



# A nonlinear collective credit allocation in scientific publications

Fenghua Wang<sup>1</sup> · Ying Fan<sup>1</sup> · An Zeng<sup>1</sup> · Zengru Di<sup>1</sup>

Received: 15 November 2018 / Published online: 25 April 2019  
© Akadémiai Kiadó, Budapest, Hungary 2019

## Abstract

Collaboration among researchers plays an important role in scientific discoveries, especially in multidisciplinary research. How to allocate credit reasonably to coauthors of a paper is a long-standing problem in the science of sciences. The collective credit allocation method (CCA method) proposed by Shen, H. W. and Barabási, A. L. provides a novel view to solve this problem, which measures the coauthors' contribution to a paper based on the citation process by the scientific community. Nevertheless, the existing collective allocation method assigns equal weights to citing papers, which is sensitive to the malicious manipulation. In this paper, we propose a nonlinear collective credit allocation method (NCCA method) that assigns different strength to citing papers according to papers' scientific impact when measuring papers' similarity. Compared to the CCA method, we find that the NCCA method assigns more credits to Nobel laureates in the Nobel-winning papers. Moreover, the NCCA method is robust against random perturbations and the malicious manipulation in both Nobel-prize papers and ordinary papers. Furthermore, the collective credit allocation method can also modify  $h$  index.

**Keywords** Citation network · Credit allocation ·  $h$  index

## Introduction

Collaboration has prompted scientific development and boosted scientific discoveries (Newman 2004; Wuchty et al. 2007; Dong et al. 2017), especially in multidisciplinary research which integrates knowledge from different scientific fields (Lawrence 2007). With the number of coauthors of each paper increasing, it is urgent for scientific community to

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11192-019-03107-x>) contains supplementary material, which is available to authorized users.

---

✉ Ying Fan  
yfan@bnu.edu.cn

✉ An Zeng  
anzeng@bnu.edu.cn

<sup>1</sup> School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China

evolve science's credit system, which is significant for hiring, funding and promotion decisions (Kennedy 2003; Zeng et al. 2017; Allen et al. 2014). For single-author papers, the credit allocation is simple. The sole author obtains the whole credit. However, the situation is more complicate for multi-authors papers because of the difficulty to quantitatively account for the discrepancy of contribution among coauthors (Sekercioglu 2008; Greene 2007; Kaur et al. 2013). The situation can be even worse when coauthors are from different fields applying different credit allocation criteria in multidisciplinary research (Lehmann et al. 2006). It is necessary for scientific community to develop a deeper understanding of the mechanism of credit allocation which could help to improve this situation and provide a perspective to measure each researcher's scientific impact (Radicchi et al. 2009; Stallings et al. 2013; Sinatra et al. 2016).

Nowadays, there are four main categories of scientific credit allocation methods in scientometrics. The first is to regard each author as a sole author and assign each author full credit. Naturally, this way leads to inflated scientific impact for multi-authors papers (Garfield 1972; Hirsch 2005). The second is to make an equal fraction of credit to co-authors of a paper (Van Hooydonk 1997; Hirsch 2007). This method neglects the fact that not all coauthors contribute equally to a paper and dilutes the credit share of the main contributor. The third is to consider the order or the role of coauthors in a paper. Based on this idea, there are four different ways to assign credit: geometric (Egghe et al. 2000), arithmetic (Trueba and Guerrero 2004), harmonic (Hagen 2008) and network-based (Kim and Diesner 2014). Different credit allocation methods are used based on different disciplines (Tschardt et al. 2007). Methods depending on author list are adopted by some fields but are not applied to others. For example, in most fields of physics, the first author or the corresponding authors get the most of credit share; while in mathematics and experimental particle physics fields, the method based on author order does not work since author names are listed alphabetically. Considering this complex situation, there is a trend to allocate authors' credit share by specified contribution declaration required by journals (Foulkes and Neylon 1996; Greene 1999). The fourth one is to treat credit allocation as a collective process (Radicchi et al. 2009; Shen and Barabási 2014; Bao and Zhai 2017). The existing collective credit allocation method (CCA method) proposed by Shen, H. W., & Barabási, A. L. allocates coauthors' credit based on the citation process by the scientific community. This model didn't consider the aging effect with the evolving system, which has been improve by Bao and Zhai (2017). The CCA method depends much on citing papers. However, the CCA method does not take citing papers' quality into consideration. They ignore the fact that papers have different scientific impact. Thus the CCA method can be easily manipulated by adding low impact citing papers. The CCA method should be improved this limitation to enhance its robustness.

The collective credit allocation method can modify  $h$  index, since one of disadvantages of  $h$  index when ranking scientists is that multiple co-authorship is not taken into account (Batista et al. 2006; Bornmann and Daniel 2007; Burrell 2007; Hirsch 2007).  $h$  index is proposed by Hirsch in 2005 to evaluate the scientific impact of a scientist, which is defined as the highest number of publications of a scientist that have been cited  $h$  or more times (Hirsch 2005). In order to address the problem of not considering multiple co-authorship, a series of works modifying  $h$  index have been done. The works include two aspects, the first is considering the number of coauthors. Schreiber, Egghe and Galam defined  $h_m$ , fractional  $h$  and  $gh$  based on fractionally credit allocation in multi authored papers (Schreiber 2008a, b, 2009; Egghe 2008; Galam 2011). The second is considering different roles of coauthors (Tschardt et al. 2007; Liu and Fang 2012; Ancheyta 2015; Ausloos 2015; Crispo 2015; Vavrycuk 2018; Hirsch 2019). The CCA method and the NCCA method have put forward

new ways to allocate credit in multi-authored papers, which provides a perspective to define  $h$  index.

