots analyzing the relationship between paper output and average academic age for each type of collaboration.

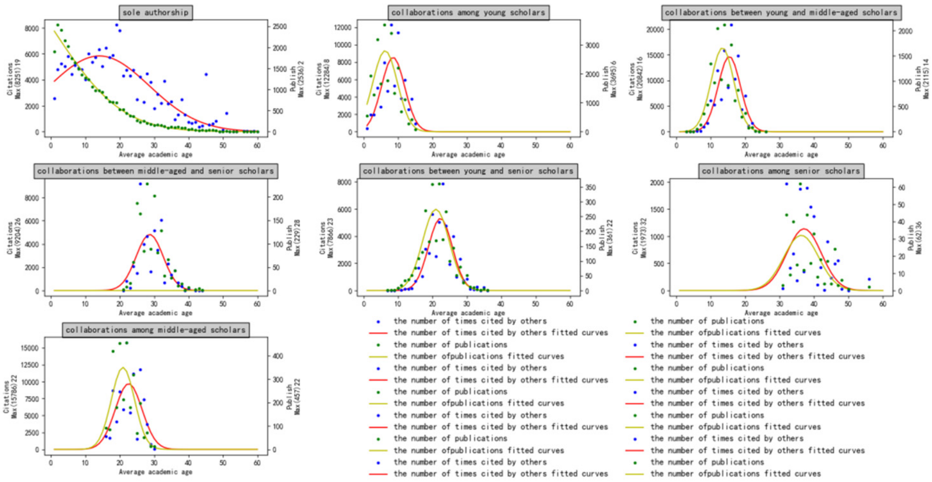

Fig. 3. Scatter plots and fitting curves of the relationship between the number of publications andcitations and the average academic age (Economics).

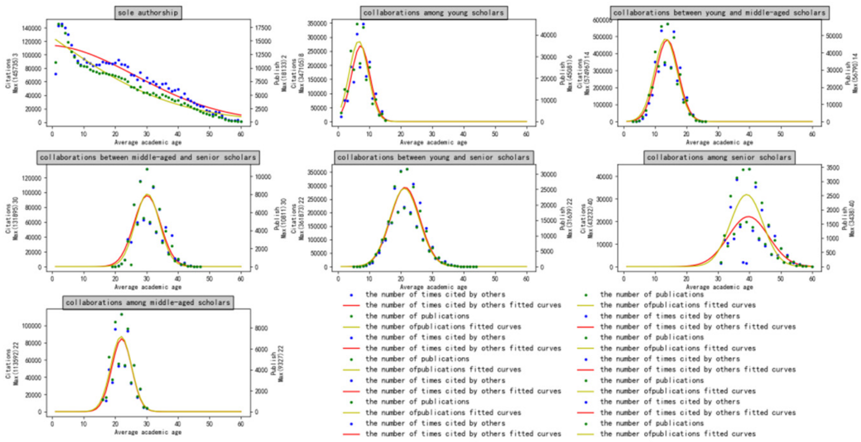

Fig. 4. Scatter plots and fitting curves of the relationship between the number of publications and citations and the average academic age (Chemistry).

The 7 graphs in Figures 3 and 4 respectively correspond to the 7 cooperative relationships proposed in this article, where the horizontal axis represents the average academic age of scholars, and the vertical axis represents the number of citations per paper. Judging from the scatter plot trends in Figures 1 and 2, they conform to a skewed distribution. Therefore, we consider using a negative exponential function to fit the curve, and then Model 1 can be improved as follows (Model 2):

$$Y_i = \sum_{j=0}^{6} a_{ij} e^{-\frac{(b_{ij}-x_j)^2}{c_{ij}}} \tag{3}$$

a_{ij}, b_{ij}, c_{ij}(j=0,1,2...6)is a parameter to be fitted. The construction idea of model2 meets the trend of scatter plots; however, its flaw is that there are too many parameters. Also, it's impossible to directly fit with the form of negative exponential functions. Therefore, we take natural logarithm for each item in Model2 and obtain an improved model as follows (Model 3)

$$\begin{cases} Y_i = y_{i0} + y_{i1} + y_{i2} + y_{i3} + y_{i4} + y_{i5} + y_{i6} \\ ln\, y_{ij} = ln\, a_{ij} - \dfrac{(b_{ij} - x_j)^2}{c_{ij}}, j = 0,1,2...6 \end{cases} \tag{4}$$

4.2 Analysis of The Relationship Between Author Collaboration and Scientific Research Output

Model calculation results. This article first targets economics discipline and proposes a correlation model (Model3) for the relationship between collaborative paper output quantity, citation quantity and average academic age. We use OLS regression model to solve related parameters of model3 while drawing relationship curve graphs between collaborative paper output and average academic age. Detailed fitting parameters are shown in Table 3.

