Mutual Influence in Citation and Cooperation Patterns

Chenbo Fu, *Member, IEEE*, Haogeng Luo, Xuejiao Liang, Yong Min, *Member, IEEE*, Qi Xuan, *Senior Member, IEEE*, and Guanrong Chen, *Life Fellow, IEEE*

Abstract*— Measuring the influence of scientists and their activities on science and society is important and indeed essential for many studies. Despite the substantial efforts devoted to exploring the influence's measures and patterns of an individual scientific enterprise, it remains unclear how to quantify the mutual impact of multiple scientific activities. This work quantifies the relationship between the scientists' interactive activities and their influences with different patterns in the AMiner dataset. Specifically, inflation treatment and field normalization are introduced to process the big data of paper citations as the scientist's influence, and then the evolution of the influence is investigated for scientific activities in the citation and cooperation patterns through the Hawkes process. The results show that elite scientists have higher individual and interaction influences than ordinary scientists in all patterns found in the study, with permutation tests verifying the significance of the new findings. Moreover, the study compares the patterns found in two largest disciplines, i.e., *STEM* and *Humanities*, revealing the higher value of individual influence in *STEM* than in *Humanities*. Furthermore, it is found that the opposite trend of *STEM* and *Humanities* in the cooperation pattern suggests different cooperation habits of scientists in different disciplines. Overall, this investigation provides a feasible approach to addressing the scientific influence issue and deepening the quantitative understanding of the mutual influence of multiple scientific activities in science and society.

***Index Terms*— Citation pattern, Hawkes process, interaction influence, science of science, scientific activity.**

I. Introduction

SCIENTIFIC activity, such as production [1], citation [2], [3], and cooperation [4], [5], is key to the development of modern science. However, the ever-increasing production makes science activities more complex and overwhelming, becoming harder and harder to analyze. Fortunately, in the last decade, the ongoing process of datafication has been continuously turning most scientific activities into computerized data. Notably, these digitized data are collected and analyzed to support related works associated with the "science of science" [6], [7], [8]. In particular, one of the current foci is the influence of scientific activities [4], [9], [10]. Despite the present good understanding about the impact of an individual scientific enterprise (i.e., scientists, journals, and institutions) on different metrics [11], [12], [13], [14], it remains unclear how to quantify the integrating impact of scientific activities among multiple scientists, e.g., citation and cooperation.

Uncovering the mechanisms of the interaction activities among scientists and their evolution is critical for understanding and evaluating the influence of scientists' activities. However, scientific influence is a complex concept that is difficult to quantify. One intuitive approach is to evaluate such influence using citation [15], [16] and some relevant metrics (e.g., H-index [11], reputation index [17], and journal impact factor [18]). In particular, the citation is specifically recorded across different disciplines. Unfortunately, with the increase in knowledge scale and complexity, the citation distribution exhibits a high degree of heterogeneity [19], [20], [21], affecting citation-related measurement methods' effectiveness. For example, the citation of a scientist with a big name continuously grows with time, showing the powerful Matthew effect [22], [23]. Substantial efforts have been devoted to extracting such influence from the complex scientific environment, e.g., distinguishing personal impacts from teamwork [24] and evaluating the influence from the visibility of scientists [25] or papers [26]. Furthermore, to offer a deep quantitative understanding of the self-organizing behaviors of scientists and unveil the influence caused by scientific activities, more and more investigations have been conducted on the behavioral patterns of scientists [27], [28], [29]. Among these scientific patterns, citing and collaborating are the two most basic scientific activities. Recent studies found that citing higher impact papers indeed has a positive feedback [30]. Moreover, cooperation will fully mobilize decentralized knowledge, allowing each scientist in the team to finish their work [31], enabling researchers in different fields to share responsibilities and risks, gain complementary knowledge, accumulate academic capital, and accelerate the exchange of

Manuscript received 29 October 2022; revised 7 April 2023, 22 May 2023, and 25 August 2023; accepted 7 October 2023. This work was supported in part by the Key Research and Development Program of Zhejiang under Grant 2022C01018, in part by the National Natural Science Foundation of China under Grant U21B2001 and Grant 61973273, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LGF21G010003 and Grant LR19F030001. *(Corresponding author: Chenbo Fu.)*

Chenbo Fu, Haogeng Luo, Xuejiao Liang, and Qi Xuan are with the Institute of Cyberspace Security and the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China, and also with the Binjiang Cyberspace Security Institute, ZJUT, Hangzhou 310056, China (e-mail: cbfu@zjut.edu.cn; 2112003091@zjut.edu.cn; 2112103145@zjut.edu.cn; xuanqi@zjut.edu.cn).

Yong Min is with the Computational Communication Research Center, Beijing Normal University, Zhuhai 519087, China, and also with the School of Journalism and Communication, Beijing Normal University, Beijing 100875, China (e-mail: myong@bnu.edu.cn).

Guanrong Chen is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China (e-mail: eegchen@cityu.edu.hk).

Digital Object Identifier 10.1109/TCSS.2023.3325264

knowledge and experience, thus accelerating the output of innovative research [32], [33]. For example, recent research shows that cooperating with top scientists has a competitive advantage in one's future career [34].

Although these studies offer clear patterns for scientific activities, the influence caused by such activities is still a challenge to understand and quantify, especially for the effect of mutual activities. Therefore, in this study, we aim to systematically quantify the influence caused by mutual scientific activities from a new perspective. Specifically, using the Hawkes process, our work quantifies the influence of mutual activities in different behavioral patterns. The Hawkes process [35], [36] is a kind of self-exciting point process, which is often used to quantify the impact between events at different time. Because the Hawkes process reveals the weak causal relationship between time and events, it has been widely used in different disciplines, for example, in modeling earthquakes and their aftershocks [37], and estimating the market risk [38]. Recently, this method has been expanded to analyze the influence of different users on social networks [39], [40], [41], [42].

With the above-mentioned studies, the main contributions of this work are summarized as follows.

1) Our work examines a total of 4.9 million papers over the past 20 years and introduces the inflation treatment and field normalization to avoid heterogeneity in calculating their contribution.
2) We quantify the influence of mutual scientific activities using the Hawkes process and investigate such influences in citation and cooperation patterns. Furthermore, we perform permutation tests to verify the significance of our results.
3) We reveal and analyze the disciplinary differences in citation and cooperation patterns.

The rest of this article is organized as follows. Section II introduces the dataset and preparation. Section III describes the inflation treatment, field normalization, and the Hawkes process. Section IV demonstrates the influence effects in the aforementioned two patterns. Section V comprises the differences in *STEM* and *Humanities*. Finally, Section VI concludes the investigation with some discussions.

II. Data Description and Preparation

In our study, we use the dataset from AMiner.[1] As the release dataset version continues to update, it has become more popular and used for analyzing the information spread [43], studying the scientific influence [44], [45], [46], building recommendations in academic networks [47], [48], researching citation and cooperation networks [49], [50], [51], [52], developing the prediction in academic networks [53], [54], and analyzing in different fields [55], [56], [57], [58]. This work adopts the twelfth version of the dataset (V12), which was last updated in 2020 and extracted from database systems and logic programming (DBLP), Association for Computing Machinery (ACM), and Microsoft Academic Graph (MAG) [59]. The dataset includes nearly 4.9 million papers from 113 887 disciplinary fields. Each data contain the paper number, paper title, scientists' information, publication year, publication location, citation relationship, and field information. The field information includes field name and fields of research weight w (called "fields of study" in [60]), and each paper is assigned to at least two fields.

TABLE I
Data Information

Object	#
Paper	3,054,175
Scientist	2,721,481
Citation	21,152,324
Cooperation	8,698,923

To quantify the continuous inference among scientists, we exclude the scientists who have not published more than two papers in three consecutive years. Furthermore, since the number of citations received by a paper will reach its maximum three years after publication and the citations can be decline after three years [61], [62]. In this work, we design the accumulation years of citations for different years. Specifically, we demonstrate the individual influence μ_m and interaction influence α_{mn} under different cumulative years of citations, i.e., three, five, and seven years. It should be noted that we focus on papers from 2000 to 2018. Thus, papers published in 2018 lack the accumulation of citations over periods of five and seven years, while those published in 2015 lack the accumulation of citations over seven years. There are totally 3 054 175 papers coauthored by 2 721 481 scientists. Table I shows the detailed information of the datasets, where cooperation means coauthorship.