In this article, we propose a nonlinear credit allocation method (NCCA method), considering citing papers' scientific impact rather than recognizing citing papers as the same weight, based on the state-of-the-art collective credit allocation method. Firstly, we get some basic statistics of the NCCA method. Then we further studying the robustness of the NCCA method, find that it is tolerant of random perturbations as well as the malicious manipulation in both Nobel-prize papers and ordinary papers. Finally, the CCA method and the NCCA method are applied to modify  $h$  index.

## Data

In this paper, the database we use contains all papers published in the journals of American Physical Society (APS), which includes Physical Review series and Reviews of Modern Physics. There are over 450000 papers and over 236000 authors, ranging from year 1893 to year 2010 in the database.

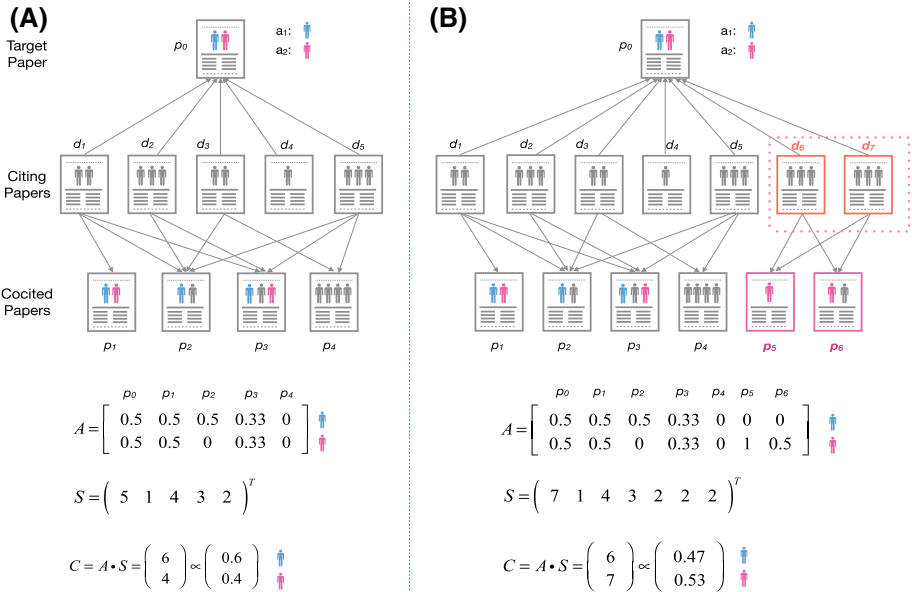
Lacking of unique author identifier in APS database, a name disambiguation process is needed to assign each paper to its true authors. In this paper, we make use of the APS database with author names disambiguated by Sinatra et al. (2016). In addition, we gather Nobel authors and papers from two different papers (Shen and Barabási 2014; Bao and Zhai 2017). As this work studies the credit allocation for multi-author papers, we delete Nobel-prize papers with only one author or more than one laureates.

## Method

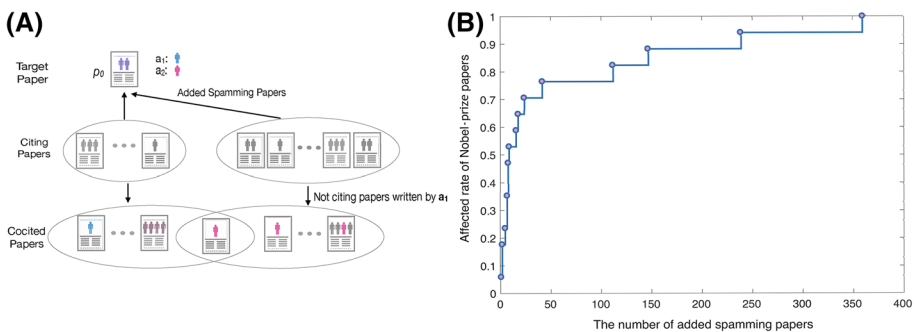
### The collective allocation method (CCA method)

The existing collective credit allocation method (CCA method) can capture the coauthors' contribution to a paper based on a credit allocation matrix and the co-citation strength of all identified co-cited papers (Shen and Barabási 2014). One of the basic assumptions of their method is that papers citing a target paper contribute equally to each co-cited paper, which makes the method sensitive to the malicious manipulation. For example, there is a paper written by author1 and author2. By the CCA method, we find that the credit of author1 is higher than that of author2 at first (Fig 1a). The credit rank of author1 and author2 can be manipulated by adding only two papers ( $d_6$  and  $d_7$ ) (Fig 1b). It is not difficult to add two papers since these two papers are not required as high quality papers. They only need to be published and cite the target paper  $p_0$  and other papers written by author2 not by author1 in the meantime. Additionally, authors of these papers can be anybody.

We try to test the sensitivity of the CCA method in Nobel papers dataset. First we find out author with the second highest credit share of a given Nobel-prize paper by the CCA method, then we add some papers which cite not only the given Nobel-prize paper but also other three papers in dataset written by that author but not by the author with the highest credit share. We count the number of added papers needed to make the credit share of the second highest author exceed the highest one (Fig. 2a). The result (Fig. 2b) shows that for 53% of 17 Nobel-prize papers, the original second highest credit share author only needs less than 10 added papers to make her or his credit share higher than the credit share of the original highest one.