Table 3. Parameters of fitted curve for collaborative paper output quantity / cited paper quantity (Economics).

a0	2.61680078e+05/ 15861.97246594	b0	1.27373372e+02/ 47.2840333	c0	1.13717957e+03/ 98.40941763
a1	5238.14131451/ 11681.79534818	b1	62.666394/ 53.33129544	c1	81.62953279/ 33.51233642
a2	2546.45051639/ 23530.31012855	b2	59.8211829/ 50.04783023	c2	138.09473844/ 58.03989036
a3	165.53712885/ 12743.08757384	b3	46.70842254/ 42.33243728	c3	186.1833593/ 59.31981316
a4	517.02911724/ 8630.31653719	b4	58.11775987/ 48.13033844	c4	153.58654204/ 82.95293698
a5	24.56414435/ 4409.80399299	b5	40.19071559/ 26.70305629	c5	1225.4328214/ 23.12996662
a6	343.00644424/ 17663.68587632	b6	48.92364122/ 44.29765096	C6	137.10272863/ 57.5827627

In Figure 3, the blue scatter points indicate the number of citations, and the green scatter points indicate the number of publications; while the red curve represents the fitted curve of the number of citations, and the yellow curve represents the fitted curve of the number of publications. It can be primarily concluded that the fitted curves are basically consistent with the distribution of scatter data, which indicates that our third model has an excellent fitting effect. While obtaining parameters for these curves, we also tested their validity. The specific test results are shown in Table 4.

Table 4. Results of parametric regression analysis of the number of published papers produced by co-authors/number of papers cited by others (Economics)

	Value	system error	t	P	
y0	-89.3162/ -6.9477	54.455/3.212	-1.640/-2.163	0.107/0.035	*/**
y1	-1.6754/ 4.3438	6.310/1.311	-0.266/3.313	0.492/0.002	*/***
y2	-931.2207/ -1.7978	611.219/2.468	-1.524/-0.728	0.134/0.470	*/*
y3	-1274.3652/ -25.0557	629.390/11.913	-2.025/-2.103	0.048/0.040	**/**
y4	8401.5667/ 41.9965	5527.065/21.562	1.520/1.948	0.135/0.057	*/**
y5	1823.1750/ 4.5432	710.661/2.004	2.565/2.267	0.013/0.028	**/**
y6	-1346.1645/ 9.1895	1104.993/2.942	-1.218/3.124	0.229/0.003	*/***

R2: 0.975/0.976 adapted one: R2: 0.972/0.973 *P<0.5 **P<0.1 ***P<0.01

By observing Table 4, we know that the value of R2 is 0.975/0.976 and adjusted R2 ‘s value is 0.972/0.973, indicating that our statistical model performs excellently in explaining and predicting variations in dependent variables. This excellent fit reflects our model's success in capturing relationships between independent and dependent variables.

From previous experimental analyses, we have found relationships between collaborations among scholars of different academic ages and research outputs in economics as follows:

From the perspective of citation count, the data for all seven combinations show a trend of increasing and then decreasing. For solo-authored papers, the peak citation count of 8,251 is reached when the average academic age is 19. When young scholars collaborate, they reach a peak citation count of 12,284 at an average academic age of 8. Collaborations between young and middle-aged scholars reach a peak citation count of 20,842 at an average academic age of 16. Young scholars working with senior scholars reach a peak citation count of 20,359 at an average academic age of 23. Middle-aged and senior scholar collaborations peak at 7,866 citations when the average academic age is 26. Senior scholar collaborations reach their peak citation count of 1,973 at an average academic age of 32, while middle-aged scholar collaborations peak at 15,786 citations with an average academic age of 22. In terms of the total number of citations, young, middle-aged, and senior scholars show better performance when collaborating with middle-aged scholars.

From the perspective of publication count, the number decreases as the author's average academic age increases for solo-authored papers but shows a trend to increase before decreasing for co-authored papers, indicating that collaboration significantly contributes to publication quantity. Young scholar collaborations reach their publication quantity peak (3,695) at an average academic age is six; young and middle-aged scholar collaborations peak (2,115) at fourteen; young and senior scholar collaborations have their highest output(361) when the authors' mean academic age is twenty-two; middle-aged and senior scholar partnerships hit their stride(229) around twenty-eight years old; senior-senior pairings max out (62) by thirty-six years old; mid-aged duos are most productive (457) around twenty-two years old. In terms of the total number of publications, young scholar collaborations and young and middle-aged scholar collaborations show better performance in scientific output.