III. Method

In this work, we chose the number of citations to quantify the scientists' contributions and influence. However, citations and papers have increased exponentially at different growth rates [63], which leads to inflation problem and causes the consequence that citations cannot be directly compared in different periods. Furthermore, papers in different disciplines have large variations in the number of citations [20]. These issues make it impossible to directly compare the numbers of citations across disciplines for different periods. To solve this problem, we introduce citation inflation treatment and field normalization, which are necessary before calculating scientists' contributions. Moreover, we attribute each paper to the first scientist to avoid double-counting and also to simplify the calculation.

A. Inflation Treatment

Because the inflation and research field have been proven to be uncorrelated, we directly adopt a method from [64] without considering the research fields. Specifically, we assume the increment

$$\Delta = ln\left(N_{(t)}\right) - ln\left(N_{(0)}\right) = I \times t \tag{1}$$

[1]https://www.aminer.cn/citation

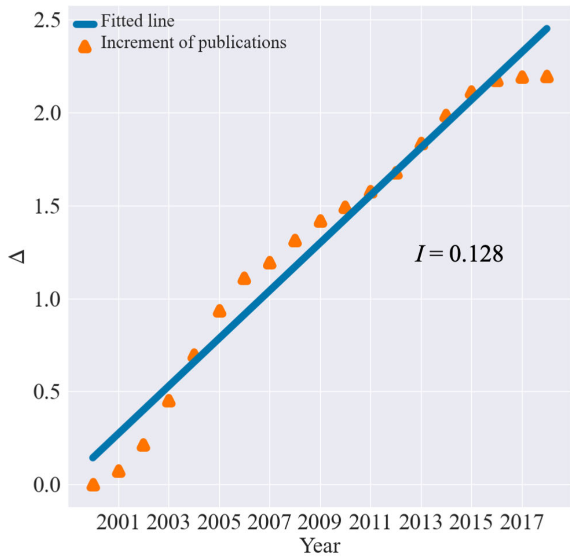


Fig. 1. Relationship between time and the increment Δ. The orange dot represents the increment of publication, and the blue solid line represents the fit result with the linear regression model. The citation inflation coefficient I is 0.128.

where $N_{(t)}$ is the total number of papers published in t years, and $N_{(0)}$ is the number of papers published in the starting year. The inflation coefficient I can be fit by the ordinary least squares (OLS). Then, the adjusted or the actual number of citations, $\overline{C(t)}$, can be calculated as $\overline{C(t)} = C(t)/(I + 1)$, where $C(t)$ is the observed citation number in year $t + 1$.

As shown in Fig. 1, the citation inflation coefficient I is 0.128 from 2000 to 2018. In this work, we set the base year to 2000 and normalize the number of citations obtained in the other years to the actual number of citations in 2000.

B. Field Normalization

To avoid the bias problem introduced by field variation, field normalization is needed. In this work, we extend the method developed in [20], to introduce a weight of the paper field given by the dataset used. To avoid amplifying the contribution of fields with low average citations, this work uses the number of publications N_j to normalize. Specifically, we define the citation contribution p of paper by

$$p = \sum_{j=1}^{n} \frac{\left(\sum_{t=2000}^{2018} \overline{C(t)}\right) * w_j}{N_j}. \tag{2}$$

Here, n is the total number of fields covered by paper i, j is one of the fields ($j = 1, 2, 3, \ldots, n$), and w_j is the weight of the jth field, which is called "fields of study" and directly provided by the dataset; the field to which the paper belongs and the field of research with weight w are given by the dataset; $\overline{C(t)}$ is the number of citations after inflation treatment in year t; and N_j is the total number of papers published in the jth field. Then, the contribution of scientist A who has m publications can be calculated as follows:

$$P_A = \sum_{i=1}^{m} p_i \tag{3}$$

where m is the total number of papers published by scientist A.

After calculation, we divide the scientists by ranking their contributions in descending order, where the top 10% are classified as elite scientists and the last 90% as ordinary scientists. Finally, we found that there are 235 537 elite scientists and 2 119 836 ordinary scientists. It should be noted that although AMiner inherits the lowest level field of MAG, this work performs field normalization for all the fields involved. Thus, the normalization results are not affected by the field level.

C. Hawkes Process

Our goal in this work is to quantify the mutual influence of multiple scientists' activities in an academic field. Traditional works focus on analyzing individual influence of scientists [63], [65], [66], ignoring the quantification of interactions among multiple scientists. However, quantifying the influence among scientists is not easy because the combined multisources' nature of the influence makes it challenging to determine the influence's origin and intensity.

Fortunately, the Hawkes process as a particular point process can explain the effects of such influence mathematically. A point process is a stochastic process that uses a collection of points to represent discrete events in a mathematical space [67], [68]. As a particular class of point processes, the Hawkes process was first proposed in 1971, which considers that the occurrence probability of the current event is dependent on the occurrence of the previous events [35], [36]. In particular, the Hawkes process treats the dependence intensity between events as a decay function in exponential form. Thus, the conditional intensity function for a Hawkes process can be calculated as follows:

$$\lambda(t) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)} \tag{4}$$

where μ is the baseline intensity of the event, t is the current time, t_i is the time at which the previous event occurs before time t, α is the excitation factor, and β is the decay rate. In this formula, α and β establish the dependence of the current event on the previous event.

In the academic network, the scientists' self-citations can be considered as a self-excitation process, where the scientists' accumulative citations can be used to quantify the total influence [expressed as $\lambda(t)$], which includes individual and historical influences. Specifically, the individual influence is an intrinsic property of a scientist (e.g., creativity [69]), and the historical influence is the excitation of the scientists' historical events (e.g., the excitation of the scientists' self-citation history). For a scientist, the individual influence is defined as μ, which is independent of the scientists' historical influence, and the historical influence is the excitation of self-citations at time t_i ($t_i < t$) as α, which decreases with the decay rate β.

However, a more common phenomenon in academic fields is that scientists are influenced by other scientists, such as collaborators, which can be considered as cross-excitation. For example, a recent report shows that teamwork has a more substantial influence than solo work [24]. The importance of the collaborators' interactions promotes our consideration of the cross-excitation process in the Hawkes process, i.e., a multivariate Hawkes process. The multivariate Hawkes process is more suitable for quantifying the influence of scientists' interactions. Compared with the univariate Hawkes process, the multivariate Hawkes process not only includes self-excitation but also considers cross-excitation. Specifically, the multivariate Hawkes process is written as follows:

$$\lambda^m(t) = \mu_m + \sum_{n=1}^{N} \sum_{t_i^n < t} \alpha_{mn} e^{-\beta_{mn}(t - t_i^n)} \tag{5}$$

where N is the total number of events in n-dimension, μ_m is the baseline intensity of event m, t is the current time, and t_i^n is the time that the nth event occurs before time t; the excitation factor α_{mn} describes the excitation of the nth event on the event m, and β_{mn} is the decay rate that explains the decaying process.

In academic fields, compared with self-excitation, the difference in the cross-excitation is the composition of the total influence. The cross-excitation considers the total influence including individual influence and interaction influence. The interaction influence is the excitation of other scientists' activities (e.g., the excitation of other scientists' citations or cooperation). Specifically, for a scientist, the interaction influence of being cited or cooperating with other scientists at time t_i ($t_i < t$) can be expressed as α_{mn}, decreasing with the decay rate β_{mn}. For example, at time t, one can observe that scientist c receives two citations at times t_1 and t_2 from scientists a and b, respectively. The conditional intensity function of scientist c, then, can be written as follows:

$$\lambda^c(t) = \mu_c + \alpha_{ca} e^{-\beta_{ca}(t - t_1)} + \alpha_{cb} e^{-\beta_{cb}(t - t_2)} \tag{6}$$

where μ_c is the baseline intensity of scientist c, $\lambda^c(t)$ is the accumulative citation of scientist c (ending at time t), α_{ca} is the excitation factor that describes the interaction influence from scientist a citing scientist c, and β_{ca} is the decay rate.

IV. Influence Effect

In this work, β is the decay rate of the excitation of interaction, and we adopt the Tree of Parzen Estimators approach based on Bayesian hyperparameter optimization to fit β in different patterns (the citation and cooperation patterns) [42], [70], [71], [72], [73]. The value of β is 0.028 and 0.29 in the citation and cooperation patterns, respectively, implying that the half-life of excitation factor α is about 25 years and 2 years. This means that the interaction influence among scientists persists over an extended time. Thus, in the following experiments, we set $\beta = 0.028$ and 0.29 in different patterns to fit the baseline intensity μ and the interaction influence α in different citation accumulation years. However, the trend in individual influence across different citation accumulation times is without significant differences. Thus, we focused on the three-year citation accumulation year. We will also show their evolution in different time windows.

