

# Identifying prize-winning scientists by a competition-aware ranking

Yuhao Zhou<sup>a</sup>, Ruijie Wang<sup>a</sup>, An Zeng<sup>b,\*</sup>, Yi-Cheng Zhang<sup>a</sup>

<sup>a</sup> Department of Physics, University of Fribourg, Fribourg 1700, Switzerland

<sup>b</sup> School of Systems Science, Beijing Normal University, Beijing, 100875, PR China

Evaluating scholars' achievements is an important problem in the science of science with applications in the evaluation of grant proposals and promotion applications. Since the number of scholars and the number of scholarly outputs grow exponentially with time, well-designed ranking metrics that have the potential to assist in these tasks are of prime importance. To rank scholars, it is important to put their achievements in perspective by comparing them with the achievements of other scholars active in the same period. We propose here a particular way of doing so: by computing the evaluated scholar's share on each year's citations which quantifies how the scholar fares in competition with the others. We assess the resulting ranking method using the American Physical Society citation data and four prestigious physics awards. Our results show that the new method significantly outperforms other ranking methods in identifying the prize laureates.

## 1. Introduction

Measuring the academic impact of scientists is an important research direction in the science of science (Zeng et al. (2017)). With the rapid development of science and technology, more and more scholars have devoted themselves to scientific research. Not only a large number of new academic papers appear every year, but also more junior scholars are emerging. So how to scientifically measure the academic influence of scientists and academic achievements are becoming more and more important. A fair measurement of academic influence can help scientific institutions select suitable scholars as faculties, promote outstanding academic staff, provide valuable references for academic awards, and evaluate grant proposals.

According to a recent review (Zeng et al., 2017), there are three main types of methods for evaluating a scholar's academic achievements. The first type is traditional static indicators. This kind of methods are the most widely used and studied, such as citation count (Garfield, 1970; Redner, 1998), H-index (Bornmann & Daniel, 2005; Light, Polley, & Börner, 2014), G-index (Egghe, 2006), and so on (Hirsch, 2005; Jin, Liang, Rousseau, & Egghe, 2007). These indicators are generally based on simple statistics, so they require less information and are more convenient to calculate. Meanwhile they also have obvious shortcomings. One example is the famous H-index, which combines the number of citations and the number of papers to measure the impact of a scholar. Although this metric is a combined measure of the quantity and quality of a scientist's publications, it is strongly dependent on the number of papers. For a scholar who publishes limited number of papers, despite that these papers have high impact, his/her H-index cannot be high. This means that although these traditional methods

\* Corresponding author.

E-mail address: [anzeng@bnu.edu.cn](mailto:anzeng@bnu.edu.cn) (A. Zeng).

are simple, there are limitations in them. The second type is network-aware methods (Bar-Ilan, 2008; Yan & Ding, 2009). In the science of science, the citation relationship can be seen as a complex network, which is called scientific network (Price, 1965). Through the node ranking algorithms, many scientific influence measurements for citation networks have been derived, such as, PageRank (Chen, Xie, Maslov, & Redner, 2007), author rank (Liu, Bollen, Nelson, & Van de Sompel, 2005), personalized PageRank (Ding, Yan, Frazho, & Caverlee, 2009), time-aware PageRank (Fiala, 2012) and diffusion-based science author rank algorithm (Radicchi, Fortunato, Markines, & Vespignani, 2009), etc. Third, some specially designed measurements for evaluating scholars, such as, dynamic method (Sinatra, Wang, Deville, Song, & Barabási, 2016), credit allocation based method (Van Hooydonk, 1997) and machine learning based methods (Abrishami & Aliakbary, 2019; Shen, Chen, Yang, & Wu, 2019).

All these methods have the same limitation, that is the time bias problem. In order to obtain a fair ranking of scholars, many research works have been done. Some scholars focus on solving time bias problems on a small scale. For example, the time-aware PageRank proposed by (Fiala, 2012), we mentioned above, solves a series of small scale time bias problems on citation relationship between scholars. Their results show that solving these time biases can help to improve the accuracy of prediction. Some other researches focus on large-scale time biases. As mentioned above, with the rapid development of science and technology, the academic environment is continuously changing, but the mainstream evaluation methods for scholars are almost static, so it's unfair when comparing scholars from different ages. For example, there is a classic time bias problem which is caused by the well-known preferential attachment mechanism (Bianconi & Barabási, 2001; Papadopoulos, Kitsak, Serrano, Boguná, & Krioukov, 2012). In the science of science, this problem exists at both the scholar level and the paper level (Price, 1965, 1976). This problem can be described as, compared to junior scholars, senior scholars have an advantage in time because they started their careers earlier. So senior scholars have longer time to accumulate attention. Due to their reputation, their work will be thought more valuable, so eventually they become famous all over the world. On the contrary, it is hard for junior scholars to get attentions. Many scholars have pointed out (Barabási & Albert, 1999; Perc, 2014) this problem. However, according to our analysis, besides the preferential attachment, there is another time bias issue in evaluating the long-term or lifetime achievements of scholars. It can be simply understood as when computing the academic achievements of a scholar, not only his/her own papers and citations should be considered, but also the peers during the same period should be taken into account. The more scholars in the same period, the more academic papers are published at this period, so it's easier to accumulate citations while more difficult to compete for an academic achievement award. Therefore, it is important to consider the academic environment when evaluating the achievements of scholars. We name this phenomenon competition. To explain and validate our theory, we use the well-known American Physical Society (APS) data set (Sinatra et al., 2016). To verify that our method can determine a scholar's academic achievement more reasonably, in this paper we use the Nobel Prize in Physics, the Wolf Prize, the Enrico Fermi Prize and the Max Planck Medal as benchmark data sets.

This paper will be composed of the following parts. In the second section, we will briefly describe the APS data set we used, define some mathematical symbols and introduce four metrics for measuring experimental results. In the third section, we first perform analysis on APS data set to introduce the competition problem in evaluating scholars. Then we explain in detail the existence and importance of the competition. Finally, we propose a new method based on the above analysis. In section 4, we illustrate the effectiveness of our method by analysing the results of experiments. In Section 5, we discuss the role of the parameter in our method. Section 6 is the conclusion of this paper.

## 2. Data, symbols and metrics

In this paper, we analyze the publication data from all journals of APS. The data contains 482,566 papers, ranging from year 1893 to year 2010. For the sake of author name disambiguation, we use the author name data set provided by Sinatra et al. which is obtained with a comprehensive disambiguation process in the APS data set (Sinatra et al., 2016). Eventually, a total number of 236,884 distinct authors are matched. We use the Nobel Prize in Physics as the main validation data set for this paper. The Nobel Prize data set includes all Nobel Prize laureates from year 1900 to year 2015, and at the same time we ensure that these Nobel Prize laureates have published at least one paper on the APS data set. Eventually, a total number of 158 distinct Nobel Prize laureates are matched. Meanwhile, we also provide the results of the Wolf Prize, the Enrico Fermi Prize, and the Max Planck Medal. It should be noted that in the following analysis, we mainly show the results on the Nobel Prize data set (The results for the other three awards are shown in a table at the end). There are three reasons:

- The Nobel Prize in Physics has the highest authority in the field of physics, which means that a laureate has a great influence on the progress of science and technology.
- The Nobel Prize in Physics has a long history, and the laureates are more widely distributed in time. This is more suitable for our research, that is, how to compare the achievement of scholars fairly in a long time span.
- The performances of the Nobel Prize laureates are uneven. Some laureates may have over 10,000 citations, while some of them only have less than 1,000, which also makes it difficult to determine the laureates of the Nobel Prize (Gingras & Wallace, 2009). Therefore, using the Nobel Prize in Physics as our benchmark set for evaluating academic achievements has a certain degree of novelty.