**Fig. 1** The collective credit allocation method (CCA method). **a** A case of the CCA method. The credit share of author  $a_1$  is 0.6 which is higher than that of author  $a_2$ . **b** A manipulation case of the CCA method. Paper  $d_6$  and papers  $d_7$  are added manipulatively. Paper  $p_5$  and papers  $p_6$  are written by author  $a_2$ , not by author  $a_1$ . After adding papers, the credit share of author  $a_2$  is 0.54 which is higher than that of author  $a_1$ . In both cases, the target paper  $p_0$  has two authors  $a_1$  and  $a_2$  colored in blue and pink, respectively;  $d_l$  ( $1 \leq l \leq 7$ ) are citing papers of  $p_0$ ; papers  $p_j$  ( $0 \leq j \leq 6$ ) were co-cited by papers that cite  $p_0$ ;  $A$  is the credit allocation matrix obtained by author list of co-cited papers;  $S$  represents the co-citation strength between co-cited papers and target paper;  $C$  is the final credit share matrix. (Color figure online)



**Fig. 2** The maliciously manipulative behavior. **a** The schematic diagram of the maliciously manipulative behavior. The target paper  $p_0$  has two authors  $a_1$  and  $a_2$ . If the credit share of  $a_1$  in paper  $p_0$  is higher than that of  $a_2$  at first according to the CCA method, this credit allocation situation can be changed by adding spamming papers which cite  $p_0$  and papers written by  $a_2$  but not by  $a_1$ . **b** The proportion of Nobel-prize papers affected by the malicious manipulation. The affected rate of Nobel-prize papers means the proportion of papers which the credit share of the original second largest one exceed the original largest one by adding a certain amount of spamming papers. In 17 Nobel-prize papers, for 53%, the original second highest credit share author only needs to add less than 10 papers to make her or his credit share higher than the credit share of the original highest one

This indicates that if we don't measure the scientific impact of citing papers, adding low scientific impact citing papers can easily manipulate the credit share of coauthors. Based on this, we propose a nonlinear collective credit method considering citing papers' scientific impact.

**The nonlinear collective allocation method (NCCA method)**

We put forward a nonlinear collective allocation method (NCCA method) by integrating the scientific impact of citing papers. We consider a paper  $p_0$  with  $M$  coauthors  $a_i (1 \leq i \leq m)$ . In order to calculate the credit share of each author, we first find all papers that cite  $p_0$  in the dataset, forming a set  $D \equiv \{d_1, d_2, \dots, d_l\}$ . Then we find all co-cited papers  $P \equiv \{p_0, p_1, \dots, p_n\}$ , representing the complete set of papers cited by papers in the set  $D$ . The relevance of each co-cited paper  $p_j (0 \leq j \leq n)$  to the target paper  $p_0$  is characterized by its co-citation strength  $s_j$  between  $p_0$  and  $p_j$ , defined as the sum of the  $f_i$  of papers by which  $p_0$  and  $p_j$  are cited together in set  $D$ . Similar to the basic idea of the well-known PageRank algorithm, we can measure a paper's scientific impact by measuring the scientific impact of papers that cite it. Citations are usually used to measure papers' scientific impact. For each paper in set  $D$ , we can compute its citations  $c_l$ , and then we calculate function  $f_l$

$$f_l = \begin{cases} 0, & \text{if } c_l = 0 \\ c_l^\theta, & \text{if } c_l \neq 0 \end{cases} \tag{1}$$

where  $\theta$  is a tunable parameter,  $\theta \geq 0$ . For example, for  $p_1$  in Fig. 3b, we have  $s_1 = f_1$  because only one paper ( $d_1$ ) cites  $p_0$  and  $p_1$  together, whereas  $s_2 = f_1 + f_2 + f_3$  as three papers ( $d_1, d_2$  and  $d_3$ ) cite  $p_0$  and  $p_2$  together. Note that the target paper  $p_0$  is also viewed as a co-cited paper of itself and the co-citation strength equals to the sum of the  $f_i$  of citing papers. By that way, we get the co-citation strength vectors (Fig. 3c). After that, we get a credit allocation matrix  $A$  by using the author list of the co-cited papers (Fig. 3d). The component  $A_{ij}$  refers to the credit share that author  $a_i$  gets from the co-cited paper  $p_j$ . The value of  $A_{ij}$  is calculated by fractional credit allocation. For example, paper  $p_2$  assigns whole credit to the sole author  $a_1$ , whereas  $p_0$  assigns equal (half) credit to authors  $a_1$  and  $a_2$  (Fig. 3d). Consequently, the total credit  $c_i$  of author  $a_i$  is the weighted sum of its local credit acquired from all co-cited papers.