A. Citation Pattern

To measure the interaction influence among different scientists, we divide scientists into elite and ordinary scientists by considering their contributions (Section II). One source of

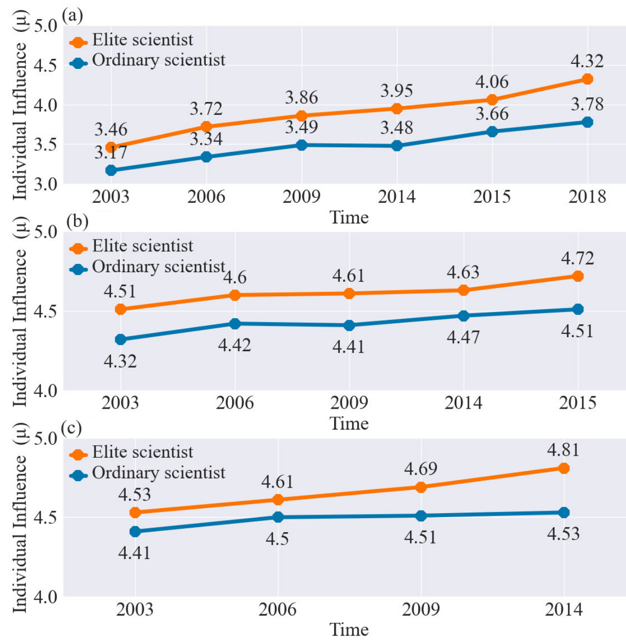


Fig. 2. Evolution of individual influence (μ) in the citation pattern. The orange line represents elite scientists and the blue line represents ordinary scientists in (a) three, (b) five, and (c) seven accumulation citation years, respectively.

the interaction influences in an academic field is the citing activities among scientists. Thus, in the citation pattern, we use the multivariate Hawkes process to quantify the evolution of the interaction influence.

1) Individual Influence: For the multivariate Hawkes process, μ_m is the baseline intensity of event m. In the citation pattern, the baseline intensity μ_m is the individual influence of scientist m, namely, the personal charisma of scientists. Fig. 2 shows comparison of the baseline intensity μ_m of elite and ordinary scientists. From the figure, one can observe an ascending trend for both elite and ordinary scientists, which implies that the scientists' individual influence is steadily improving in the citation pattern. Furthermore, comparing Fig. 2(a) with Fig. 2(b), one can find that the individual influence of elite scientists is more significant than that of ordinary scientists.

2) Interaction Influence: In addition to the citations influenced by the personal charisma of scientists, the mutual citing activities among scientists can also bring new citations. In the citation pattern, this interaction influence is represented by α_{mn}, which comes from the citing activities of scientist n, i.e., citing scientist m's papers. Fig. 3 shows comparison of the interaction influence α_{mn} of elite and ordinary scientists. Similarly, one can find an ascending trend of interaction influence α_{mn} for both elite and ordinary scientists, suggesting that the interaction influence among scientists is growing steadily. Moreover, comparing Fig. 2 with Fig. 3, the value of interaction influence α_{mn} is more significant than the baseline intensity μ_m, which implies that in the citation pattern, compared with scientists' charisma, mutual citation behavior is more helpful to the scientists' total influence. Furthermore, from Fig. 3(a) and (b), one can see that these citations from

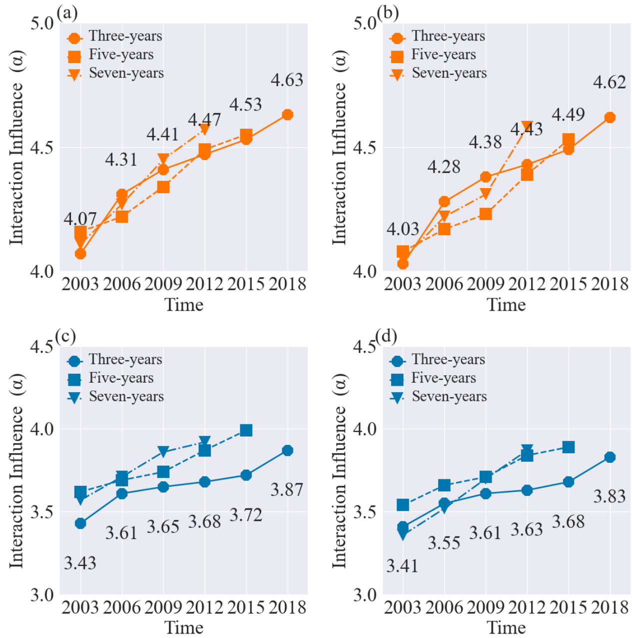


Fig. 3. Evolution of interaction influence (α) in the citation pattern. The orange line represents the interaction influence of elite scientists on (a) elite scientists and (b) ordinary scientists, respectively. The blue line represents the interaction influence of ordinary scientists on (c) elite scientists and (d) ordinary scientists, respectively. Where dot, square, and triangle represent three, five, and seven accumulation citation years, respectively.

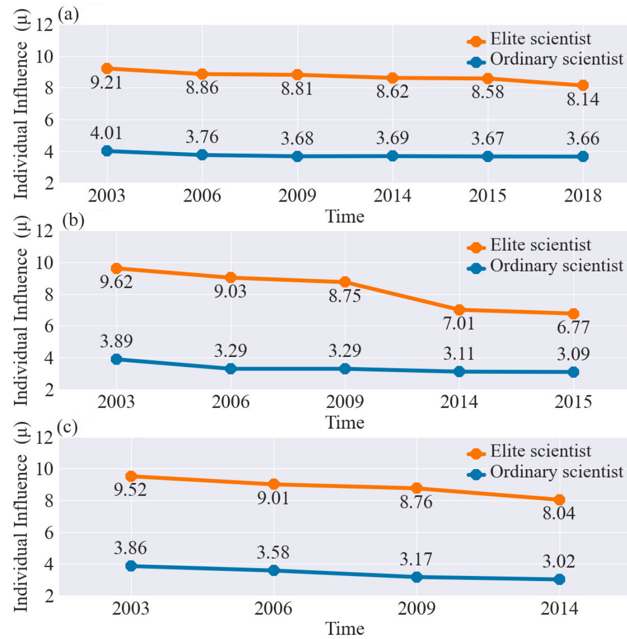


Fig. 4. Evolution of individual influence (μ) in the cooperation pattern. The orange line represents elite scientists and the blue line represents ordinary scientists, in (a) three, (b) five, and (c) seven accumulation citation years, respectively.

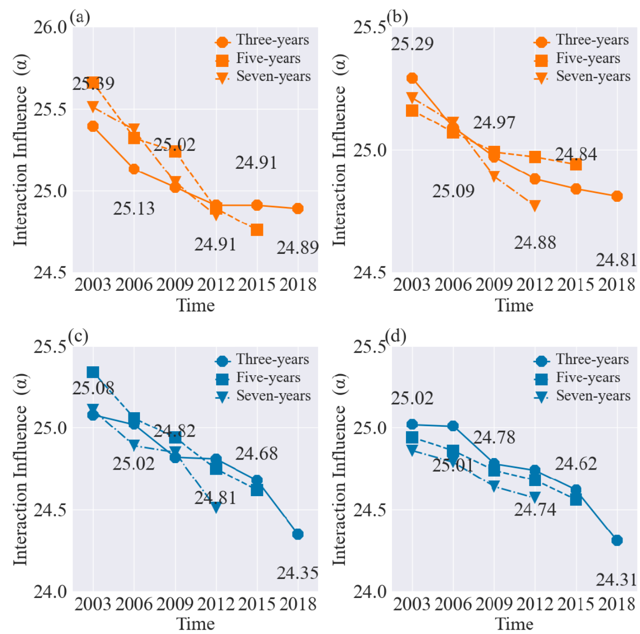


Fig. 5. Evolution of interaction influence (α) in the cooperation pattern. The orange line represents the interaction influence of elite scientists on (a) elite scientists and (b) ordinary scientists, respectively. The blue line represents the interaction influence of ordinary scientists on (c) elite scientists and (d) ordinary scientists, respectively. Where dot, square, and triangle represent three, five, and seven accumulation citation years, respectively.

the elite scientists strongly impact the target scientists, for both elite and ordinary scientists. This seems due to the fact that elite scientists receive more attention than ordinary scientists, thereby indirectly increasing the attention of those papers cited by elite scientists. Furthermore, comparing Fig. 3(a) and (c), the influence of elite scientists citing elite scientists' papers is greater than that of ordinary scientists citing elite scientists' papers. That is mainly because elite scientists gain greater attention through their high visibility, and papers cited by elite scientists also have advantages in receiving citations. Comparing Fig. 3(c) and (d), the interaction influence of ordinary scientists citing elite scientists' papers is almost at the same level as that of ordinary scientists citing ordinary scientists' papers, which implies the weak interaction influence of ordinary scientists.