**Table 1**

A simple and concrete example for calculating metrics

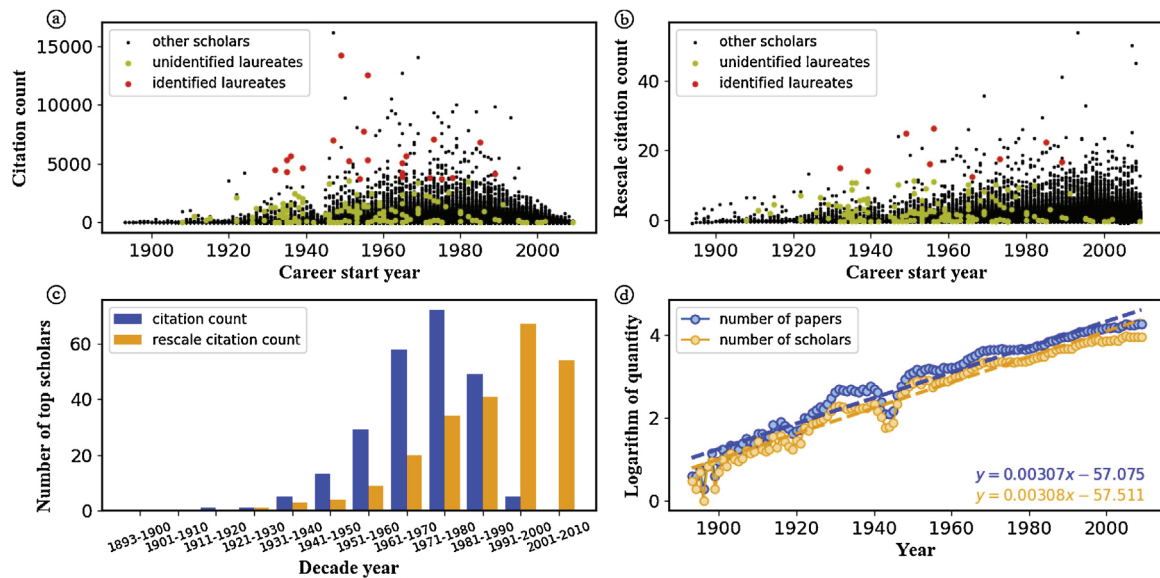
Name	Original score	Original ranking	New score	New ranking
<i>a</i>	10	1	2	6
<i>b</i>	9	2	6	2
<i>c</i>	8	3	4	4.5
<i>d</i>	7	4	7	1
<i>e</i>	6	5	5	3
<i>f</i>	5	6	4	4.5
<i>g</i>	4	7	1	7
Precision		0.33		0.67
AR		4		2.5
RIR			0.67	

We first give some definitions of important symbols here. All lowercase letters indicate the attributes of a paper, and all uppercase letters indicate the attributes of the an author. The time of paper  $i$  was published is  $t_i$ . The debut time of author  $i$  is  $T_i$ , which means the time when the scholar  $i$  published the first paper in the APS data set.  $N_i$  represents the author  $i$ 's collection of all papers in the APS data set, and we use  $|N_i|$  to indicate the number of all his papers. We use  $c_i$  to indicate the total number of citations for paper  $i$ , and use  $C_i$  to represent the total number of citations of author  $i$  (It should be noted that our citation relationship is limited to the APS data set we used), so we have  $C_i = \sum_{j \in N_i} c_j$ . In this paper we use the second subscript to indicate the component of the indicator in a certain year. For example,  $N_{it}$  represents the number of papers published by author  $i$  in year  $t$ , so we have  $N_i = \sum_t N_{it}$ . Similarly,  $C_{it}$  represents the number of citations the author  $i$  received in the year  $t$ , so we also have  $C_i = \sum_t C_{it}$ . In addition, we use  $R_t$  to indicate the total number of references for the year  $t$ , that is, the sum of the number of references of the papers published in year  $t$ .

In this paper, we aim to identify among all scholars who are eligible for the Nobel Prize and compare the results with the real Nobel Prize data set. In order to quantitatively compare the identification performance of different methods, we use the following four metrics:

- Precision.  $Precision = \frac{|q \cap p|}{|q|}$ . The precision is widely used in identification tasks (Lü & Zhou, 2011),  $q$  is the set of real laureates and  $p$  is the result predicted by the algorithm, where  $|q| = |p|$ . We call  $q \cap p$  as identified laureates and complementary set  $\complement_q p$  as unidentified laureates. This index is the most straightforward index to measure the identification result. However, this index cannot accurately measure algorithms when the size of test data set is small, such as the Nobel Prize data set we used in this paper, so we provide the second measurement.
- Average ranking (AR). In order to solve the low-resolution problem, we use the average ranking. This metric is also used in Radicchi et al. (2009) to measure the quality of their identification method. Given an evaluation index, after we have the rankings for all scholars according to their scores calculated by the index, we average rankings of all real laureates. It should be noted that when some scholars get the same scores, the ranking of these scholars will all be replaced by the mean value of their rankings. Unlike precision, AR considers the ranking of all laureates, so it is not affected by the size of the test data set. However, when we use it, we find that this metric also has a problem. When the average ranking result of an algorithm is particularly high, there may be two reasons. It may be because the algorithm has a better overall performance, or it may be because the ranking of a few data points increase sharply, resulting in an increase in the average ranking. In order to avoid this kind of misjudgment, we introduce another metric RIR.
- Ranking increase rate (RIR).  $RIR = \frac{1}{N} \sum_i \delta(i)$ . When comparing the ranking of the two ranking methods, beside the average ranking, we can use RIR to measure how many Nobel laureates in our case are improved in ranking.  $\delta$  is comparing function. Here we directly abbreviate to  $\delta(i)$ , representing whether the new method can improve the ranking for scholar  $i$ . If the ranking is higher than the original method,  $\delta(i)$  is equal to 1, otherwise it is 0. It can be seen that the RIR does not care about the optimization degree of the ranking, but focuses on the scope of optimization.
- Top-n%. Finally, in order to enrich the results, we use the Top-n% metric to further analyze the results. This metric was also mentioned and used in Radicchi et al. (2009).  $n$  represents the ranking threshold. In this paper, we use four levels of  $n$ : 0~0.1%, 0.1%~1%, 1%~10%, 10%~100%. Then in each interval we count how many real laureates are included. It is easy to know that the more the real laureates are assigned in the higher rankings interval, the better the identification result will be.