$$c_i = \sum_j A_{ij} s_j \tag{2}$$

or in the matrix form

$$C = AS \tag{3}$$

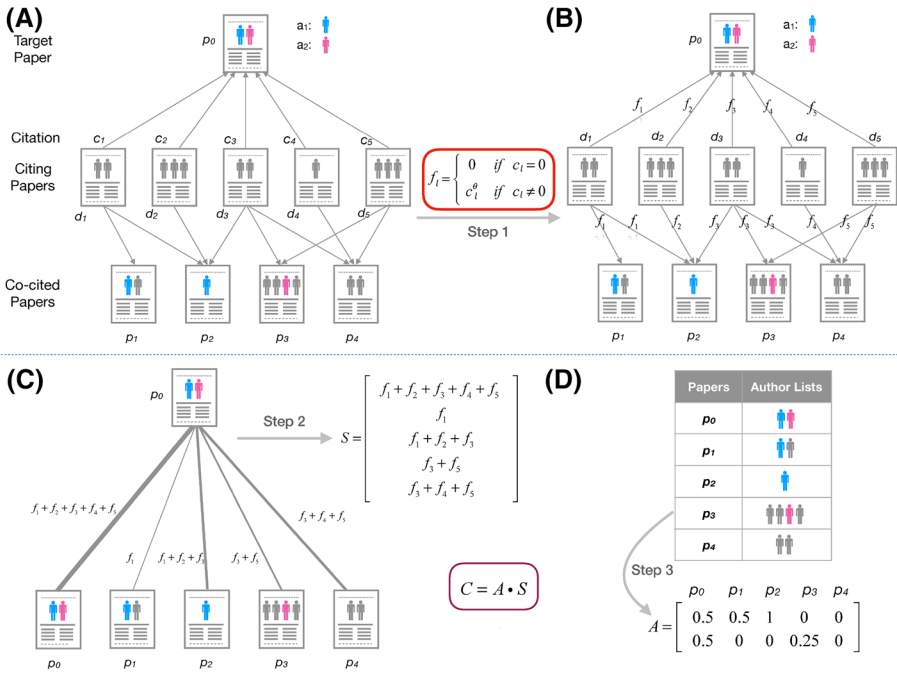
The credit of all authors of target paper  $p_0$  is provided by the vector  $C$ . By normalizing  $C$ , the fractional credit share among coauthors can be obtained.

**Computing the  $h_c$  index**

According to Hirsch's original definition, the  $h$  index can be obtained by

$$h = \max_r (r \leq c(r)) \tag{4}$$

$r$  is the rank attributed to the paper when the publication list is in decreasing order based on the number of citations  $c(r)$ . In Eq. (4) each paper is fully counted for the calculation of its rank,



**Fig. 3** Illustrating the nonlinear credit allocation process. **a** The target paper  $p_0$  has two authors,  $a_1$  and  $a_2$ , colored in blue and pink, respectively. We also show the citing papers  $d_l$  ( $1 \leq l \leq 5$ ), the citation of citing papers  $c_l$  ( $1 \leq l \leq 5$ ) and the co-cited papers  $p_j$  ( $0 \leq j \leq 4$ ) that were cited by these citing papers together with  $p_0$ . **b** Calculating the function  $f_i$ . **b** The co-citation network of  $p_0$  constructed from **a**, where the weights of links denote the co-citation strengths between the co-cited papers and the target paper  $p_0$ . We get the matrix  $S$ . **d** The credit allocation matrix  $A$  obtained from the author lists. The matrix  $A$  provides for each co-cited paper the authors' credit share. For example, because  $p_1$  has  $a_1$  as one of its two authors but it lacks the author  $a_2$ , it votes 0.5 for author  $a_1$  and 0 for author  $a_2$ . With the matrix  $A$  and the co-citation strength matrix  $S$ , the credit share of the two authors of  $p_0$  is computed according to Eq. 2 or 3 with a normalization. (Color figure online)

$$r = \sum_{r'=1}^r 1 \tag{5}$$

An effective rank can be yielded by counting a paper with the credit share  $c(r)$  from the CCA method or the NCCA method of the author,

$$r_{eff}(r) = \sum_{r'=1}^r c_r \tag{6}$$

Which can be used to define the  $h_c$  index,

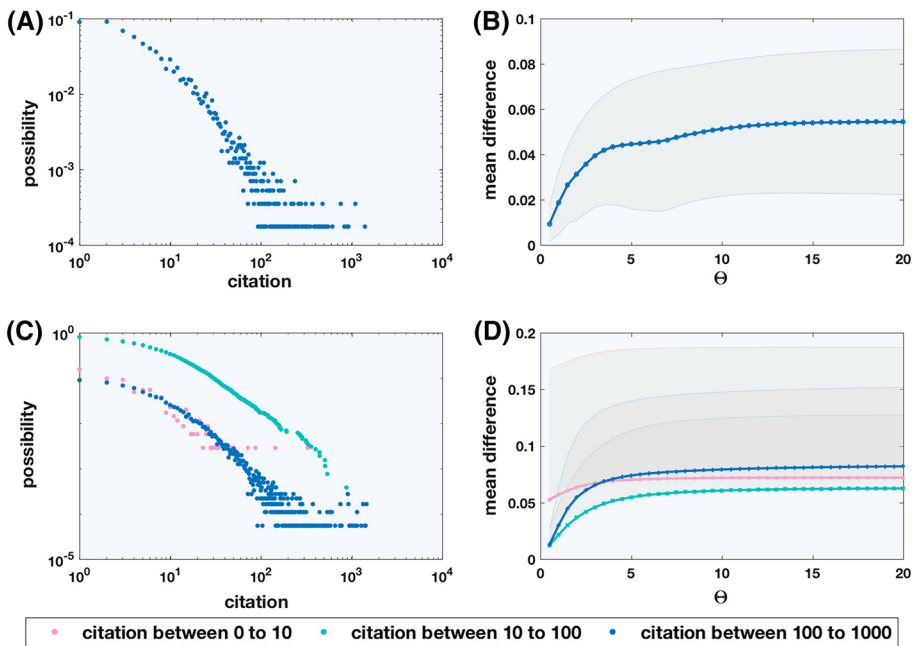
$$h_c = \max_r (r_{eff}(r) \leq c(r)) \tag{7}$$