In general, in the citation pattern, the elite scientist plays a more important role, which helps promote the influence of other scientists by their influences.

B. Cooperation Pattern

Another common source of influence in academic fields is the cooperative activities among scientists. We further investigate the impact of the cooperation pattern by focusing on the cooperative activities between elite and ordinary scientists. For simplicity, we attribute each paper to the first author and consider the interaction influence as the influence of other collaborators on the first author. Then, we examine whether any detectable variation in scientists' interaction influences under the cooperation pattern exists.

1) Individual Influence: Same as the citation pattern, let μ_m be the baseline intensity of scientist m. Fig. 4 shows comparison of the baseline intensity μ_m of elite and ordinary scientists. We can observe a descending trend of individual influence for both elite and ordinary scientists. However, since the total influence $\lambda^m(t)$ is growing over time, this declining trend implies that cooperative activities may play an important role in the cooperation pattern.

2) Interaction Influence: The interaction influence α_{mn} in the cooperation pattern comes from scientist n coauthored with scientist m. Fig. 5 shows comparison of the interaction

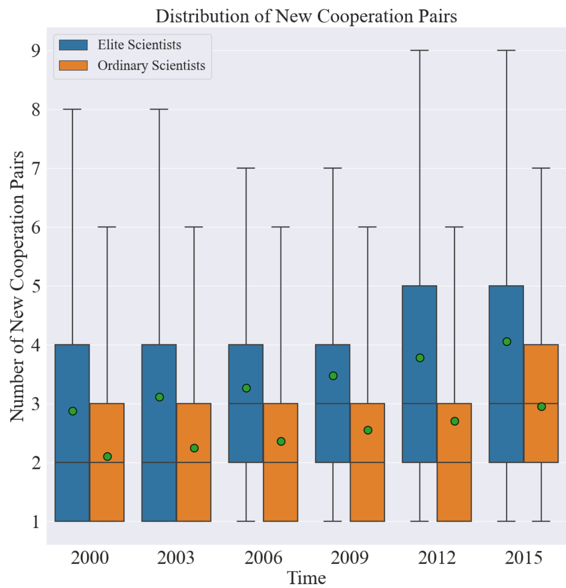


Fig. 6. Distribution of new cooperation pairs. The blue and orange boxes represent the distributions of new cooperation pairs of elite and ordinary scientists, respectively. The green dot represents the average number of new cooperation pairs. The black line in boxes represents the median number of new cooperation pairs.

influence α_{mn} of elite and ordinary scientists. Interestingly, the evolution of the interaction influence in the cooperation patterns (Fig. 5) shows an opposite trend compared with that in the citation pattern (Fig. 3). The decrease in the influence of single pair of cooperation and the increase in the total interaction influence may be caused by the number of increasing cooperation times. Recent research has found that too much new cooperation could be harmful [74], including the high cost of forming new ties [75], new cooperation members' adaption [76], and less trust and familiarity [77], which may be the reason for the decreasing of interaction influence.

To understand the potential source that might reduce the interaction influence α_{mn}, we further investigate the new cooperation pairs of the scientists. Fig. 6 compared the distribution of new cooperation, which increases steadily in each period, for both elite and ordinary scientists. Comparing Fig. 5 with Fig. 4, the interaction influence α_{mn} is significantly larger than the baseline intensity μ_m, implying that the scientists' total influence depends more on cooperation. Furthermore, from Fig. 5(a) and (b), one can see that cooperating with elite scientists influences the target scientist greatly, for both elite and ordinary scientists, suggesting the high impact of the elite scientists. By comparing Fig. 5(a) and (c), it can be seen that the interaction influence of elite scientists on elite scientists is more significant than that of ordinary scientists on elite scientists. This phenomenon may be caused by the "star effect" [78] of elite scientists, e.g., the papers published by elite scientists get more attention than that published by ordinary scientists. Moreover, there is a "win–win" situation in the cooperation among elite scientists, and the influence which is generated is more significant than simply summing their effects together. Comparing Fig. 5(c) and (d), one can see that the interaction influence of ordinary scientists on elite scientists is slightly greater than that of ordinary scientists on ordinary scientists, implying that ordinary scientists' participation in the cooperation provides limited interaction influence on the other scientists.

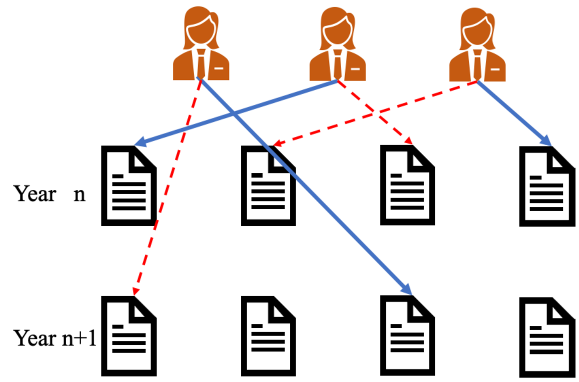


Fig. 7. Simple illustration of the permutation test. The blue solid line represents the original relationship and the red dotted line indicates the result after permutation. The permutation keeps the occurrence time and the total number of activities constant but destroys the original relationships between papers and scientists.

In general, cooperative activities play an important role in the cooperation pattern, which brings greater impact than individuals do. However, too much new cooperation may decrease the trend of interaction influence. Furthermore, elite scientists can bring a greater impact on cooperation, and ordinary scientists can benefit from cooperating with elite scientists, which is consistent with the recent reports in the literature on scientists' cooperation [5].

C. Permutation Tests

To verify the significance of the excitation effects, we conduct permutation tests of citation and cooperation patterns. Specifically, we randomly permute the association of activity types (citations or cooperation by scientists) to generate a null model. Fig. 7 demonstrates an example of possible permutation. It should be noted that our procedure keeps the occurrence time and the number of activities constant but destroys the original event relationship. After that, we refit the parameters of the multivariate Hawkes processes over the null model and compare the differences in Hawkes process parameter values, as shown in Fig. 8. The comparison results show that the ascending or descending trend in the original citation and cooperation patterns are either significantly attenuated or eliminated in the null model. Furthermore, several model lines are below null model lines because the permutation breaks the citation and cooperation relationship between scientists. For example, the value of the null model is higher than the model in Fig. 8(c) and (d) which may be because the citation and cooperation relationship between ordinary scientists and other scientists replaces with the relationship between the elite scientists and other scientists. Thus, the null model produces a greater effect than the original model.

To further verify the significance of the influence of citation or cooperation, we use the Z-test to compare the differences in the interaction influence α_{mn} between the null model and the original ones. Specifically, denote by Z_o the original interaction influence α and by Z_p the value after permutation,

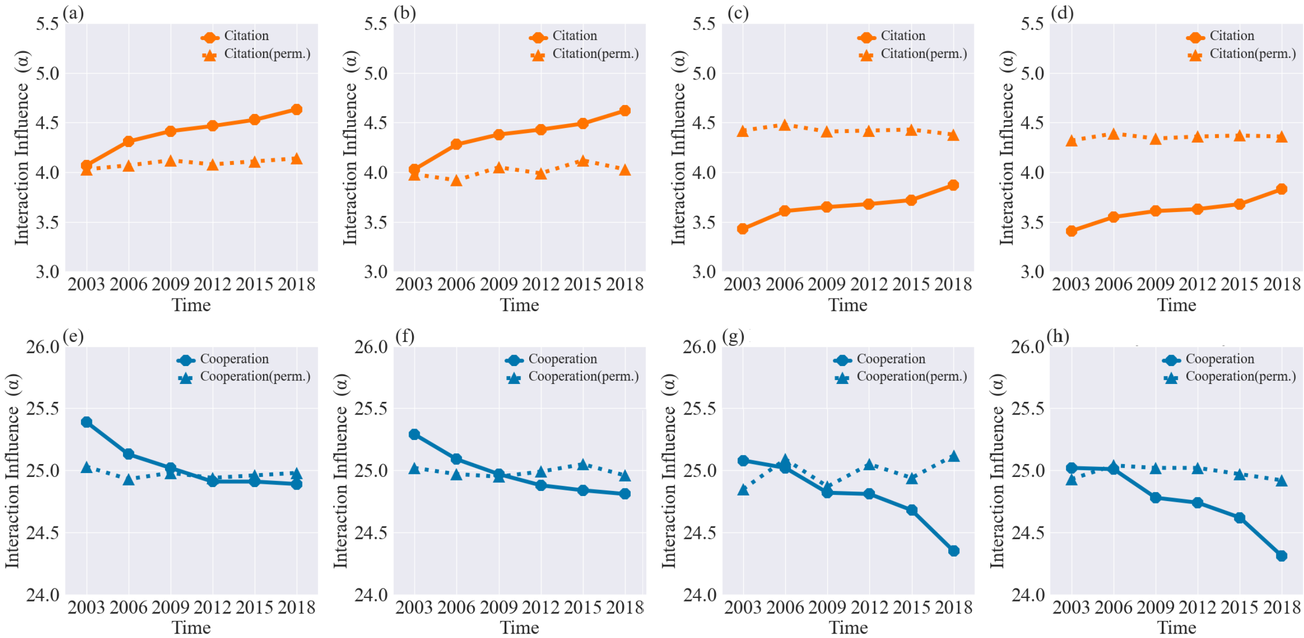


Fig. 8. Results of permutation tests in (a)–(d) citation pattern (denoted by orange solid and dotted lines) and (e)–(h) cooperation pattern (denoted by blue solid and dotted lines).