An example of calculating concrete values for both AR and RIR is given below. For example, here we have seven scholars, named 'a' to 'g'. Their original scores and original rankings are scores and rankings calculated by original method respectively. And their new score and new ranking are calculated by new method respectively. Detailed values are given in Table 1. We assume that 'b', 'd' and 'f' are real laureates. Using the original method for prediction, the top three scholars obtained are 'a', 'b' and 'c', so the precision is 0.33, while the prediction results of the new method are 'd', 'b' and 'e', and the precision is 0.67. Based on the original method, the average ranking (AR) is 4. After using the new method, the AR of the new method is improved to 2.5. Finally, both 'd' and 'f' are improved in their rankings, so the RIR is equal to 0.67.



**Fig. 1.** Statistical results on the APS data set. Subgraph (a) is the scatter plot of the career start year of researchers versus the total citations to them. (b) is the scatter plot of the career start year of researchers and their rescaled citation count. (c) Histogram of top-0.1% scholars ranked by rescaled citation count and citation count. Here we assign all top-0.1% scholars into several different decade periods instead of every single year. (d) The number of new authors (or new papers) in each year for the APS data. The y-axis is the logarithm of the quantity. We can see that the number of scholars and publications increases exponentially over time.

### 3. Model

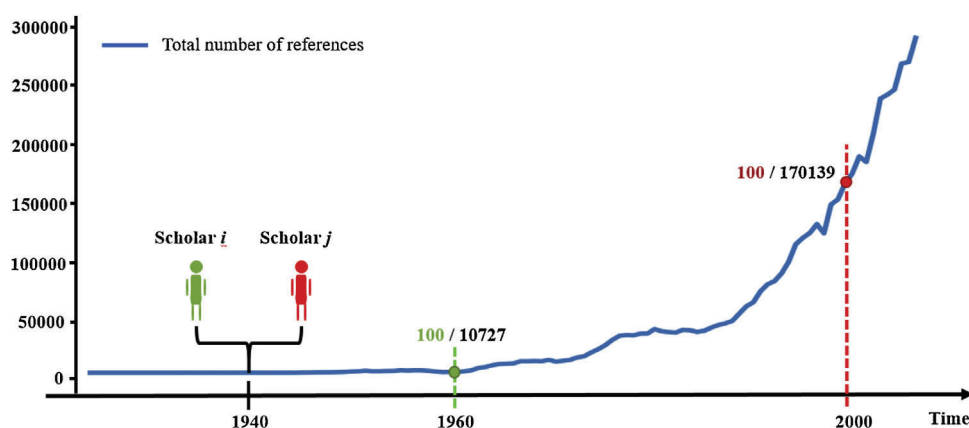
In this section, we will propose our new method for identifying the prize-winning scientists by considering competition in scientists ranking. We start with the basic method, citation count (CC), which can be computed by directly counting the total citations of all scholars in APS data set. As shown in Fig. 1(a). Each dot represents a scholar. We can see that the result distribution of CC can be divided into a rising phase and a decreasing phase (except the obvious gap caused by the World War II around 1945 year). These two phases are caused by competition and preferential attachment respectively.

As shown in Fig. 1(d), the academic population increases over time, which directly leads to higher citations of scholars, shown in Fig. 1(a). This makes the distribution of top scholars seriously uneven over time. We can see the blue histogram in Fig. 1(c), by 1980, the distribution of top-0.1% of scholars has a positive correlation with time. After the peak around 1980, scholars' citation count and the number of top scholars decrease sharply year by year. This trend is mainly caused by preferential attachment. Many methods have been proposed for correcting the bias of the preferential attachment problem, such as rescaled citation count (Newman, 2009), rescaled PageRank (Mariani, Medo, & Zhang, 2016), citation rank (Walker, Xie, Yan, & Maslov, 2007), etc. Taking rescaled PageRank (Mariani et al., 2016) as an example, this effect can be eliminated by only comparing a scholar with his peers, and the result can be seen in Fig. 1(b) and the orange histogram in Fig. 1(c). After using the rescaled method to eliminate the effect of preferential attachment, the downward trend has gone, leaving only a clear upward trend. It means that, although we have solved the preferential attachment problem, the effect of competition still exists. Scholars from different time periods are still not at the same level to be compared. Moreover, the distribution shape of three other popular indicators are same as CC. Results can be seen in Fig. 3. This means that those three indicators do not consider competition either.

Competition can be illustrated with a simple example, the number of scholars who debuted in year 1970 is 2,127, the average citations of scholars in that year is 111.1, and the standard deviation of citations is 295. According to the  $3\sigma$  principle, there are 48 outstanding scholars (that is, their citations are greater than the mean value plus 3 times of standard deviation). In 1910, only 27 scholars debuted in this year, and only one person can be thought an outstanding scholar with total 1,003 citations, even less than the minimum of the 48 scholars from 1970. Therefore, if the Nobel Prize uses citation as their criterion, all scholars from 1910 will have no chance to be nominated. In reality, for the outstanding scholars in 1970 it was 48 times more difficult to win the prize than for those in 1910, because of the competition between a large number of peers from 1970. This phenomenon was also observed by Gingras and Wallace (2009).

In summary, with the academic environment continuously changing, traditional static academic indicators are ineffective in identifying scholars' achievements (Fire & Guestrin, 2019), because they usually ignore the competition effect. Therefore, we need to design a new method for evaluating the overall achievement of scholars, which can tackle the competition problem.

We use the Fig. 2 to explain our idea. As shown in the figure, we assume that both scholar  $i$  and scholar  $j$  debuted in year 1940, and the total citations of them are 100 until 2010. The 100 citations of scholar  $i$  came from year 1960, and the citations



**Fig. 2.** The illustration of the competition effect when evaluating the achievement of two scholars. We assume there are two scholars  $i$  and  $j$ . Until 2010, they both have 100 citations, but  $i$  received these citations in 1960 and  $j$  received these citations in 2000. If we evaluate the two scholars from the perspective of competition, scholar  $i$  are more outstanding than scholar  $j$ , as it is more difficult to obtain 100 citations in 1960.

of scholar  $j$  came from year 2000. According to the APS data set, the total number of references from all papers published in year 1960 was 10,727, while the total number of references in year 2000 was 170,139. So getting 100 citations in 2000 is much easier than that in year 1960. In order to solve this difference, we propose the following formula.

$$C_i^* = \sum_t \frac{C_{it}}{(R_t)^\alpha} \quad (1)$$

Where  $C_i^*$  is our new index. In order to calculate  $C_i^*$ , we consider the author's citations in each single year as  $C_{it}$ . According to the above analysis, the achievements of a scholar in year  $t$  do not depend on the absolute citations received in this year, but on a relative value. In order to increase the generality of the method, we add an adjustable parameter  $\alpha$ . A larger  $\alpha$  indicates that our method gives more weight to competition. It is natural that when  $\alpha = 0$ , the Eqn (1) degenerates into citation count. Therefore, we can regard CC as a specific case of our method. We can also see that if we use traditional methods, such as CC, H-index (assuming that both scholars have only one paper in the same time) or rescaled method we mentioned above, it is impossible to distinguish scholar  $i$  and scholar  $j$ , because they all ignore the competition effect.

## 4. Results

In this section, we mainly present the experiment results on the Nobel Prize, because of the three advantages mentioned above. Meanwhile, to enrich the results, we also briefly show the results of three other prizes at the end of this section.