## Results

### The basic statistics of the NCCA method

The fundamental difference between the NCCA method and the CCA method is that the NCCA method gives different weights to the citing papers of a given paper when using them to allocate credit to authors. In this paper, citation is defined as the scientific impact. As an empirical support of the NCCA method, we compute the citation distributions of citing papers of Nobel-prize papers and three kinds of ordinary papers (Fig. 4a, c). We find that the citation distributions have fat tails, which indicates that most citing papers have low scientific impact. Due to the significant heterogeneous impact of the citing papers, they should be considered differently when relying on them to allocate credit to authors of a given paper.

Parameter  $\theta$  is important in the NCCA method. To investigate the influence of  $\theta$  on the results, we calculate the mean difference of authors' credit share between the CCA method



**Fig. 4** The basic statistics of the NCCA method. **a** The distributions of citing papers' citations of Nobel-prize papers. It indicates that there are lots of low scientific impact citing papers. **b** The mean difference of authors' credit share between the CCA method and the NCCA method with the change of the value of  $\theta$  in the NCCA method in Nobel-prize papers. The mean difference substantially increases before  $\theta < 5$ . **c** The distributions of citing papers' citations of ordinary papers. The pink dots represent papers whose citation range is between 0 and 10. The green dots represent papers whose citation range is between 10 and 100. The blue dots represent papers whose citation range is between 100 and 1000. The phenomenon is that most citing papers' scientific impact is low in ordinary papers too. **d** The mean difference of authors' credit share between the CCA method and the NCCA method with the change of the value of  $\theta$  in the NCCA method in ordinary papers. The pink line represents papers whose citation range is between 0 and 10. The green line represents papers whose citation range is between 10 and 100. The blue line represents papers whose citation range is between 100 and 1000. The mean difference substantially increases before  $\theta < 5$  too. (Color figure online)





corresponding to laureates by the NCCA method is 75%. However, the accuracy of the CCA method is only 70% (Supplementary Materials: Table 3).

### **The robustness of the NCCA method against random perturbations**

Considering that the results of collective credit allocation methods may be influenced by the reference of some papers missing, we validate the robustness of the CCA and NCCA methods against random perturbations, removing a certain proportion of citing papers of a paper. With perturbations, we apply the CCA method and the NCCA method to Nobel-prize papers and ordinary papers whose citation range are between 10 to 100 and between 100 to 1000.

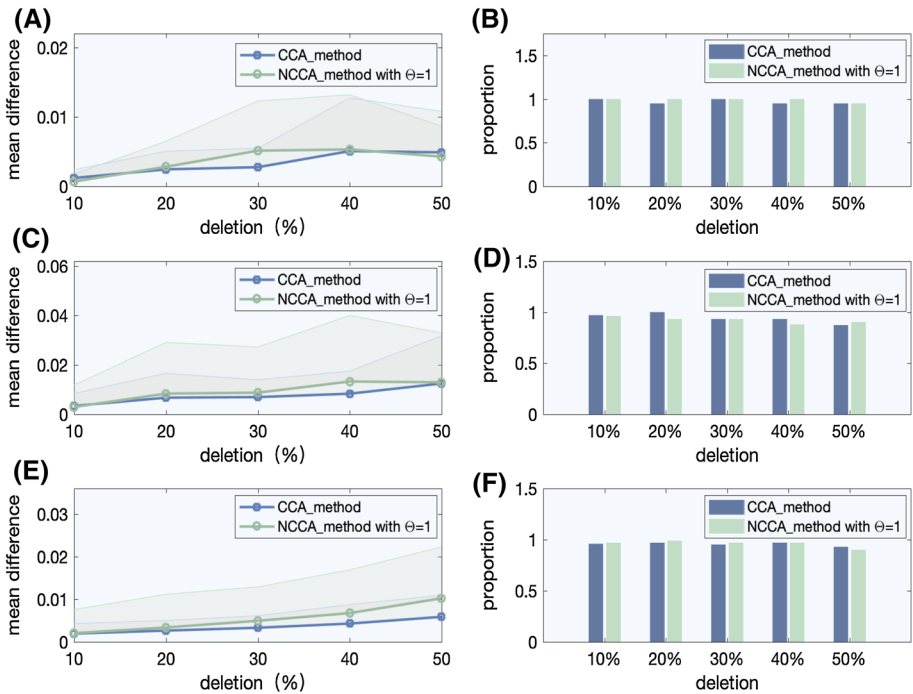
Firstly, we calculate the change of credit share of coauthors after the deletion of citing papers in each paper. We compute the mean difference of credit share in each kind of data. Then, we investigate whether the author with the highest credit share is the same one after removing a proportion of citing papers in each paper. We calculate the proportion of each kind of papers in which the author with highest credit share does not change after perturbations. From the results (Fig. 6), we find that not only in Nobel-prize papers but also in ordinary papers, the mean difference of credit share of coauthors is small and changes slightly with the proportion of citing papers removed increasing. Furthermore, the proportions of papers are above 95 percent satisfying the condition that authors with highest credit share are the same after perturbations. These results indicate that the CCA method and the NCCA method both are robust against random perturbations.