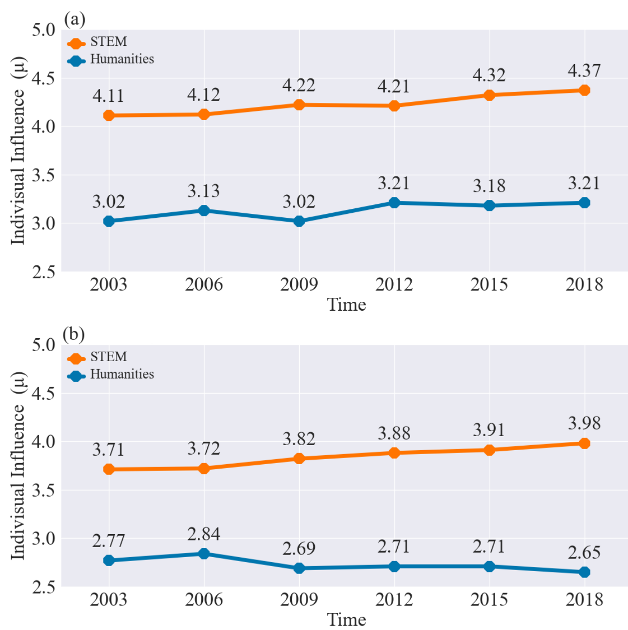


Fig. 9. Evolution of individual influence (μ) of the citation pattern in *STEM* (orange line) and *Humanities* (blue line). (a) Individual influence of elite scientists. (b) Individual influence of ordinary scientists.

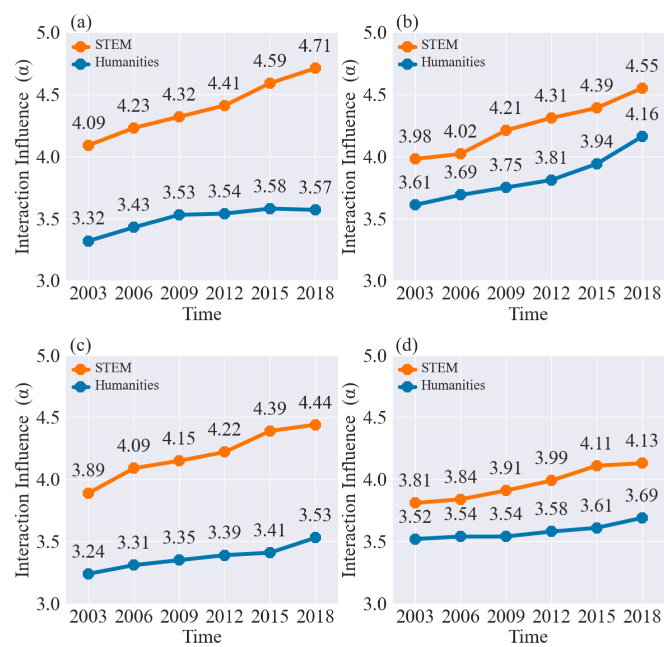


Fig. 10. Evolution of interaction influence (α) of the citation pattern in *STEM* (orange line) and *Humanities* (blue line). (a) Elite on elite scientists. (b) Elite on ordinary scientists. (c) Ordinary on elite scientists. (d) Ordinary on ordinary scientists.

where σ is the standard deviation among them. Then, the Z-score is calculated by

$$Z_{\text{score}} = \frac{Z_o - Z_p}{\sigma}. \tag{7}$$

The results show that all the p-values are smaller than 0.01, indicating that the influence caused by citing or cooperative activities is a significant phenomenon in the academic fields, which confirms that interaction among scientists is essential to developing scientists' influences.

V. Comparison in Disciplines

Since citation and collaboration habits are field-dependent, many features exhibit different characteristics in different disciplines, e.g., Sleeping Beauty features [79], Novelty features [80], and Coauthor Rate [81]. Here, we further investigate the citation and cooperation patterns in the two largest disciplines, i.e., *STEM* and *Humanities*. For simplicity, each major discipline contains ten subdisciplines with the highest number of publications. It should be noted that it is difficult to ensure a similar quantity of papers for all subdisciplines.

TABLE II

DISCIPLINES' CHARACTERISTICS

Discipline	Sub-discipline	Range of #Papers	#Sub-disciplines
STEM	Cluster analysis, Key distribution in wireless sensor networks, Image segmentation, Cloud computing, Wireless sensor network, The Internet, Computer science, Energy consumption, Scale-space segmentation, Artificial neural network.	[6399, 13781]	10
Humanities	Population, Humanities, Philosophy of logic, Music information retrieval, Mandarin Chinese, Chinese characters, French, Art history, Readers–writers problem, German.	[28, 7752]	10

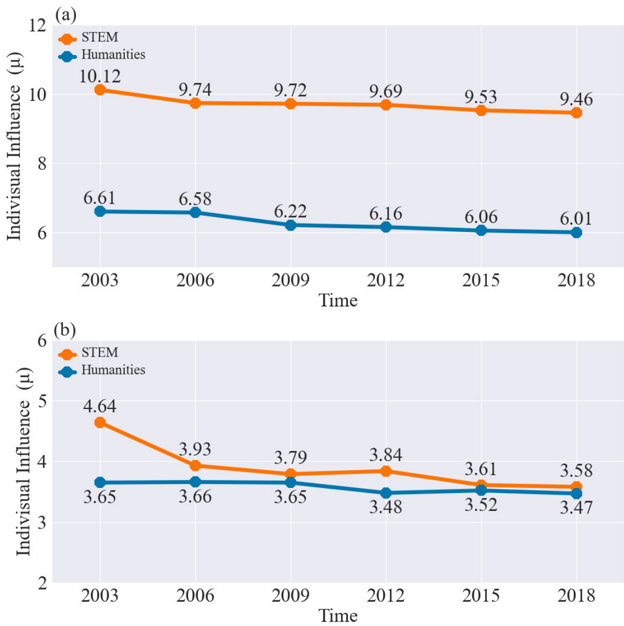


Fig. 11. Evolution of individual influence (μ) of the cooperation pattern in *STEM* (orange line) and *Humanities* (blue line). (a) Individual influence of elite scientists. (b) Individual influence of ordinary scientists.

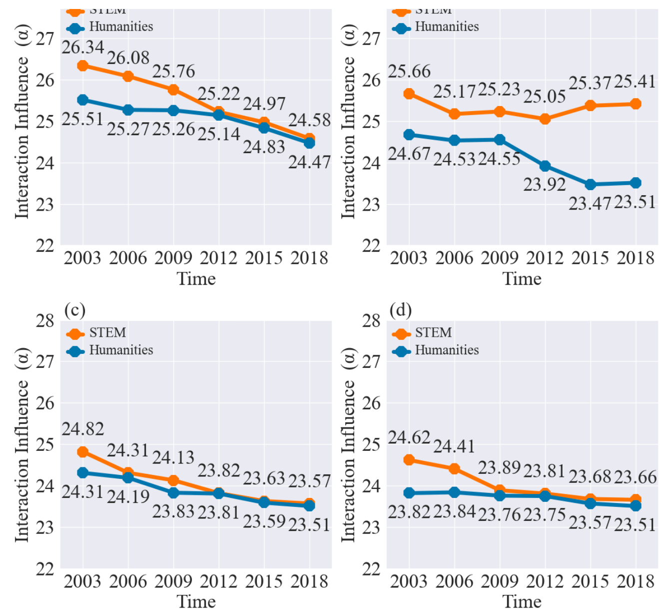


Fig. 12. Evolution of interaction influence (α) of the cooperation pattern in *STEM* (orange line) and *Humanities* (blue line). (a) Elite on elite scientists. (b) Elite on ordinary scientists. (c) Ordinary on elite scientists. (d) Ordinary on ordinary scientists.