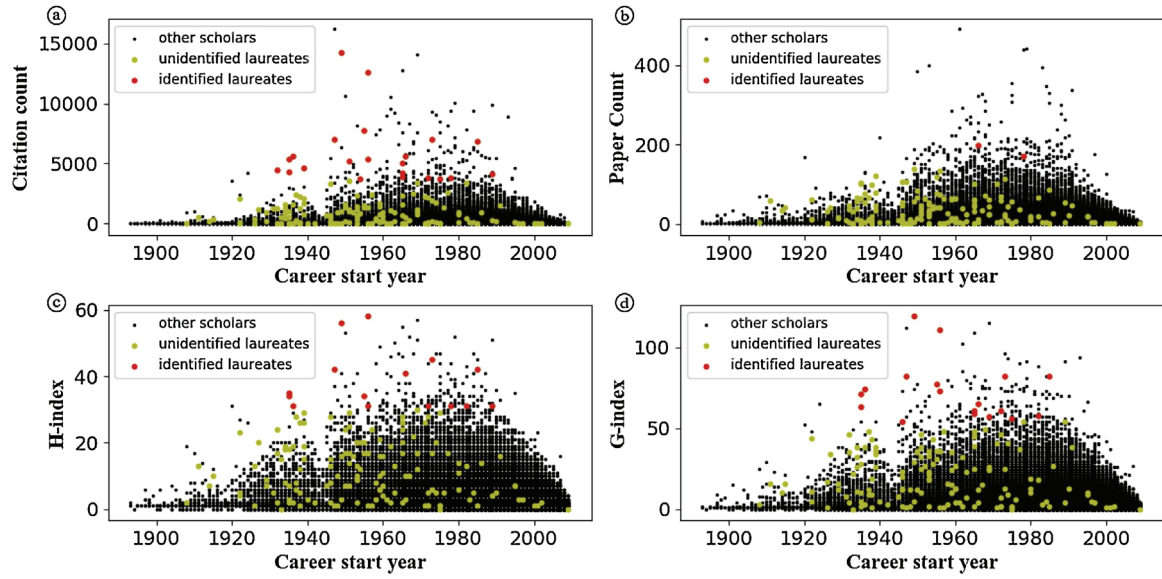
### 4.1. Methods

First, we introduce all methods we used for making a comparison. As mentioned above, a plethora of quantitative metrics exist and could be compared in principle, but our goal is to show that our methods can solve the competition effect problem effectively. Therefore, in this paper our focus is narrowed to traditional statistical methods. So we first briefly introduce four traditional famous metrics:

- Citation count (CC). CC is a basic academic statistical indicator, which can straightforwardly reflect the attention of scholars' scientific work.
- Paper count (PC). That is, up to a certain year, how many publications the scholar has. This indicator is also called the academic productivity, and it can measure scholars' academic activity and productivity, but ignore their publications quality.
- H-index (Hirsch, 2005). Currently, H-index is a mainstream evaluation method for scholars, which combines academic productivity and academic quality to evaluate a scholar comprehensively. H-index treats all papers of a scholar equally, which will ignore the importance of high quality papers, so it is not fair for evaluating the scientific performance for scholars with a few highly cited papers. To overcome it, G-index was proposed.
- G-index. Egghe (2006) proposed the G-index, which is defined as the largest number  $g$  of individual publications that together have at least  $g^2$  citations. So G-index can give more weight to papers with high citations than H-index.

Then we introduce our two methods:





**Fig. 3.** Scatter plot of four basic indicators versus the career start year of researchers (a:citation count, b:paper count, c:H-index, d:G-index). In each subgraph, each point represents a scholar. The red dots and the green dots represent the Nobel Prize laureates. The red dots represent the scholars that are correctly identified as the Nobel Prize laureates while the green dots represent the failed case. All black dots are scholars who have not won the Nobel Prize.

- Competition-aware citation count (CCC), see formula(1). As a variant of CC, the results of CCC are mainly compared with CC. Actually, the Eqn (1) can be generalized into a framework, that is, we can use any academic indicators to replace CC, then split it by year, and divide by the corresponding competition factor, namely the total references in one year. Therefore, in order to further illustrate the effectiveness of our method, we also apply it to the number of papers (which is called the academic productivity of scholars) as an additional example.
- Competition-aware paper count (CPC). It is natural to consider that the competition factor for the academic productivity of scholars is the total number of papers published in the one year. We thus propose a competition-aware paper count method, mathematically expressed as Eqn (2). We can see that when  $\alpha = 0$ ,  $P_i^*$  degenerates into a pure academic productivity of a scholar. Therefore, the Eqn (2) is an improved version of the scholar's academic productivity. If CPC works well, it can not only explain the existence of the competition effect and the effectiveness of our method, but also exhibits the robustness of our method. After reducing a lot of useful information (without information of the citations of papers compared with CC and PC), our method can still have high accuracy, which means any academic statistical indicators can be improved by introducing the competition concept. As a variant of PC, the CPC results are mainly compared with PC below.

$$P_i^* = \sum_t \frac{P_{it}}{(N_t)^\alpha} \quad (2)$$

#### 4.2. Analysis results

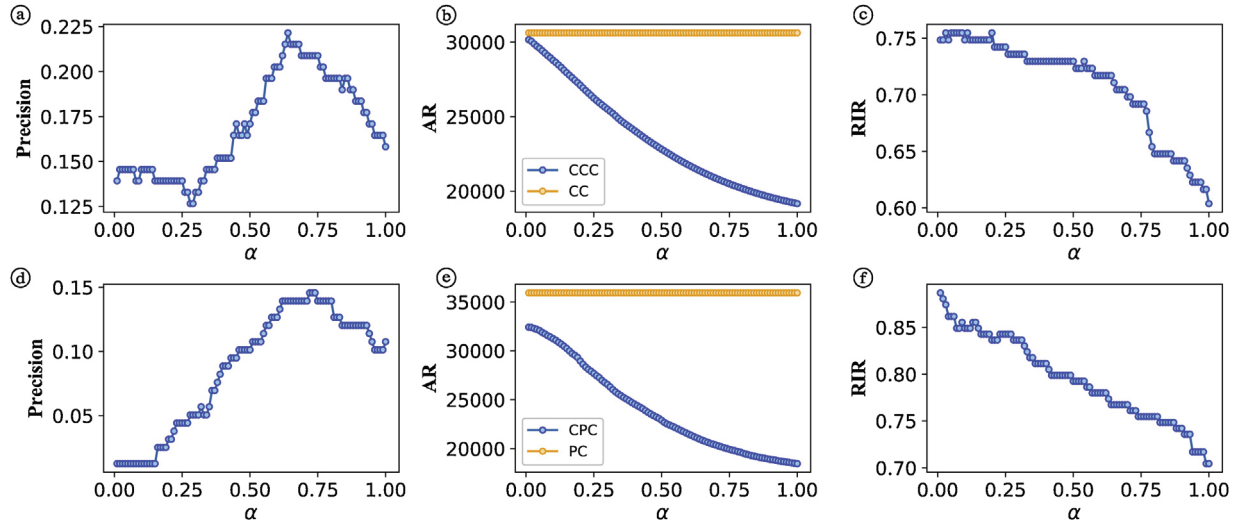
As shown in Fig. 3, we present the results of four classical methods (CC, PC, H-index and G-index). First, Fig. 3 gives us an intuitive feeling about the distribution of all scholars and the ranking positions of all real laureates. It can be clearly seen that although the Nobel Prize laureates are generally recognized as the scholars with the largest contributions, this does not mean that their indicators are always the best. Actually many Nobel Prize laureates are at very low positions. Of course, this may be caused by the limited data set and incomplete information, but more importantly, this means that determining Nobel Prize laureates is a very difficult task, and the precision will not be very high. Then, as we mentioned above, the distribution trend of the four indicators is basically the same, all of which show serious biases on time. This means that these static indicators can no longer evaluate academic achievements fairly, due to the fact that the academic environment changes continuously (Fire & Guestrin, 2019). The detailed simulation results of the four methods are shown in Table 2. From the perspective of precision and average ranking, we can see that G-index performs slightly better than H-index, and the CC method is the best among them. Unsurprisingly, the paper count is the worst index, because PC only considers the productivity but ignores the quality.

