Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

First and Others credit-assignment schema for evaluating the academic contribution of coauthors*

Li WEIGANG
(*TransLab, Department of Computer Science, University of Brasilia, Brasilia-DF 70910-900, Brazil*)
E-mail: weigang@unb.br

Received Feb. 8, 2016; Revision accepted Apr. 25, 2016; Crosschecked Jan. 20, 2017

Abstract: Credit-assignment schemas are widely applied by providing fixed or flexible credit distribution formulas to evaluate the contributions of coauthors of a scientific publication. In this paper, we propose an approach named First and Others (F&O) counting. By introducing a tuning parameter α and a weight β, two new properties are obtained: (1) flexible assignment of credits by modifying the formula (with the change of α) and applying preference to the individual author by adjusting the weights (with the change of β), and (2) calculation of the credits by separating the formula for the first author from others. With formula separation, the credit of the second author shows an inflection point according to the change of α. The developed theorems and proofs concerning the modification of α and β reveal new properties and complement the base theory for informetrics. The F&O schema is also adapted when considering the policy of 'first-corresponding-author-emphasis'. Through a comparative analysis using a set of empirical data from the fields of chemistry, medicine, psychology, and the Harvard survey data, the performance of the F&O approach is compared with those of other methods to demonstrate its benefits by the criteria of lack of fit and coefficient of determination.

Key words: Bibliometrics; Credit of coauthorship; *H*-index; Informetrics; Scholar information
http://dx.doi.org/10.1631/FITEE.1600991 **CLC number:** TP391

1 Introduction

Since the *H*-index was proposed by Hirsch (2005) with a clear concept and simple expression, it has become a basic metric to evaluate an individual's scientific research output. Bibliometric counting methods, together with the *H*-index, play an important role in academia, as they are used to determine the level of contribution of researchers by their publications. Major citation databases such as Google Scholar, Microsoft Academic Search, Scopus, Web of Science, and even some academic social networks (e.g., ResearchGate) are using the *H*-index and related metrics to evaluate the academic impact of a scientist.

Normally, when a published article is referenced by other researchers, it earns a new citation. Each author of this article will then receive a score of 1 as credit under the equal contribution (EC) norm (Boas, 1964; Endersby, 1996; Maciejovsky *et al.*, 2009). This statistical method is named whole counting (WC), which is referred to as the *N*/*N* evaluation model in this paper. Although many scientists and researchers have accepted this simple and effective model, it fails to take into account the difference in contribution from coauthors and may encourage new articles to list more authors without restriction. Because the *H*-index is based on the assumption that each author has contributed equally to the project, it is difficult to accurately reflect an individual's actual academic level and contribution by using the *N*/*N* model. Additionally, government administrations and commercial institutions frequently use this scoring system to review applications for promotions or research funding. Inaccurate scoring systems may then

---

* Project supported by the Brazilian National Council for Scientific and Technological Development (No. 304903/2013-2)
ORCID: Li WEIGANG, http://orcid.org/0000-0003-1826-1850

lead to an inefficient allocation of resources based on over (or under) estimation of the academics' contributions.

The Nature Index 2015 Global is calculated using the weighted-scoring method, and also follows the EC norm. For a paper with N authors, if it earns a new citation, the system will assign each author a score of $1/N$ (May and Brody, 2015) using the fractional counting (FC) (Narin, 1976; Lindsey, 1980; Gauffriau *et al.*, 2007), which is referred to as the $1/N$ evaluation model in this paper. When compared to the N/N model, the $1/N$ model generates a more reasonable representation of the contributions of the authors and their institutions for the development of global science and technology. Other researchers (He *et al.*, 2012; Lin *et al.*, 2013) also studied the patterns of author orders for the evaluation of the influence at the institution level.