Our treatment only picked the most popular fields, which may approximately represent the current development trends in *STEM* and *Humanities*. Their characteristics are shown in Table II.

A. Comparison by Disciplines in the Citation Pattern

We first investigate the differences between disciplines in the citation pattern. Figs. 9 and 10 show plots of the evolution of scientists' individual and interaction influences in *STEM* and *Humanities* over time in the citation pattern, respectively. As can be seen from Figs. 9 and 10, the general trends of individual influence (μ) and interaction influence (α) are similar across all the disciplines, in both *STEM* and *Humanities*. However, a larger value of the individual influence in *STEM* than *Humanities* still indicates slight differences in the ability of scientists from different disciplines to attract publications. Furthermore, as can be seen from Fig. 10, the impact of citing activities on scientists has significant difference between *STEM* and *Humanities*, i.e., the interaction influence of *STEM* grows faster than *Humanities*.

B. Comparison by Disciplines in the Cooperation Pattern

To understand the impact of collaborative activities across different disciplines, we now investigate the differences between disciplines in the cooperation pattern. Figs. 11 and 12 show plot of the evolution of scientists' individual and interaction influences in *STEM* and *Humanities* over time in the cooperation pattern, respectively. The general trends of individual influence and interaction influence are similar across all the disciplines, in both *STEM* and *Humanities*, which further implies the importance of cooperation. However, compared with ordinary scientists, the individual influence of elite scientists in *STEM* is significantly stronger than that in *Humanities*, underlining the different abilities of elite scientists from different disciplines to attract publications (Fig. 11). Furthermore, comparing Figs. 11 and 12, interaction influence is significantly stronger than individual influence, implying a higher impact of cooperation in *STEM* or *Humanities*. Moreover, the interaction influence between elite scientists in different disciplines gradually converges after 2012 [Fig. 12(a)], while the interaction influence between elite scientists and ordinary scientists gradually diverges after 2009 [Fig. 12(b)], suggesting that the cooperation patterns of elite–elite scientists are similar in different disciplines, but there are still large differences in ordinary–elite scientists' collaboration. The cooperation habits in different disciplines may be the cause of this phenomenon, i.e., the significant differences in cooperation rates for different disciplines [82] and the smaller team sizes in *Humanities* [8]. In addition, the interaction influence coming from ordinary scientists' cooperation is still at a low level [Fig. 12(c) and (d)]. Comparing with elite scientists [Fig. 12(a) and (b)], the influence of ordinary scientists is weak, for both *STEM* and *Humanities*.

VI. Conclusion and Discussion

To summarize, our present work investigates the influence of scientists in different patterns using the multivariate Hawkes process. Specifically, by dividing the scientist's influence into individual and interaction influences, we quantify the scientist's influence in the citation and cooperation patterns, respectively. Our results show that elite scientists have a greater impact than ordinary scientists in both the patterns, demonstrating the significant effect brought about by the academic interactions of elite scientists. Our results also validate the recent studies about top scientists, e.g., the competitive advantages of cooperating with top scientists [34] and the powerful Matthew effect of the top scientists [23]. Moreover, our permutation tests show the significance of the excitation effects, highlighting that the interaction activities among scientists (i.e., citation and cooperation) play an important role in academic fields. Furthermore, our comparisons in *STEM* and *Humanities* show that elite scientists' individual and interaction influences are more significant than ordinary scientists in all the disciplines and patterns. Besides, the opposite trends of interaction influence evolution of *STEM* and *Humanities* in elite–elite and ordinary–elite scientists' cooperation suggest that the latent cooperation habits for different disciplines are different, e.g., cooperation rates or team size. Generally speaking, our work provides a feasible view for the development of academic fields, showing the main sources of scientists' influences in different patterns and revealing the main differences across scientists and disciplines. Our results also uncover the weak causal relationship between excitation effects in academic fields and their temporal evolution.

However, our work still has some limitations. First, our work is based on the AMiner dataset and has yet to expand or supplement other datasets, which means that the dataset's quality may lead to our conclusion containing some bias. For example, the AMiner dataset focuses on *STEM* papers and includes some papers in *Humanities*, which is unfair when discussing the differences between *STEM* and *Humanities*. However, none of the existing large-scale datasets focuses on *Humanities*, whether AMiner or MAG. Furthermore, the AMiner dataset only labels the lowest level of fields, which may also be too fine-grained, and the underlying technology of the delineation of the fields (i.e., clustering method for big data) is controversial in terms of its accuracy and reliability [83]. Second, we extend the method developed in [20] to avoid enlarging the contribution in the fewer publications' fields. Through this method, although the impact of low average citations is eliminated, the contributions of all the papers are normalized to a lower level. In future work, we will consider a more reasonable method of field normalization to calculate the contributions of papers in various fields. Third, considering disciplinary differences, fewer papers are published in *Humanities* than in *STEM*, leading to relatively less research in *Humanities*. Thus, our future work will adopt more appropriate classification methods or more suitable datasets for comparing fields, especially *STEM* and *Humanities*. Finally, evaluating author contributions is a challenging task, which is also a problem encountered by the existing scientific credit system [84], [85], [86]. However, scientific works are the result of the joint efforts of all the collaborators, and this work only considers the contribution of the paper to the first author, which potentially inflates the contribution of the first author and ignores the efforts of other collaborators. Thus, our future work will try to find a more equitable method for allocating contributions to reveal better the potential influence of the citation and cooperation between scientists. Hopefully, our findings could be further expanded to uncover scientists' interaction influences in other patterns of the academic fields, e.g., the interdisciplinarity pattern and the differences in interaction influences among scientists in different career stages.

References

[1] B. F. Reskin, "Scientific productivity and the reward structure of science," *Amer. Sociol. Rev.*, vol. 42, no. 3, pp. 491–504, 1977.
[2] S. E. Cozzens, "Taking the measure of science: A review of citation theories," *Newslett. Int. Soc. Sociol. Knowl.*, p. 16, May 1981.
[3] J. Westbrook, "Identifying significant research: Literature citation counting is evaluated as a means for identification of significant research," *Science*, vol. 132, no. 3435, pp. 1229–1234, 1960.
[4] S. Wuchty, B. F. Jones, and B. Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, vol. 316, no. 5827, pp. 1036–1039, May 2007.
[5] G. J. Borjas and K. B. Doran, "Which peers matter? The relative impacts of collaborators, colleagues, and competitors," *Rev. Econ. Statist.*, vol. 97, no. 5, pp. 1104–1117, Dec. 2015.
[6] A. Zeng et al., "The science of science: From the perspective of complex systems," *Phys. Rep.*, vols. 714–715, pp. 1–73, Nov. 2017.
[7] A. Clauset, D. B. Larremore, and R. Sinatra, "Data-driven predictions in the science of science," *Science*, vol. 355, no. 6324, pp. 477–480, Feb. 2017.
[8] S. Fortunato et al., "Science of science," *Science*, vol. 359, no. 6379, p. 185, 2018.
[9] Y.-H. Eom and H.-H. Jo, "Generalized friendship paradox in complex networks: The case of scientific collaboration," *Sci. Rep.*, vol. 4, no. 1, pp. 1–6, Apr. 2014.
[10] J. Li, Y. Yin, S. Fortunato, and D. Wang, "Scientific elite revisited: Patterns of productivity, collaboration, authorship and impact," *J. Roy. Soc. Interface*, vol. 17, no. 165, Apr. 2020, Art. no. 20200135.
[11] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, Nov. 2005.
[12] S. Lehmann, A. D. Jackson, and B. E. Lautrup, "Measures for measures," *Nature*, vol. 444, no. 7122, pp. 1003–1004, 2006.
[13] E. Garfield, "The history and meaning of the journal impact factor," *J. Amer. Med. Assoc.*, vol. 295, no. 1, pp. 90–93, 2006.
[14] L. Waltman et al., "The Leiden ranking 2011/2012: Data collection, indicators, and interpretation," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 12, pp. 2419–2432, Dec. 2012.
[15] D. J. D. S. Price, "Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front," *Science*, vol. 149, no. 3683, pp. 510–515, 1965.
[16] L. Waltman, "A review of the literature on citation impact indicators," *J. Informetrics*, vol. 10, no. 2, pp. 365–391, May 2016.
[17] A. M. Petersen et al., "Reputation and impact in academic careers," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 43, pp. 15316–15321, Oct. 2014.
[18] E. Garfield, "Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies," *Science*, vol. 178, no. 4060, pp. 471–479, 1972.
[19] Q. Zhang, N. Perra, B. Gonçalves, F. Ciulla, and A. Vespignani, "Characterizing scientific production and consumption in physics," *Sci. Rep.*, vol. 3, no. 1, pp. 1–9, Apr. 2013.
[20] F. Radicchi, S. Fortunato, and C. Castellano, "Universality of citation distributions: Toward an objective measure of scientific impact," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 45, pp. 17268–17272, Nov. 2008.