The results of two new methods CCC and CPC are shown in Fig. 4. It should be noted that since our new methods contains a tunable parameter  $\alpha$ , in our experiments we set  $\alpha$  from 0 to 1, and calculate every 0.01 step. The result of CCC is shown in the first row of Fig. 4. First, from the perspective of precision, compared with the original CC method (when  $\alpha = 0$ ), most of the  $\alpha$  values can improve the identification precision (except for a small range around 0.3), and the optimal precision

**Table 2**

Quantitative results for all awards. For the sake of fairness when comparing all awards, we adjust  $\alpha = 0.64$  for CCC method (Experimental results show that when  $\alpha = 0.64$ , CCC has the best prediction precision on the Nobel Prize data set. In order to compare performance without loss of generality, we will fix this parameter on the three other data sets).

Prize	CCC			CC		H-index		G-index		PC	
	Precision	AR	RIR	Precision	AR	Precision	AR	Precision	AR	Precision	AR
Nobel Prize	<b>0.222</b>	<b>21,370</b>	0.717	0.139	30,617	0.095	33,761	0.114	32,440	0.013	35,921
Enrico Fermi Prize	<b>0.120</b>	<b>28,145</b>	0.820	0.000	44,736	0.000	47,582	0.020	45,762	0.020	46,160
Max Planck Medal	<b>0.069</b>	<b>20,768</b>	0.724	0.034	37,737	0.017	48,652	0.017	46,287	0.017	53,448
Wolf Prize	<b>0.100</b>	<b>15,917</b>	0.560	0.060	18,108	0.060	23,062	0.060	24,505	0.020	29,953

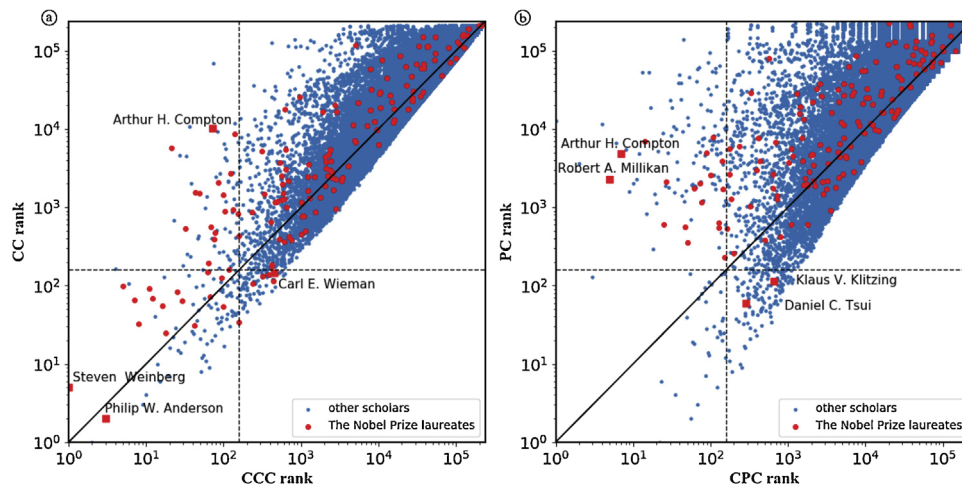


**Fig. 4.** Dependence of the results on the parameter  $\alpha$ . The subgraphs a,b and c are the results of our CCC method, and the subgraphs d,e and f are the results of the CPC method. The left, middle, and right sides represent precision, average rank, and ranking increase rate respectively. The best precision of CCC is 0.222, and the corresponding  $\alpha$  is 0.64. And the best precision for CPC is 0.146 with the  $\alpha$  equals to 0.72. As the  $\alpha$  increases, the average ranking of the two methods are monotonically increasing, while both the ranking increase rates show monotonous decreasing trends.

0.222 is reached when alpha is set to 0.64. As we can see, the average ranking shows a monotonous increasing pattern as  $\alpha$  increases. However, this does not mean that the greater  $\alpha$  is the better the result will be. As we can see from the RIR curve, it shows a monotonous decreasing trend. This means when we increase  $\alpha$ , the rise in average rankings is due to a small number of Nobel laureates moving up significantly, not overall optimization. This phenomenon also indicates that AR is not an effective ranking method for identifying prize-winning scientists. To sum up, our CCC method shows a good performance, which indicates the advantage of introducing the competition mechanism into the author ranking task.

As for CPC, it can also prove the robustness of our method. It is well known that there is a positive correlation between CC and PC (Bayer & Folger, 1966). Compared to CC, PC only accounts for the quantity of papers of scholars but not quality information. If our method still has good performance on PC, it indicates that our method is robust against low quality data. The results of CPC are shown in the second row of Fig. 4. Similar to the results of CCC, the precision of CPC increases first and then decreases. We can see that the RIR value stays above 70% all the time, which means that our method can effectively improve the rank of most Nobel Prize laureates. It should be noted that there is a gap at the left side of the two AR curves in Fig. 4(e). This is because ranking scholars by using PC may make many scholars indistinguishable. For example, there are 892 scholars owning 20 papers in our data set. However after using our new indicator, these scholars' scores will be completely different and the rankings of laureates become better. It is worth noting that there is a great improvement of precision in Fig. 4(d) in which the optimal value of the precision is 0.146 (when  $\alpha = 0.72$ ). This precision is 11 times higher than using the productivity to identify Nobel laureates (when  $\alpha = 0$ , precision=0.0127), and it is even better than the precision of CC (precision=0.139). This result shows that after considering the competition effect, if we only use productivity to identify, it can also achieve the similar effect as the citation count.