### **The robustness of the NCCA method against the malicious manipulation**

As we find the CCA method is sensitive to manipulation in Nobel-prize papers, we try to testify whether the NCCA method can suppress the manipulative behavior, adding low scientific impact citing papers. What we calculate is the number of added papers needed to make credit share of the second highest author exceed the original highest one. For the malicious manipulation, we suppose that each added paper randomly cites three papers written by the second highest credit share author but not by the highest one. In 20 Nobel-prize papers, we remove three papers and finally focus on 17 Nobel-prize papers. Since there are two Nobel-prize papers in which authors have the same credit share and there is another Nobel-prize paper in which the author with the second highest credit share does not have three papers satisfying the above hypothesis. For the parameters in the NCCA method, the value of  $\theta$  is 1 and we assume the citation of added papers is 1 or 5. This is because papers in APS database receive 1.38 and 4.17 on average in 1 and 3 years after they are published.

The results of different methods are listed in Table 1. Compared with the CCA method, the NCCA method need more added papers, which means that the NCCA method has the better performance in suppressing the effect of added papers, especially for those added papers with low scientific impact.

In Nobel-prize papers dataset, the NCCA method performs well in suppressing the influence of added citing papers with low scientific impact. To verify the NCCA method is also effective to ordinary papers, we again compute the number of added papers needed to make credit share of the second highest author exceed the highest one. For the parameters in the NCCA method, the values of  $\theta$  are set to 0.5, 0.8 and 1, respectively. We assume citations of added papers are 1, 5 and 10.38. The reason why the citation is 10.38 is that the



**Fig. 6** The robustness of the NCCA method against random perturbations. **a** The mean difference of authors’ credit share in Nobel-prize papers after removing a certain proportion of citing papers. **b** The proportion of Nobel-prize papers in which the highest credit share author does not change after removing a certain proportion of citing papers. **c** The mean difference of authors’ credit share in ordinary papers after removing a certain proportion of citing papers, whose citation range is between 10 and 100 (randomly choosing 100 papers). **d** The proportion of papers in which the highest credit share author does not change after removing a certain proportion of citing papers. The citation range of those papers is between 10 and 100 (randomly choosing 100 papers). **e** The mean difference of authors’ credit share in ordinary papers after removing a certain proportion of citing papers, whose citation range is between 100 and 1000 (randomly choosing 100 papers). **f** The proportion of papers in which the highest credit share author does not change after removing a certain proportion of citing papers. The citation range of papers is between 100 and 1000 (randomly choosing 100 papers)

average degree of citation network is 10.38 in the APS dataset. To every kind of ordinary papers, we randomly select 50 papers from that kind of papers in APS database and do it 50 times, then we calculate the number of added papers needed on average. The result is showed in Fig. 7. We find that in all kinds of ordinary papers, the NCCA method always needs more added papers than the CCA method. Besides, more added papers are needed for ordinary papers with larger citations range. Overall, the NCCA method is more robust against the malicious manipulation compared with the CCA method.

**The  $h_c$  index**

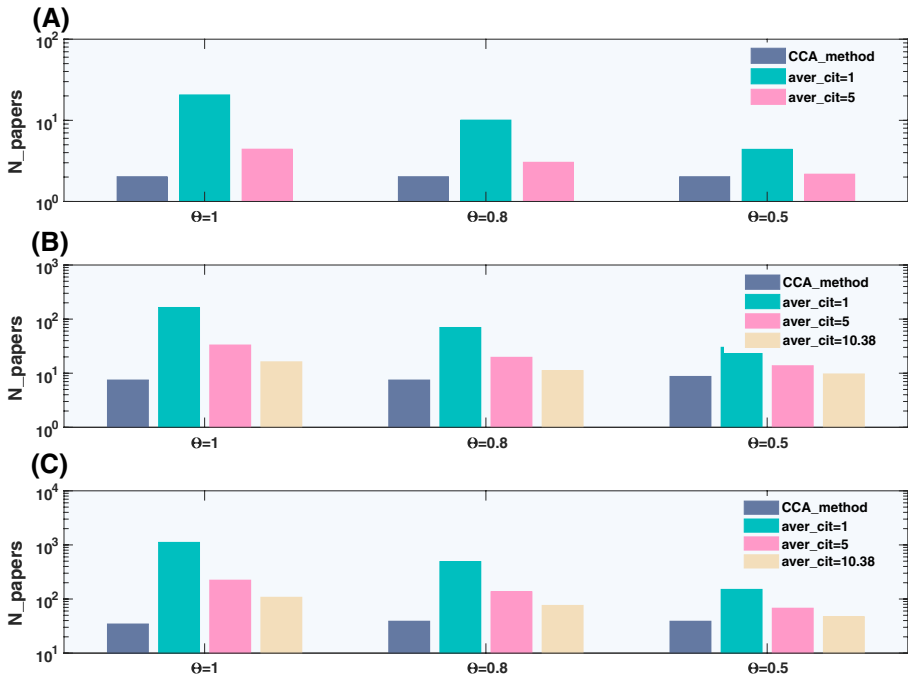
Schreiber proposed  $h_m$  index which is modified by counting the papers fractionally based on the number of authors. The  $h_m$  index can result in a different rank compared with the original h-index. According to the way to compute  $h_m$  index, we put forward  $h_c$  index modified by the