[21] L. Waltman, N. J. van Eck, and A. F. J. van Raan, "Universality of citation distributions revisited," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 1, pp. 72–77, Jan. 2012.

[22] R. K. Merton, "The Matthew effect in science: The reward and communication systems of science are considered," *Science*, vol. 159, no. 3810, pp. 56–63, 1968.

[23] J. Brainard, "Reviewers award higher marks when a paper's author is famous," *Science*, vol. 377, no. 6612, p. 1251, 2022.

[24] M. Ahmadpoor and B. F. Jones, "Decoding team and individual impact in science and invention," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 28, pp. 13885–13890, Jul. 2019.

[25] I. Tahamtan, A. Safipour Afshar, and K. Ahamdzadeh, "Factors affecting number of citations: A comprehensive review of the literature," *Scientometrics*, vol. 107, no. 3, pp. 1195–1225, Jun. 2016.

[26] N. A. Ebrahim, H. Salehi, M. A. Embi, F. H. Tanha, H. Gholizadeh, and S. M. Motahar, "Visibility and citation impact," *Int. Educ. Stud.*, vol. 7, no. 4, pp. 120–125, Mar. 2014.

[27] T. Jia, D. Wang, and B. K. Szymanski, "Quantifying patterns of research-interest evolution," *Nature Human Behaviour*, vol. 1, no. 4, pp. 1–7, Mar. 2017.

[28] A. Zeng et al., "Increasing trend of scientists to switch between topics," *Nature Commun.*, vol. 10, no. 1, pp. 1–11, Jul. 2019.

[29] C. Fu, X. Yue, B. Shen, S. Yu, and Y. Min, "Patterns of interest change in stack overflow," *Sci. Rep.*, vol. 12, no. 1, pp. E4671–E4680, Jul. 2022.

[30] S. Mammola, D. Fontaneto, A. Martínez, and F. Chichorro, "Impact of the reference list features on the number of citations," *Scientometrics*, vol. 126, no. 1, pp. 785–799, Jan. 2021.

[31] F. M. H. Neffke, "The value of complementary co-workers," *Sci. Adv.*, vol. 5, no. 12, Dec. 2019, Art. no. eaax3370.

[32] F. W. Kabo, N. Cotton-Nessler, Y. Hwang, M. C. Levenstein, and J. Owen-Smith, "Proximity effects on the dynamics and outcomes of scientific collaborations," *Res. Policy*, vol. 43, no. 9, pp. 1469–1485, Nov. 2014.

[33] E. Leahey, "From sole investigator to team scientist: Trends in the practice and study of research collaboration," *Annu. Rev. Sociol.*, vol. 42, no. 1, pp. 81–100, Jul. 2016.

[34] W. Li, T. Aste, F. Caccioli, and G. Livan, "Early coauthorship with top scientists predicts success in academic careers," *Nature Commun.*, vol. 10, no. 1, pp. 1–9, Nov. 2019.

[35] A. G. Hawkes, "Point spectra of some mutually exciting point processes," *J. Roy. Stat. Soc., Ser. B Methodol.*, vol. 33, no. 3, pp. 438–443, Oct. 1971.

[36] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971.

[37] A. G. Hawkes, "Cluster models for earthquakes-regional comparisons," *Bull. Int. Stat. Inst.*, vol. 45, no. 3, pp. 454–461, 1973.

[38] V. Chavez-Demoulin, A. C. Davison, and A. J. McNeil, "Estimating value-at-risk: A point process approach," *Quant. Finance*, vol. 5, no. 2, pp. 227–234, Apr. 2005.

[39] M. Dubey, P. Srijith, and M. S. Desarkar, "HAP-SAP: Semantic annotation in LBSNs using latent spatio-temporal Hawkes process," in *Proc. 28th Int. Conf. Adv. Geographic Inf. Syst.*, 2020, pp. 377–380.

[40] S. Li, X. Gao, W. Bao, and G. Chen, "FM-Hawkes: A Hawkes process based approach for modeling online activity correlations," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 1119–1128.

[41] P. Bao, "Modeling and predicting popularity dynamics via an influence-based self-excited Hawkes process," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 1897–1900.

[42] T. Santos, S. Walk, R. Kern, M. Strohmaier, and D. Helic, "Self- and cross-excitation in stack exchange question & answer communities," in *Proc. World Wide Web Conf.*, May 2019, pp. 1634–1645.

[43] S. Molaei, S. Babaei, M. Salehi, and M. Jalili, "Information spread and topic diffusion in heterogeneous information networks," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, Jun. 2018.

[44] A. Fronczak, M. J. Mrowinski, and P. Fronczak, "Scientific success from the perspective of the strength of weak ties," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, Mar. 2022.

[45] T. Amjad, Y. Rehmat, A. Daud, and R. A. Abbasi, "Scientific impact of an author and role of self-citations," *Scientometrics*, vol. 122, no. 2, pp. 915–932, Feb. 2020.

[46] L. Wan, M. Zhang, X. Li, L. Sun, X. Wang, and K. Liu, "Identification of important nodes in multilayer heterogeneous networks incorporating multirelational information," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 6, pp. 1715–1724, Dec. 2022.

[47] X. Liu, K. Wu, B. Liu, and R. Qian, "HNERec: Scientific collaborator recommendation model based on heterogeneous network embedding," *Inf. Process. Manage.*, vol. 60, no. 2, Mar. 2023, Art. no. 103253.

[48] Y. Li, R. Wang, G. Nan, D. Li, and M. Li, "A personalized paper recommendation method considering diverse user preferences," *Decis. Support Syst.*, vol. 146, Jul. 2021, Art. no. 113546.

[49] R. W. Wang and F. Y. Ye, "Simplifying weighted heterogeneous networks by extracting h-structure via s-degree," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019.

[50] G. Siudem, B. Żogała-Siudem, A. Cena, and M. Gagolewski, "Three dimensions of scientific impact," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 25, pp. 13896–13900, Jun. 2020.

[51] F. Mercorio, M. Mezzanzanica, V. Moscato, A. Picariello, and G. Sperlí, "DICO: A graph-DB framework for community detection on big scholarly data," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 4, pp. 1987–2003, Oct. 2021.

[52] H. Shakibian and N. Moghadam Charkari, "Mutual information model for link prediction in heterogeneous complex networks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–16, Mar. 2017.

[53] E. Bütün, M. Kaya, and R. Alhajj, "A supervised learning method for prediction citation count of scientists in citation networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Jul. 2017, pp. 952–958.

[54] X. Mo, R. Tang, and H. Liu, "A relation-aware heterogeneous graph convolutional network for relationship prediction," *Inf. Sci.*, vol. 623, pp. 311–323, Apr. 2023.

[55] B. Laufer, S. Jain, A. F. Cooper, J. Kleinberg, and H. Heidari, "Four years of FAccT: A reflexive, mixed-methods analysis of research contributions, shortcomings, and future prospects," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, 2022, pp. 401–426.

[56] O. Vásárhelyi, I. Zakhlebin, S. Milojević, and E.-Á. Horvát, "Gender inequities in the online dissemination of scholars' work," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 39, Sep. 2021, Art. no. e2102945118.

[57] M. Shin, A. Soen, B. T. Readshaw, S. M. Blackburn, M. Whitelaw, and L. Xie, "Influence flowers of academic entities," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol. (VAST)*, Oct. 2019, pp. 1–10.

[58] L. Gienapp, W. Kircheis, B. Sievers, B. Stein, and M. Potthast, "A large dataset of scientific text reuse in open-access publications," *Scientific Data*, vol. 10, no. 1, p. 58, Jan. 2023.

[59] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "ArnetMiner: Extraction and mining of academic social networks," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 990–998.

[60] A. Sinha et al., "An overview of Microsoft academic service (MAS) and applications," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 243–246.