In order to compare CC and CCC from a more intuitive perspective, we use the ranking results of the two indicators to make a scatter plot, as shown in Fig. 5(a). If a scholar is above the black diagonal line, it means that his/her CCC ranking is better than CC ranking, and the farther the deviates from the line, the more his/her ranking is improved by CCC. As can be seen from Fig. 5, the majority of Nobel Prize laureates (red dots) are above the diagonal line, indicating that CCC has improved the ranking of most Nobel Prize laureates, which is the result of overall improvement rather than local improvement. Although some of the Nobel laureates (red dots) are below the black solid line, the deviations are much smaller than those for the dots above it. That is actually because, those scholars whose rankings are not promoted do not drop significantly, and they



**Fig. 5.** The comparison of ranking results of different methods. It should be note that we take logarithm for all rankings. (Left) The x-axis is the ranking of a scholar calculated by CCC ( $\alpha=0.64$ ), and the y-axis is the ranking of a scholar by using the CC. (Right) The x-axis is ranking of a scholar calculated by CPC ( $\alpha=0.72$ ), and the y-axis is the ranking of a scholar by using the PC. In these two figures, each dot represents a scholar. The red markers and blue dots indicate the Nobel Prize laureates and other scholars respectively. Black solid lines are  $y=x$ , indicating that the dot on it has the same rankings in two methods. The black dashed lines indicate the boundaries of identification, that is  $x=\ln 158$  and  $y=\ln 158$ , so the left side of the vertical dashed line indicates the identification result of our new indicator, and the lower side of the horizontal dashed line indicates the identification result of the traditional method.

can still retain their original rankings. Taking two scholars as examples, the first scholar ‘Arthur H. Compton’ and the second scholar ‘Carl E. Wieman’ who are marked in the figure. Among all the Nobel Prize laureates identified by using CCC, ‘Arthur H. Compton’ has the lowest CC ranking, his CCC ranking and CC ranking are 72 and 10,141 respectively. This is a significant improvement in the ranking, which lifts ‘Arthur H. Compton’ into top-1% group. Meanwhile, among all the Nobel Prize laureates identified by using CC, ‘Carl E. Wieman’ has the lowest CCC ranking, his CCC ranking and CC ranking are 453 and 145 respectively. Although there is a drop in his ranking, he still remains in the top-1% group.

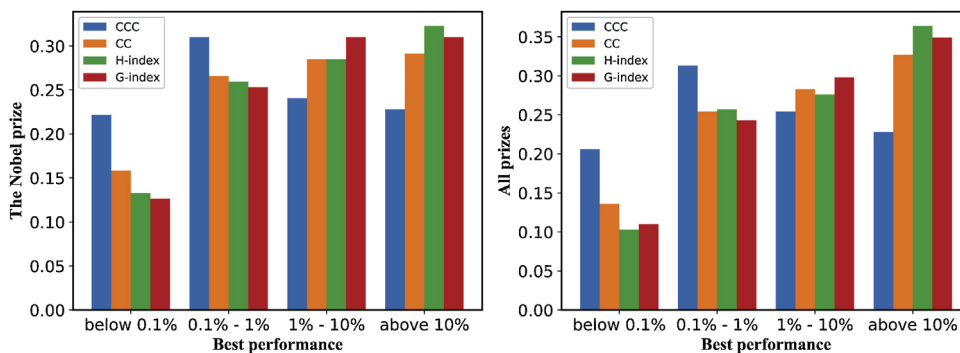
In addition, as we can see that the figure is divided into four regions by two dashed lines. The lower left area is the common identification area of two methods, that is, the scholars identified by the two methods at the same time. The lower right area and upper left area represent the additional identification area of CC and CCC respectively. We can see that CCC can retain the most of laureates identified by CC (only five red dots located at the lower right area). As shown in the figure, take ‘Steven Weinberg’ and ‘Philip W. Anderson’ as two examples, no matter which method is used, these two scholars are always ranked very high. At the same time, in the upper left corner, we can see CCC successfully identify other 18 Nobel Prize laureates. This shows that CCC is a successful extension method of CC. The results of CPC and PC are shown in Fig. 5. The same as the left one, most of the Nobel Prize laureates (red dots) are located above the black solid line, and their deviations are much larger than the red dots below it (take the four scholars marked in the figure as examples). There are no Nobel Prize laureates in the common identification area of PC and CPC, mainly because of the poor performance of the original PC method. In other words, the new method does not inherit any useful results from PC.

Finally, we use the Top-n% metric to compare the results of various methods. In order to enrich the results, we also use three other famous physics awards, the Wolf Prize (50 laureates), the Enrico Fermi prize (50 laureates) and the Max Planck Medal (58 laureates). Compared with the Nobel Prize, these three prizes have a smaller number of laureates, and the distribution of the winning time is not as wide as the Nobel Prize. Here we do not plot these three prizes separately, but put all the awards together (specific results can be found in Table 2). Finally we have 272 laureates from four awards, some of whom may won multiple awards. In the top-n% metric, we expect that the better the algorithm performs, the more awarded authors will be found in the top rank percentage. As we can see from the Fig. 6, our competition-aware method always performs better. The two categories ‘below 0.1%’ and ‘0.1%-1%’ always contain more laureates compared to other methods. Meanwhile ‘1%-10%’ and ‘above 10%’ always have fewer laureates. This means that our method would identify most of the award winning authors by placing those at a relatively high ranks. Here we also use a table to show more details of performance on the four awards, see Table 2. We adjust  $\alpha$  to 0.64. From the results, we can see that our method performs well on both four data sets, and parameter  $\alpha$  seems to be universal (Although in a specific award, 0.64 may not be the best choice, it still works better).

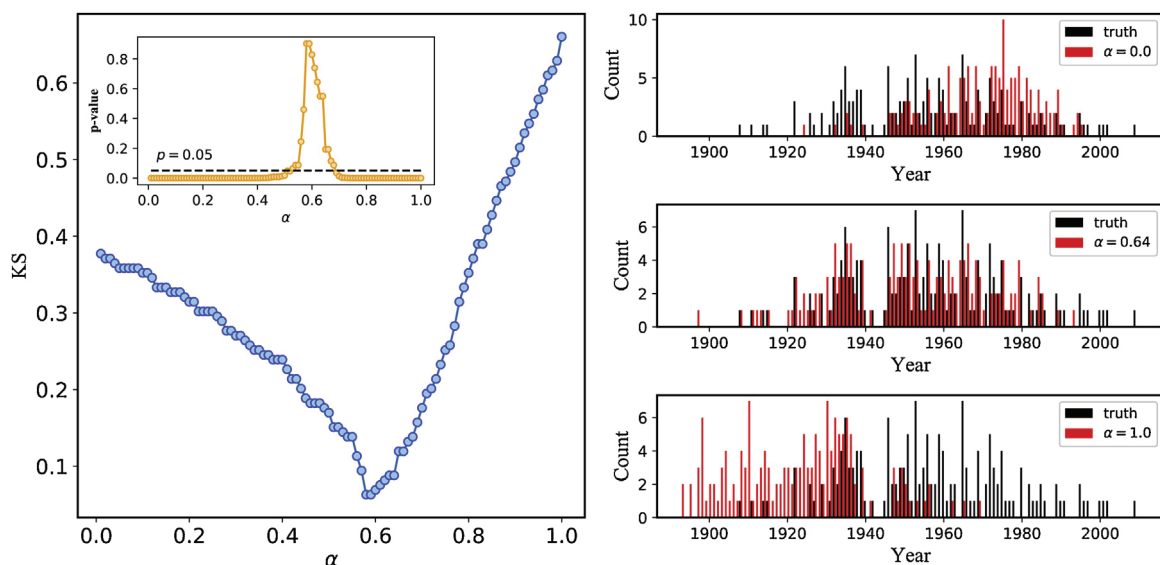
#### 4.3. The role of $\alpha$

When we introduce the Eqn (1) proposed in section 3, we regard  $\alpha$  as the parameter that is used for adjusting the degree of attention of the competition. A larger  $\alpha$  indicates that stronger effect of competition is considered, which is the most intuitive interpretation of  $\alpha$ . At the same time, according to the results in section 4.2, we can find that although most  $\alpha$  can