In order to explore the differences in the relationship between collaborative research output and average academic age in different disciplines, we used the above model to calculate the data for the field of chemistry. We also used OLS regression models to solve the relevant parameters of Model 3 and drew a curve graph of the relationship between collaborative paper output and average academic age. Detailed fitting parameters are seen in Table 5.

Table 5. Fitting curve parameters of the number of published papers produced by coauthors/number of cited paper (Chemistry).

a0	1.48394938e+04/ 1.43911741e+05	b0	3.29051182e+00/ 1.77513553e+01	c0	1.57526764e+03/ 9.47329482e+02
a1	5.63366849e+04/ 4.46291064e+05	b1	7.06555886e+00/ 7.75862070e+00	c1	1.64939567e+01/ 1.58178609e+01
a2	1.20420865e+05/ 8.30312127e+04	b2	1.36615026e+01/ 5.73708603e+03	c2	2.69137776e+01/ -1.76901372e +07
a3	25214.14228234/ 4.10856443e+05	b3	30.56355844/ 3.06609381e+01	c3	37.7653113/ 4.00021366e+01
a4	8.29453830e+04/ 1.21215113e+06	b4	2.12378441e+01/ 2.16058311e+01	c4	5.11331992e+01/ 4.80411377e+01
a5	8329.03519327/ 1.31517137e+05	b5	40.08271885/ 4.04858677e+01	c5	52.22702349/ 5.45337924e+01
a6	16850.46493681/ 2.55403405e+05	b6	22.20607771/ 2.23390608e+01	C6	19.62401991/ 1.84329522e+01

In Figure 4, the blue scatter points represent the number of citations and the green scatter points represent the number of p publications; while the red curve represents the fitted curve of the citation numbers, and the yellow curve represents the fitted curve of the publication numbers. A preliminary conclusion can be drawn that the fitted curves are basically consistent with the distribution of scatter data, which indicates that the fitting effect of the third model proposed in this paper is excellent. While obtaining parameters for fitting curves, we also tested their rationality. The specific test results are shown in Table 6.

Table 6. Results of parametric regression analysis of the fitted curve by the number of published papers/citations (Economics)

	Value	system error	t	P	
y0	-905.2067/ -11.5169	219.416/1.732	-4.126/-6.651	0.000/0.000	***/***
y1	6.1522/ 8.3935	12.002/0.467	0.513/17.984	0.410/0.000	*/***
y2	440.9344/ -0.7940	129.586/ 0.207	3.403/-3.844	0.001/0.000	***/***

y3	-55.0454/ 16.1615	200.005/1.338	-0.275/12.076	0.484/0.000	*/***
y4	236.8014/ 1.9243	164.877/1.357	1.436/1.418	0.157/0.162	**/**
y5	1789.2204/ 9.8521	724.537/2.216	2.469/4.446	0.017/0.000	**/***
y6	301.6621/ 29.9730	124.465/2.282	2.424/13.135	0.019/0.000	**/***

R2: 0.995/0.997 adapted one: R2: 0.994/0.997 *P<0.5 **P<0.1 ***P<0.01

According to the test results, the R2 value is 0.995/0.997, and the adjusted R2 value is 0.994/0.997. All P values are less than 0.5, indicating that the proposed statistical model is excellent in explaining and predicting the variability of the dependent variable. This excellent fit value reflects that the model in this paper successfully captures the relationship between the independent variables and dependent variables.

Through the previous experimental analysis, we obtained the following relationship between collaborative research output and average academic age among authors in the field of chemistry.

From the perspective of citation count: when authored independently, the citation count decreases with the increase in average academic age. The number of papers published by co-authors first increases and then decreases, indicating that collaboration contributes significantly to citations. When young scholars collaborate, the citation count peaks at 347,105 papers at an average academic age of 8; it reaches 574,967 papers at an average age of 14 when young and middle-aged scholars collaborate; it hits a peak of 361,873 papers at an average age of 22 for collaborations between young and senior scholars; for middle-aged and senior scholars collaborating, the peak is 131,895 papers at an average age of 30; senior scholars collaborating reach a peak of 43,232 papers at an average age of 40; middle-aged scholars collaborating peak at 113,592 papers at an average age of 22. In terms of total citations, young scholars, middle-aged scholars, and senior scholars show good performance in scientific research output when collaborating with young scholars.