[61] P. D. B. Parolo, R. K. Pan, R. Ghosh, B. A. Huberman, K. Kaski, and S. Fortunato, "Attention decay in science," *J. Informetrics*, vol. 9, no. 4, pp. 734–745, Oct. 2015.

[62] M. Golosovsky and S. Solomon, "Runaway events dominate the heavy tail of citation distributions," *Eur. Phys. J. Special Topics*, vol. 205, no. 1, pp. 303–311, May 2012.

[63] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, "Quantifying the evolution of individual scientific impact," *Science*, vol. 354, no. 6312, Nov. 2016, Art. no. aaf5239.

[64] A. M. Petersen, R. K. Pan, F. Pammolli, and S. Fortunato, "Methods to account for citation inflation in research evaluation," *Res. Policy*, vol. 48, no. 7, pp. 1855–1865, Sep. 2019.

[65] P. Vinkler, "An attempt of surveying and classifying bibliometric indicators for scientometric purposes," *Scientometrics*, vol. 13, nos. 5–6, pp. 239–259, May 1988.

[66] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, Oct. 2013.

[67] M. S. Bartlett, "Statistical estimation of density functions," *Sankhyā, Indian J. Statist., Ser. A*, vol. 25, no. 3, pp. 245–254, 1963.

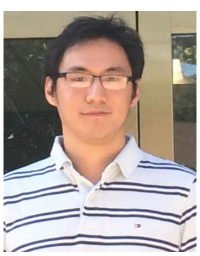
[68] M. S. Bartlett, "The spectral analysis of point processes," *J. Roy. Stat. Soc., Ser. B Methodol.*, vol. 25, no. 2, pp. 264–281, 1963.

[69] B. F. Jones and B. A. Weinberg, "Age dynamics in scientific creativity," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 47, pp. 18910–18914, Nov. 2011.

[70] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quart. Appl. Math.*, vol. 2, no. 2, pp. 164–168, 1944.

[71] M. Pelikan, *Bayesian Optimization Algorithm: From Single Level to Hierarchy*. Champaign, IL, USA: University of Illinois at Urbana-Champaign, 2002.

[72] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "BOA: The Bayesian optimization algorithm," in *Proc. Genetic Evol. Comput. Conf. (GECCO)*, vol. 1, 1999, pp. 525–532.
[73] J. Bergstra, D. Yamins, and D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 115–123.
[74] M. Liu et al., "Team formation and team performance: The balance between team freshness and repeat collaboration," 2022, *arXiv:2205.08756*.
[75] S. E. Jackson, V. K. Stone, and E. B. Alvarez, "Socialization amidst diversity: The impact of demographics on work team oldtimers and newcomers," *Res. Organizational Behav.*, vol. 15, pp. 45–109, Jan. 1992.
[76] G. Chen, "Newcomer adaptation in teams: Multilevel antecedents and outcomes," *Acad. Manage. J.*, vol. 48, no. 1, pp. 101–116, Feb. 2005.
[77] G. S. van der Vegt, S. Bunderson, and B. Kuipers, "Why turnover matters in self-managing work teams: Learning, social integration, and task flexibility," *J. Manage.*, vol. 36, no. 5, pp. 1168–1191, Sep. 2010.
[78] P. Azoulay, J. S. G. Zivin, and J. Wang, "Superstar extinction," *Quart. J. Econ.*, vol. 125, no. 2, pp. 549–589, May 2010.
[79] Q. Ke, E. Ferrara, F. Radicchi, and A. Flammini, "Defining and identifying sleeping beauties in science," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 24, pp. 7426–7431, Jun. 2015.
[80] J. Wang, R. Veugelers, and P. Stephan, "Bias against novelty in science: A cautionary tale for users of bibliometric indicators," *Res. Policy*, vol. 46, no. 8, pp. 1416–1436, Oct. 2017.
[81] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. suppl_1, pp. 5200–5205, Apr. 2004.
[82] D. Henriksen, "The rise in co-authorship in the social sciences (1980–2013)," *Scientometrics*, vol. 107, no. 2, pp. 455–476, May 2016.
[83] H. Peng, Q. Ke, C. Budak, D. M. Romero, and Y.-Y. Ahn, "Neural embeddings of scholarly periodicals reveal complex disciplinary organizations," *Sci. Adv.*, vol. 7, no. 17, Apr. 2021, Art. no. eabb9004.
[84] J. Angrist and G. Imbens, "Identification and estimation of local average treatment effects," NBER, Tech. Work. Paper 118, 1995.
[85] D. L. Thistlethwaite and D. T. Campbell, "Regression-discontinuity analysis: An alternative to the Ex post facto experiment," *J. Educ. Psychol.*, vol. 51, no. 6, p. 309, 1960.
[86] C. Hsiao, *Analysis of Panel Data*. Cambridge, U.K.: Cambridge Univ. Press, 2022.

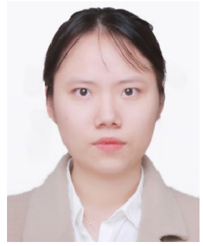


Chenbo Fu (Member, IEEE) received the B.S. degree in physics from the Zhejiang University of Technology, Hangzhou, China, in 2007, and the M.S. and Ph.D. degrees in physics from Zhejiang University, Hangzhou, in 2009 and 2013, respectively.

He was a Post-Doctoral Researcher with the College of Information Engineering, Zhejiang University of Technology, and a Visiting Scholar with the Department of Computer Science, University of California at Davis, Davis, CA, USA, in 2014. In 2017, he was a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, SAR, China. He is currently a Lecturer with the Institute of Cyberspace Security, Zhejiang University of Technology. His research interests include network-based algorithms' design, social networks, data mining, chaos synchronization, network dynamics, and machine learning.

Haogeng Luo received the B.S. degree in electrical engineering and automation from the Northern University of China, Taiyuan, China, in 2019. He is currently pursuing the M.S. degree in control theory and engineering with the College of Information and Engineering, Zhejiang University of Technology, Hangzhou, China.

His current research interests include science of science and causal inference, especially intersection of the two fields.



Xuejiao Liang received the B.S. degree in information management and information systems from the Northeast Forestry University, Harbin, China, in 2020. She is currently pursuing the M.S. degree in control theory and engineering with the College of Information and Engineering, Zhejiang University of Technology, Hangzhou, China.

Her current research interests include science of science and causal inference, especially intersection of the two fields.

Yong Min (Member, IEEE) received the B.S. degree in material engineering and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2003 and 2010, respectively.

He was a Post-Doctoral Researcher with the College of Life Sciences, Zhejiang University, in 2011. He was a Visiting Scholar with the Department of Environmental Science, University of Quebec (Montreal), Montreal, QC, Canada, in 2017. He is an Associated Professor with the Computational Communication Research Center, Beijing Normal University, Beijing, China. His research interests include network-based algorithms' design, social network analysis, data mining, computational communication, network dynamics, and machine learning.

Qi Xuan (Senior Member, IEEE) received the B.S. and Ph.D. degrees in control theory and engineering from Zhejiang University, Hangzhou, China, in 2003 and 2008, respectively.

He was a Post-Doctoral Researcher with the Department of Information Science and Electronic Engineering, Zhejiang University, from 2008 to 2010, respectively, and a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, in 2010 and 2017. From 2012 to 2014, he was a Post-Doctoral Fellow with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently a Professor with the Institute of Cyberspace Security, College of Information Engineering, Zhejiang University of Technology, Hangzhou, with the PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen, China, and also with the Utron Technology Company Ltd., (as a Hangzhou Qianjiang Distinguished Expert), Hangzhou. His current research interests include network science, graph data mining, cyberspace security, machine learning, and computer vision.

Guanrong Chen (Life Fellow, IEEE) received the M.Sc. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 1981, and the Ph.D. degree in applied mathematics from Texas A&M University, College Station, TX, USA, in 1987.

Since 2000, he has been a Chair Professor and the Founding Director of the Centre for Complexity and Complex Networks, City University of Hong Kong, Hong Kong. He is currently the Hong Kong Shun Hing Education and Charity Fund Chair Professor in Engineering. His research interests are in the fields of complex networks, nonlinear dynamics, and control systems.

Prof. Chen has been a member of the Academy of Europe since 2014 and a fellow of The World Academy of Sciences since 2015. He was Awarded the 2011 Euler Gold Medal from Russia and conferred Honorary Doctor Degrees by the Saint Petersburg State University, Russia, in 2011, and the University of Normandy, France, in 2014. He has been a Highly Cited Researcher in engineering continuously for ten years according to Clarivate Web of Science.