**Fig. 6.** The top- $n\%$  measurement on 4 awards in physics (the Nobel Prize, the Max Planck Medal, the Wolf Prize, and the Enrico Fermi Prize). We fix  $\alpha = 0.64$  for CCC. The performance of a scholar depends on the percentile of the scholar's ranking. The lower this percentage is, the better the performance of the considered scientist is. The height of the bar indicates the proportion of laureates. The left figure only considers the Nobel Prize, while the right figure takes into account all four prizes. We can find no matter the Nobel Prize or all prizes we considered, our competition-aware method performs better than others.



**Fig. 7.** We use Kolmogorov-Smirnov test (see main text for explanation) to explain the role of parameter  $\alpha$ . The figure on the left side is the curve of KS scores with parameter  $\alpha$ , and the small inset graph is the curve of  $p$ -values. The three subgraphs on the right side are the comparisons between the distribution of the identification results under three different parameter values ( $\alpha = 0, 0.64, 1$ ) and the distribution of the real Nobel Prize laureates. Black bar represents the distribution of the Nobel Prize laureates and red bar represents the distribution of our results.

improve the identification effect, there is always a best choice of  $\alpha$ . Why is there always a best result after adjusting  $\alpha$  to a specific value, and how does  $\alpha$  affect the identification results? In this section, we aim to illustrate the effect of  $\alpha$ .

$\alpha$  is actually helping our method to find the best output distribution, which is the closest to the distribution of real Nobel Prize laureates. To better illustrate this point, we use Kolmogorov-Smirnov test (KS test) (Hodges, 1958). This test is a nonparametric test of the equality of continuous (or discontinuous), one-dimensional probability distributions that can be used to compare two samples. The larger the KS test score, the larger the distance between the two distributions; the smaller the score, the more similar the two distributions are. Meanwhile, the KS score also gives the corresponding  $p$ -value for analyzing the significance of result. The procedure of experiment is the same as in section 4.2, where  $\alpha$  is from 0 to 1, and the distance between the distribution of the identification set and the distribution of the real Nobel Prize set is calculated in every 0.01.

The results are shown in Fig. 7. We can clearly see that under the change of  $\alpha$ , the KS score decreases first and then increases sharply. The minimum value of the distance appears around  $\alpha = 0.6$ , and a smaller score means the distribution of our identification can match the distribution of the real Nobel Prize laureates better. In the three histograms on the right side of Fig. 7, we show the distributions of our results under three special cases of  $\alpha$  and the distribution of the real Nobel Prize laureates. In the first histogram, we set  $\alpha = 0$ , which means that the red bar shows the distribution of the CC result. It can be seen that identification tends to concentrate around 1970. This is easy to understand, since we have analysed in Fig. 3 in section 4 that the distributions of the four traditional indicators are basically the same, with peaks around 1970. Therefore, the distributions of top scholars calculated by traditional indicators always appear near 1970. When  $\alpha = 0.64$ , the identification

**Table 3**  
Logistic regression results

Cross validation		CV-1	CV-2	CV-3	CV-4	Average
A	AR	8,242	9,829	6,405	6,105	7,646
	F1-score	0.016	0.015	0.016	0.016	0.016
	Precision	0.225	0.200	0.100	0.051	0.144
B	AR	7,012	7,651	5,840	5,061	6,391
	RIR	0.525	0.650	0.625	0.550	0.588
	F1-score	0.026	0.022	0.026	0.021	0.024
C	Precision	0.250	0.250	0.150	0.077	0.182
	AR	<b>5,840</b>	<b>5,356</b>	<b>4,305</b>	4,580	<b>5,020</b>
	RIR	<b>0.725</b>	<b>0.825</b>	<b>0.725</b>	0.667	<b>0.735</b>
CCC	F1-score	<b>0.030</b>	<b>0.026</b>	<b>0.031</b>	<b>0.030</b>	<b>0.029</b>
	Precision	<b>0.275</b>	<b>0.300</b>	0.175	<b>0.103</b>	<b>0.213</b>
	AR	6,542	6,528	5,013	<b>3,232</b>	5,329
CCC	RIR	0.625	<b>0.825</b>	0.700	<b>0.692</b>	0.711
	Precision	0.250	<b>0.300</b>	<b>0.200</b>	0.053	0.201

precision reaches the optimal value (0.222). As can be seen from the second histogram, the matching results of the red bar and the black bar are significantly better than the first case and the distribution of red bar becomes wider. It is particularly noteworthy that between 1900 and 1940, CC predicts that the number of top scholars in this period is almost zero, and this situation is improved very well after increasing  $\alpha$  to 0.64. In the third histogram we show the results when  $\alpha = 1.0$ . It can be seen that the algorithm pays too much attention to the competition, and the distribution of the prediction mainly concentrates between 1900 and 1940. It can also be seen from the curve of KS score that when the competition plays too much role in our method while the citations are ignored, the KS score increases rapidly to 0.67.

In general, through the above analysis, we believe that the parameter  $\alpha$  is actually used to adjust the distribution of the identification result. An optimal  $\alpha$  makes sure that there is no obvious bias on time when we make identification, and the output distribution is more close to the distribution of real laureates.

#### 4.4. Further exploration with machine learning

At the end of the paper, we want to illustrate that our competition-aware method can also be combined with machine learning methods, which deserves more exploration. We take the logistic regression (Owen, 2007) as an example. We consider the task of predicting Nobel Prize laureates as a special binary classification task. We use label 1 to indicate the Nobel Prize winner, and label 0 to indicate scholar who does not win the Nobel Prize. This prediction task is very special because its training data set is extremely imbalanced. The size of the positive samples in training data set is only 158 but the size of the negative samples is about 230,000. So during the training phase, we use weighted methods to give greater weight to the positive samples to solve this problem.

To illustrate the importance of the competition mechanism, we give three training strategies for machine learning experiments.

- (A) Citation count. Take the citation count as the unique feature of sample to make prediction.
- (B) Take citations in each single year as a single feature. So in this case we use vectors to represent different samples.
- (C) Based on case B, we use the concept of competition to normalize all feature vectors. See Eqn (1) for details.

In addition, in order to compare prediction results more comprehensively, we use four fold cross validations and four metrics AR, RIR, Precision, and F1-score to measure the performance of algorithm.