It is difficult to evaluate the contribution of coauthors because of various subjective and objective factors (Price, 1981; Larsen, 2008). There are many possibilities of each author's level of contribution toward a published article. For example: (1) the coauthors delineate their contributions in the acknowledgement section or in another section of the article; (2) the coauthors implicitly indicate their contributions by listing themselves in alphabetical order (equal contribution, Waltman (2012)); (3) no coauthor indicates equal or unequal contribution, and the corresponding author is mentioned. Due to the large variation in author contribution, N/N and $1/N$ models are not sufficient for analyzing all scenarios, or determining accurate contribution levels in all cases. This serves as the motivation for this study.

In general, sequence-determines-credit (SDC) approaches (Tscharntke *et al.*, 2007; Yang *et al.*, 2014) can be effectively used to analyze most of the above-mentioned scenarios. In this paper, all SDC approaches are classified in the S/N evaluation model. Cole and Cole (1973) argued that contribution from the first author should be emphasized, and in certain research fields, the straight counting for the first author is applied when evaluating contribution. They proposed Coles counting to attribute first author with score 1 and others with 0. As this method is very simple, it makes the data collection and analysis process much easier.

Vinkler (1993; 2000) analyzed the contribution, authorship, and team cooperativeness with Central Research Institute for Chemistry (CRIC) counting. Egghe *et al.* (2000) presented the geometric index as a method for accrediting publications to authors or countries. Kalyane and Vidyasagar Rao (1995), van Hooydonk (1997), Zhang (2009), and Abbas (2011) applied arithmetic counting or its variants to calculate weighted citations. Harmonic counting was proposed by Hodge and Greenberg (1981) and applied by Hagen (2010; 2013). Assimakis and Adam (2010) developed the p-index as a new author's productivity index. Stallings *et al.* (2013) used an axiomatic approach with a collaboration index to assign relative credits to the coauthors. All of these schemas support the hypothesis of decreasing the contribution index from the first author, but they lack the flexibility of adjusting the weights for different authors.

On the other hand, the contribution of the corresponding author should also be considered (Wren *et al.*, 2007). Huang *et al.* (2011) developed a straight counting method to focus on the corresponding author. However, it is difficult to determine appropriate weights to apply for a given field, discipline, or even a single article. The first-last-author-emphasis (FLAE) method was investigated by Trueba and Guerrero (2004) and Tscharntke *et al.* (2007). FLAE and other corresponding author emphasis methods are classified as the U/N model, which will be defined in this paper.

Due to the large quantity of scientific publication data in the Web (Dong *et al.*, 2015), developing efficient evaluation methods is an emergent area of study. As the objective of this study, a new credit-assignment schema is proposed under the S/N model group, hereinafter referred to as 'First and Others' (F&O) counting. A tuning parameter α is introduced to change the formula for flexibility consideration, and the weight β (or the weight vector **BETA**=$[\beta_2, \beta_3, \ldots, \beta_N]$) is designed to assign the preference to any individual author if necessary. Two main properties of this schema are: (1) flexible assignment credits by modifying the formula (with changing α) and applying the preference to an individual author by adjusting the weight (with changing β); (2) calculation of the credits by separating the formula for the first author from others. With this separation, the credit of the second author shows an inflection point according to the change of α. This adjustment defines the ratio of credits, ρ_i, between authors, especially between the first and second authors, to analyze the property of the counting and make F&O counting more applicable. With the incorporation of α and β, the F&O approach generalizes some existing credit assignment approaches such as straight and harmonic counting. The theorems and proofs related to the F&O and associated parameters provide the base theory for studying the distribution of credits in the informatics field. Considering the special contributions of the first author and corresponding author, we also modify F&O counting to improve upon the first-corresponding-author-emphasis solution (FCAE).

2 Flexible formula and weight preference of credit-assignment schemas

The main contribution of a scientific paper is in sharing new academic ideas and technology developments. Scientific research is a collaborative endeavor, and the participants are normally listed as the coauthors of the publication. During the past decade, a very large number of citation impact indicators have been introduced (Waltman, 2015).

For better conducting the discussion, some selected credit-assignment schemas and their formulas are listed in Table 1. The last column shows the ratio between the credits of the ith and $(i+1)$th authors, which is defined as Credit(i)/Credit(i+1). This ratio is a useful parameter to present an important feature of the credit distribution.