From the perspective of publication count: when working alone, the publication count decreases with the increase in average academic age. The number of co-authored publications also shows a trend of first increasing then decreasing over time. Young scholar collaborations reach a peak publication count of 45,081 articles when their mean academic age is six years; collaborations between young and mid-aged rscholars reach their zenith with a total article count reaching up to 56,790 when their mean academic age is fourteen years old; partnerships involving both young and middle-aged scholars hit their pinnacle with as many as 31,639 articles when averaging twenty-two years. Collaborations between middle-aged and senior scholars see their highest output totaling 10,811 articles upon reaching thirty years into their careers on average. When senior scholars collaborate, the peak number of publications is 3,438 when the average academic age is 40; when middle-aged scholars collaborate, the peak number of publications is 9,372 when the average academic age is 22. In terms of total publications, young scholars, middle-aged scholars, and senior scholars show good performance in scientific research output when collaborating with young scholars.

5 Conclusion and Inspiration

5.1 Research Conclusions

Based on the statistics of papers of economics and chemistry in Microsoft Academic Graph (MAG) database for 50 years, this paper analyzes the relationship between academic age and scientific research output, and uses cluster analysis to effectively group the data. A negative exponential function model of the relationship between the number of papers published and author collaboration types is constructed by using OLS regression and F-scale method, and the results are statistically tested. The conclusions obtained are of high reliability.

For the discipline of economics, in terms of the number of citations, young scholars, middle-aged scholars and senior scholars contribute more to the scientific research output when collaborating with middle-aged scholars. The peak contribution ages for collaborations are 16 (young-middle), 22 (middle-middle), and 28 (middle-senior). Young scholars excel in paper publication when collaborating amongst themselves or with middle-aged scholars at average academic ages of 6 (young-young) and 14 (young-middle).

For chemistry, in terms of the number of citations, young, middle-aged, and senior scholars contribute more when collaborating with young scholars. The peak contribution ages are 8 (young-young), 14 (young-middle), and 22 (young-senior). Young scholars also perform well in paper publication when collaborating with all age groups at these same average academic ages.

5.2 Discussion and Insights

Due to database boundary limitations in this study, there's a margin of error for the academic ages of young and senior scholars which might affect calculation accuracy. Therefore, we selected a large sample size to minimize this effect, including 120,000 authors in economics, over 70,000 articles, and over 500,000 other citations, as well as 2,000,000 authors, 1,500,000 articles, and 15,000,000 citations in chemistry. The result is that the larger the amount of data, the smaller the impact of the boundary. From the perspective, research trends are feasible, but model accuracy still needs improvement—a direction for our future work. For example, we can expand database boundaries by ten years, selecting intermediate data for modeling, and use data correction methods.

5.3 Limitations and Prospects

Our data from the Microsoft Academic Graph database covers only economics and chemistry subjects; thus it has certain limitations in depth and breadth due to its single source nature. This article only considers the effects of collaboration between authors in different academic ages on research output, and some other factors are not taken into account. Future research directions include considering factors like gender, professional title, educational background, etc., on scientific and technological cooperation and research output. Moreover, we should expand the research scope across different countries, regions and fields [27–30]. In addition, we should integrate machine learning and big data technologies to build more effective mathematical models to achieve more refined predictions and management of research collaboration and output [31–33].