**Table 1** The number of papers (N\_papers) needed to make credit share of the original second highest credit share author exceed the original highest one in Nobel-prize papers by using different kinds of collective credit allocation methods. ‘ $\theta = 1, \text{cit}=5$ ’ means the NCCA method with parameter  $\theta = 1$  and the citation of added spamming papers is 5; ‘ $\theta = 1, \text{cit}=1$ ’ means the NCCA method with parameter  $\theta = 1$  and the citation of added spamming papers is 1

DOI	Publishing year	Awarding year	N_papers		
			CCA	$\theta = 1, \text{cit} = 5$	$\theta = 1, \text{cit} = 1$
10.1103/PhysRevLett.61.169	1988	2013	1	49	245
10.1103/PhysRevLett.77.4887	1996	2012	2	2	8
10.1103/PhysRevLett.58.1490	1987	2002	2	19	95
10.1103/PhysRev.112.1940	1958	1964	5	33	164
10.1103/PhysRev.83.333	1951	1994	7	24	119
10.1103/PhysRev.73.679	1948	1981	7	27	135
10.1103/PhysRev.69.37	1946	1952	8	71	355
10.1103/PhysRevLett.84.5102	2000	2005	8	99	494
10.1103/PhysRevLett.61.826	1988	1997	9	113	565
10.1103/PhysRevLett.55.48	1985	1997	16	194	966
10.1103/PhysRevLett.76.1796	1996	2012	18	106	529
10.1103/PhysRevLett.20.1205	1968	2002	24	107	532
10.1103/PhysRevLett.84.3232	2000	2005	42	204	1020
10.1103/PhysRev.122.345	1961	2008	112	725	3621
10.1103/PhysRevLett.57.2442	1986	2007	147	711	3551
10.1103/PhysRevLett.75.3969	1995	2001	239	2084	10,416
10.1103/PhysRevLett.61.2472	1988	2007	360	1784	8920

collective credit allocation method. In order to find out whether the  $h_c$  index can make any significance when ranking scientists, we select top ten authors according to the number of coauthors in APS dataset and compute their  $h$  index,  $h_m$  index,  $h_{c0}$  index(the CCA method),  $h_{c1}$  index( the NCCA method with  $\theta = 0.5$ ) and  $h_{c2}$  index( the NCCA method with  $\theta = 1$ ). The result is shown in Table 2.

We find that  $h_c$  index can lead to a different ranking of authors. Although some authors have the same rank in  $h$  index and  $h_m$  index, the rank of these authors in  $h_c$  index changes greatly, like author9 and author10. Compared with  $h$  index and  $h_m$  index, the value is smaller in  $h_c$  index and a few authors have same values in  $h_c$  index. In addition, the rank in  $h_{c0}$ ,  $h_{c1}$  and  $h_{c2}$  are almost same. The credit share of authors calculated by the CCA or NCCA method can have huge difference in different papers even though papers have the same number of authors. The collective process in the scientific community determines the credit allocation in the CCA or NCCA method. That’s why the  $h_c$  index is more reasonable than  $h_m$  index, since the credit allocation is more fair in collective credit allocation method than other methods calculating with the inverse of the number of coauthors.



**Fig. 7** The number of papers needed to make credit share of the second highest author exceed the highest one in ordinary papers with different parameters in the NCCA method. ‘CCA\_method’ means the CCA method; ‘aver\_cit = 1’ means the citation of added papers is 1 in the NCCA method; ‘aver\_cit = 5’ means the citation of added papers is 5 in the NCCA method; ‘aver\_cit = 10.38’ means the citation of added papers is 10.38 in the NCCA method. **a** Ordinary papers with the range of citation between 0 and 10. **b** Ordinary papers with the range of citation between 10 and 100. **c** Ordinary papers with the range of citation between 100 and 1000

**Table 2** The  $h$  index, the  $h_m$  index, the  $h_{c0}$  index(the CCA method), the  $h_{c1}$  index( the NCCA method with  $\theta = 0.5$ ) and the  $h_{c2}$  index( the NCCA method with  $\theta = 1$ ) of top ten authors. “author 1” to “author 10” are ranked by the number of coauthors

Authors	$h$	rank( $h$ )	$h_m$	rank( $h_m$ )	$h_{c0}$	rank( $h_{c0}$ )	$h_{c1}$	rank( $h_{c1}$ )	$h_{c2}$	rank( $h_{c2}$ )
Author1	16	1	11.67	1	1.35	3	1.35	3	1.36	3
Author2	14	5	7.13	6	0.74	5	0.75	6	0.76	6
Author3	15	2	8.33	5	1.07	4	1.04	4	1.01	4
Author4	14	5	9.33	4	4.29	1	4.20	1	4.07	1
Author5	5	10	2.83	10	0.06	9	0.06	9	0.05	9
Author6	14	5	7.13	6	0.58	7	0.61	7	0.64	7
Author7	14	5	6.66	8	0.46	8	0.53	8	0.58	8
Author8	12	9	6.66	8	0.03	10	0.03	10	0.03	10
Author9	15	2	10.33	2	3.10	2	3.06	2	3.02	2
Author10	15	2	10.33	2	0.74	5	0.78	5	0.80	5