It can be seen from the Table 3 that strategy C achieves the best performance, and strategy A has the worst result. Compared with strategy A, the training set in strategy B is upgraded such that each sample gets extra information about time. This extra information is very helpful to improve the prediction precision. Compared to strategy B, we can see that in strategy C the prediction precision of the machine learning algorithm has been further improved after the competition mechanism is performed. In fact, for strategy C, the role of the competition mechanism is used as a data normalization process. This procedure is conducive to improving the performance of machine learning methods, which indicates that considering the competition effect is also helpful for designing training sets for machine learning algorithms.

At the end of the table, we also show the prediction results of the simple competition-aware citation count (CCC). It can be seen from the results that the performance of CCC is significantly better than that of strategy B, which means that our competition mechanism is useful itself. Strategy C has the highest accuracy, which also means that machine learning algorithms are superior to traditional methods in ranking or prediction tasks. In summary, all the results show that the competition is a crucial factor for identifying prize winning scientists.

## 5. Conclusion

In this paper, we mainly discuss how to measure the academic achievements for scholars more reasonably. We introduce the competition mechanism through analysis on the American Physical Society (APS) data set. With the rapid development of modern science and technology, the number of new scholars and new academic papers are increasing exponentially. This means that scientists are facing more and more competitors in academia. Therefore, junior scholars may find that it will be more difficult for them to obtain those lifetime achievement awards over time. Meanwhile, this also means that traditional academic indicators may have lost their meaning when evaluating scholars from different periods (Fire & Guestrin, 2019). Compared with senior scholars, junior scholars are in a larger and more active academic environment, it is easier for them to obtain more citations and attentions, but it does not mean that their performances are better. Therefore, if we do not take the competition effect into account, it is difficult for us to evaluate scholars' achievement fairly.

Then we use a series of experiments to prove our idea. From the results, we learn that after taking the competition factor into account, the predictive power of our new method shows a great improvement. Take citation count as an example, the prediction precision increases from 0.139 to 0.222. The average ranking of all real laureates increases from 30,617 to 21,370. Moreover, according to the indicator RIR, 71.7% of the laureates' rankings are improved by using our new method. Our new method is the optimal method among the traditional indicators we have listed for comparison. At the same time, we also made a further step by using the paper count to prove that our method has enough robustness and generality. The results show that when the number of papers is the only known information to determine the Nobel Prize laureates, our new method still has a significant improvement. The precision 0.146 is reached when competition is taken into account. The average ranking of all real laureates rises from 35,921 to 18,462. Moreover, according to the indicator RIR, 76.1% of the real laureates' rankings are improved. In summary, the aim of our method is not to replace other ranking techniques, optimized and almost perfected in the course of many years. We want to illustrate the existence of the competition factor and solve it by a simple way. Our research has important practical value, such as providing valuable references for academic awards, selecting suitable scholars as faculties, and promoting outstanding academic staff.

There are also numerous meaningful extensions for the future research. First, one can try to combine our idea with other methods, such as the diffusion based models discussed in introduction. This kind of methods can use more structure information of citation network, which will benefit the results. Similar to competition-aware citation count, we could adjust the impact of each paper according to the competition mechanism. Taking Pagerank as an example, we can give each out edge of a paper a weight based on its publication year, and this weight is determined by the total number of references or the number of papers in that year. Second, it is well known that there are many branches in the physics today. The impact of competition within these branches is different from each other. Therefore, when comparing the scholars from different academic domains, it will be fairer, if we perform analysis to distinguish different domains. For example, as mentioned by Zeng et al. (2019), one can use the community detection (Fortunato & Hric, 2016) to detect different research branches on citation network. The size of those communities can represent the degree of competition. So besides the competition between different periods, we can also calculate the competition inside different academic directions. These two factors together decide the weight of edges in the citation network. In addition to improving the prediction, one can also analyze the strength of competition between different fields according to the fitting results.

## CRedit authorship contribution statement

**Yuhao Zhou:** Software, Validation, Writing - original draft, Writing - review & editing, Formal analysis. **Ruijie Wang:** Writing - original draft, Writing - review & editing. **An Zeng:** Conceptualization, Methodology, Writing - review & editing, Formal analysis, Data curation. **Yi-Cheng Zhang:** Conceptualization, Supervision.

## Acknowledgment

The authors would like to thank Matúš Medo for helpful discussion. This work was supported by the Swiss National Science Foundation (Grant No. 200020-156188).

## References

- Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, 13, 485–499.
- Bar-Ilan, J. (2008). *Informetrics at the beginning of the 21st century: a review*. *Journal of informetrics*, 2, 1–52.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286, 509–512.
- Bayer, A. E., & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of education*, 381–390.
- Bianconi, G., & Barabási, A. L. (2001). Bose-einstein condensation in complex networks. *Physical review letters*, 86, 5632.
- Bornmann, L., & Daniel, H. D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65, 391–392.
- Chen, P., Xie, H., Maslov, S., Redner, S. 2007. Finding scientific gems with google's pagerank algorithm.
- Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60, 2229–2243.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69, 131–152.
- Fiala, D. (2012). Time-aware pagerank for bibliographic networks. *Journal of Informetrics*, 6, 370–388.
- Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: observing goodhart's law in action. *GigaScience*, 8, giz053.
- Fortunato, S., & Hric, D. (2016). Community detection in networks. *A user guide*. *Physics reports*, 659, 1–44.

- Garfield, E., et al. (1970). Citation indexing for studying science. *Nature*, 227, 669–671.
- Gingras, Y., & Wallace, M. (2009). Why it has become more difficult to predict nobel prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901–2007). *Scientometrics*, 82, 401–412.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102, 16569–16572.
- Hodges, J. 1958. The significance probability of the smirnov two-sample test. *Arkiv för Matematik* 3, 469–486.
- Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The r-and ar-indices: Complementing the h-index. *Chinese science bulletin*, 52, 855–863.
- Light, R. P., Polley, D. E., & Börner, K. (2014). Open data and open code for big science of science studies. *Scientometrics*, 101, 1535–1551.
- Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H. 2005. Co-authorship networks in the digital library research community. *Information processing & management* 41, 1462–1480.
- Lü, L., Zhou, T. 2011. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications* 390, 1150–1170.
- Mariani, M. S., Medo, M., & Zhang, Y. C. (2016). Identification of milestone papers through time-balanced network centrality. *Journal of Informetrics*, 10, 1207–1223.
- Newman, M. E. (2009). The first-mover advantage in scientific publication. *EPL (Europhysics Letters)*, 86, 68001.
- Owen, A. B. (2007). Infinitely imbalanced logistic regression. *Journal of Machine Learning Research*, 8, 761–773.
- Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguná, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, 489, 537.
- Perc, M. (2014). The matthew effect in empirical data. *Journal of The Royal Society Interface*, 11, 20140378.
- Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27, 292–306.
- Price, D. J. D. S. (1965). Networks of scientific papers. *Science*, 150, 510–515.
- Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80, 056103.
- Redner, S. (1998). How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4, 131–134.
- Shen, Z., Chen, F., Yang, L., & Wu, J. (2019). Node2vec representation for clustering journals and as a possible measure of diversity. *Journal of Data and Information Science*, 4, 79–92.
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354, aaf5239.
- Van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, 48, 944–945.
- Walker, D., Xie, H., Yan, K. K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, 2007, P06010.
- Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, 60, 2107–2118.
- Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature communications*, 10, 3439.
- Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, 714, 1–73.