References

[1] Milojevi´c S. How are academic age, productivity and collaboration related to citing behavior of researchers? PloS one 2012, 7(11), e49176.
[2] Liang Z, Ba Z, Mao J, Li G. Research complexity increases with scientists' academic age: Evidence from library and information science. Journal of Informetrics 2023, 17(1), 101375.
[3] Zhe C, Lu X, Xiong X. Analysis of influence factors on the quality of international collaboration research in the field of social sciences and humanities: The case of Chinese world class universities (2015–2019) Sage Open 2021, 11(4), 21582440211050381.
[4] Hou L, Pan Y, Zhu J. Impact of scientific, economic, geopolitical, and cultural factors on international research collaboration. Journal of Informetrics 2021, 15(3), 101194.
[5] Jung I. International Collaboration in Educational Technology Research: A Personal Reflection on Research Process, Experience, and Outcomes. Information and Technology in Education and Learning 2022, Volume 2 ,Issue 1, PP Inv-p003-Inv-p003.
[6] Zhang J, Yang X, Hu X, Li T. Author Cooperation Network in Biology and Chemistry Literature during 2014-2018: Construction and Structural Characteristics. Information 2019, Volume 10,Issue7, PP 236-236.
[7] Lee, J., Bozeman, B. The impact of research collaboration on scientific productivity. Social Studies of Science 2005, 35(5), 673-702.
[8] L. Ductor.Does co-authorship lead to higher academic productivity, Oxf. Bull. Econ. Stat 2015, 77(3), 385–407.
[9] K.C. Garg, S. Dwivedi.Pattern of collaboration in the discipline of Japanese encephalitis. DESIDOC J. Libr. Inf. Technol 2014, 34(3),241–247.
[10] A. ˇCavušoˇglu, Türker. Patterns of collaboration in four scientific disciplines of the Turkish collaboration network. Physica A 2014, 413, 220–229.
[11] Kim, H., Jeong, D. Collaboration patterns, productivity, and citation impacts of aging scientists. A case study of Korean researchers 2017, 12(4).
[12] MIAO Yajun, QI Wei, ZHONG Qi. Study on the academic age characteristics of scientists: based on the two dimensional of academic productivity and academic influence. Studies in Science of Science 2013, 31(2), 177-183.
[13] M. Krapfi. Age and complementarity in scientific collaboration. Empir. Econ 2015, 49(2), 751–781.
[14] W. Wang, S. Yu, T.M. Bekele, X. Kong, F. Xia. Scientific collaboration patterns vary with scholars' academic ages. Scientometrics 2017, 112 (1), 329–343.
[15] Liu Junwan. Social Age Analysis of the Influence of Outstanding Scholar Papers. Journal of the China Society for Scientific and Technical Information 2010, 29(1), 121-127.
[16] Liu Junwan, Jin Bihui. An age distribution of productivity of scientific papers for highly cited scientists. Science Research Management 2009, 30(3), 96-103.
[17] Hirsch J E. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences 2005,102(46), 16569-16572.
[18] Du Jian, Zhang Bin, Li Yang, Tang Xiaoli, Xu Peiyang. Optimization of the Evaluation Indicators of Scholars'Research Impact and Comparative Analysis Between National and International Academic Behaviors of Researchers. Library and Information Service 2011, 55(10) , 98-102.
[19] Falagas M E, Ierodiakonou V, Alexiou V. At what age do biomedical scientists do their best work? . The FASEB Journal 2008, 22(12), 4067-4070.

[20] Bao Wei, Jin Honghao, Tian Mingzhou. The Relationship Between the Age Structure of Factly and the Research Performance of Research Universities of China . Journal of Higher Education 2020, 41(5), 54-62.
[21] GINGRASY,LARIVIERE V,MACALUSO B ,et al. The Effects,of,Aging on Researchers' Publication and Citation Patterns. PloS one 2008, 12, e4048.
[22] ZUCKERMAN H, MERTON R K. Age,Aging,and Age Structure in Science. Higher Education 1972, 2, 1-4.
[23] LEHMAN H C. Age and Achievement Princeton University Press;2017; pp.3-21.
[24] DIAMONDA M. The Life-cycle Research Productivity of Mathematicians and Scientists. Journal of Gerontology 1986, 4, 520-525.
[25] LEVIN S G,STEPHAN PE. Age and Research Productivity of Academic Scientists. Research in Higher Education 1989, 5, 531-549.
[26] Sinatra, D. Wang, P. Deville, C. Song, A. Barabasi. Quantifying the evolution of individual scientific impact. Science 2016, 354(6312),aaf5239.
[27] C. Zhe, X. Lu, X. Xiong. Analysis of influence factors on the quality of international collaboration research in the field of social sciences and humanities: The case of Chinese world class universities (2015–2019). SAGE Open 2021, 11(4), 21582440211050381.
[28] Esva B, Jccd, E, Aurora A, Teixeira A. Which distance dimensions matter in international research collaboration? A cross-country analysis by scientific domain. Informetr 2022, 16, 101259.
[29] A. Abbasi, A. Jaafari. Research impact and scholars' geographical diversity. Informetr 2013, 7(3), 683–692.
[30] U. Finardi. Scientific collaboration between BRICS countries. Scientometrics 2015, 102(2), 1139–1166.
[31] W. Wang, F. Xia, J. Wu, Z. Gong, H. Tong, B.D. Davison. Scholar2vec: vector representation of scholars for lifetime collaborator prediction. ACM Trans. Knowl. Discov. Data (TKDD) 2021, 15(3), 1–19.
[32] Zhang X, Zhang Z, Chen X. Collaborative Features of Authors Based on Academic Journal Papers and Their Influence on Scientific Research Output-Taking Highly Published Authors of International Medical Informatics as an Example. Journal of the China Society for Scientific and Technical Information 2019, 38(1), 29-37.
[33] Z. Chinchilla-Rodríguez, B. Vargas-Quesada, Y. Hassan-Montero, A. González-Molina, F. Moya-Anegóna. New approach to the visualization of international scientific collaboration. Information Visualization 2010, 9(4),277–287.