## Discussion and conclusions

In this paper, we propose a nonlinear collective credit allocation method (NCCA method) to assign credits to authors in a given paper. Similar to the existing method (CCA method), the NCCA method is time dependent and is determined by papers citing the target paper. The author whose papers are more co-cited with the target paper will receive higher credit. Compared to the CCA method, the modification is that the NCCA method takes into account the impact of citing papers when evaluating the co-citation relations. This modification can significantly enhance the robustness of the NCCA method with respect to random perturbations and the malicious manipulation. The comparison of the CCA method and the NCCA method is conducted in both Nobel-prize papers as well as ordinary papers. The results indicate that the credit share assigned by the NCCA method is much less influenced by adding citing papers with low scientific impact. In addition, the CCA and NCCA method can be also applied to modify  $h$  index.  $h_c$  index can result in a different rank and measure the scientific impact of scientists more fairly. However, the procedure of the NCCA method is complicated and thus its wide applicability is questionable.

Additionally, several promising extensions can be made in the future. For example, with different definitions of papers' scientific impact such as PageRank values, we may find one that can improve the results of the NCCA method. APS dataset only has the physics field data. We need to validate the results of the NCCA method using other fields data.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (Grant Nos. 61603046 and 61374175) and Natural Science Foundation of Beijing (Grant No. L160008).

## References

- Allen, L., Brand, A., Scott, J., Altman, M., & Hlava, M. (2014). Credit where credit is due. *Nature*, *508*(7496), 312–313.
- Ancheyta, J. (2015). A correction of h-index to account for the relative importance of authors in manuscripts. *International Journal of Oil Gas and Coal Technology*, *10*, 221–232.
- Ausloos, M. (2015). Assessing the true role of coauthors in the h-index measure of an author scientific impact. *Physica A*, *422*, 136–142.
- Bao, P., & Zhai, C. (2017). Dynamic credit allocation in scientific literature. *Scientometrics*, *112*(1), 595–606.
- Batista, P. D., Campiteli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, *68*(1), 179–189.
- Bornmann, L., & Daniel, H. D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385.
- Burrell, Q. L. (2007). Should the h-index be discounted. *ISSI Newsletter*, *5*(5), 65–68.
- Crispo, E. (2015). A new index to use in conjunction with the h-index to account for an author's relative contribution to publications with high impact. *Journal of the Association for Information Science and Technology*, *66*, 2381–2383.
- Dong, Y., Ma, H., Shen, Z., & Wang, K. (2017, August). A century of science: Globalization of scientific collaborations, citations, and innovations. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1437–1446). ACM.
- Egghe, L. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology*, *59*, 1608–1616.
- Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the Association for Information Science and Technology*, *51*(2), 145–157.
- Foulkes, W., & Neylon, N. (1996). Redefining authorship. Relative contribution should be given after each author's name. *BMJ: British Medical Journal*, *312*(7043), 1423.

- Galam, S. (2011). Tailor based allocations for multiple authorship: A fractional gh-index. *Scientometrics*, 89, 365–379.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471–479.
- Greene, M. (1999). Policy on Papers' contributors. *Nature*, 399(6735), 393.
- Greene, M. (2007). The demise of the lone author. *Nature*, 450(7173), 1165.
- Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS ONE*, 3(12), e4021.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569.
- Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49), 19193–19198.
- Hirsch, J. E. (2019).  $h_a$ : An index to quantify an individual's scientific leadership. *Scientometrics*, 118(2), 673–686.
- Kaur, J., Radicchi, F., & Menczer, F. (2013). Universality of scholarly impact metrics. *Journal of Informetrics*, 7(4), 924–932.
- Kennedy, D. (2003). Multiple authors, multiple problems. *Science*, 301(5634), 733.
- Kim, J., & Diesner, J. (2014). A network-based approach to coauthorship credit allocation. *Scientometrics*, 101(1), 587–602.
- Lawrence, P. A. (2007). The mismeasurement of science. *Current Biology*, 17(15), R583–R585.
- Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, 444(7122), 1003.
- Liu, X. Z., & Fang, H. (2012). Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution. *Journal of Informetrics*, 6, 557–565.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5200–5205.
- Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80(5), 056103.
- Schreiber, M. (2008a). To share the fame in a fair way, hm modifies h for multi-authored manuscripts. *New Journal of Physics*, 10, 040201.
- Schreiber, M. (2008b). A modification of the h-index: The hm-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2, 211–216.
- Schreiber, M. (2009). A case study of the modified Hirsch index hm accounting for multiple coauthors. *Journal of the American Society for Information Science and Technology*, 60, 1274–1282.
- Sekercioglu, C. H. (2008). Quantifying coauthor contributions. *Science*, 322(5900), 371–371.
- Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences*, 111(34), 12325–12330.
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(aa6312), 5239.
- Stallings, J., Vance, E., Yang, J., Vannier, M. W., Liang, J., Pang, L., et al. (2013). Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences*, 110(24), 9680–9685.
- Trueba, F. J., & Guerrero, H. (2004). A robust formula to credit authors for their publications. *Scientometrics*, 60(2), 181–204.
- Tscharnitke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. *PLoS Biology*, 5(1), e18.
- Van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, 48(10), 944–945.
- Vavrycuk, V. (2018). Fair ranking of researchers and research teams. *PLoS ONE*, 13(4), e0195509.
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
- Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., et al. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, 714, 1–73.