As mentioned by Xu *et al.* (2016) after analyzing 15 counting approaches, author-credit allocation schemas have been studied and categorized by Abbas (2011), Liu and Fang (2012), and Du and Tang (2013). In Xu *et al.* (2016), these approaches were classified into three types: linear, curve, and other credit-assignment schemes. Kim and Kim (2015) categorized the schemas of unequal coauthorship credit allocation into fixed and flexible ones. According to this criterion, the first four approaches in Table 1 (Eqs. (1)–(4)) are the formulas with fixed parameters. If i and N (author number and the total number of authors, respectively) are defined, then Credit(i)/Credit(i+1) is unchanged. In a special case of geometric counting, this ratio is always 2. The last five approaches in Table 1 (Eqs. (5)–(9)) are flexible to obtain the variant assignment credits depending on the additional parameters. There are still two basic classes: flexible formula with changeable parameters and flexible weight to address the preference for the individual coauthor.

2.1 Flexible formula with changeable parameters

In case of the corrected contribution scores (CCS) model (Lukovits and Vinkler, 1995), the credit of the first author is separated from others. The ratio, Credit(i)/Credit(i+1), depends on the value of T, which is a percentage of the coauthor contribution threshold. For a paper with fewer than 10 coauthors ($N\leq 10$), T=10%; for $N\leq 20$, T=5%. As in most cases, N is less than 10, T is fixed at 10%, and when N=5, Credit(1)/Credit(2)=2.

The approach proposed by Liu and Fang (2012) can be presented in various forms with the variant of tuning parameter q. When q=0, Eq. (6) is identical to the fractional credit $1/N$; when q=1, it is identical to harmonic counting. This method is flexible: changing q from 0 to 1 to meet the needs of the application. When q is determined, the credit formula is fixed. Credit(i)/Credit(i+1) depends only on q.

Kim and Diesner (2014) introduced a novel network-based approach (NBA) as a robust and flexible framework for coauthorship credit allocation (Eq. (7)). This model generates a different set of credits depending on the distribution factor d. In Kim and Kim (2015) the parameters can be tuned for model flexibility (including arithmetic, geometric, and harmonic models) and for better performance against empirical data. All these approaches belong to the category listed in Section 2.1 because when i, N, and d are determined, Credit(i)/Credit(i+1) is fixed.

2.2 Flexible weight for individual coauthor preference

Trueba and Guerrero (2004) tried to eliminate the undesirable property of a fixed formula. They developed a schema with the capability to modify the weights to assign preference to the individual coauthors. They used a basic formula for credit-assignment first, and then assigned arbitrarily a fixed fraction of total credit, f ($0\leq f\leq 1$), to a few favored coauthors. The sum of the credits according to the basic formula is not 1 but $1-f$. In the case where these favored authors are the first, second, and last, the first will obtain additional credit as $c_1 f$, the second $c_2 f$, and

Table 1 Selected credit-assignment schemas and their formulas

Counting	Reference	Formula	Credit(i)/Credit(i+1)
Harmonic	Hodge and Greenberg (1981); Hagen (2010)	$\dfrac{1/i}{1+1/2+\cdots+1/N}$ (1)	$\dfrac{i+1}{i}$
Arithmetic	Kalyane and Vidyasagar Rao (1995); van Hooydonk (1997)	$\dfrac{2(N-i+1)}{N(N+1)}$ (2)	$\dfrac{N-i+1}{N-i}$
Geometric	Egghe *et al.* (2000)	$\dfrac{2^{N-i}}{2^N-1}$ (3)	2
A index	Stallings *et al.* (2013)	$\dfrac{1}{N}\sum_{j-i}^{N}\dfrac{1}{j}$ (4)	$\dfrac{\frac{1}{i}+\frac{1}{i+1}+\ldots+\frac{1}{N}}{\frac{1}{i+1}+\ldots+\frac{1}{N}}$
Corrected contribution scores (CCS)	Lukovits and Vinkler (1995)	$\begin{cases} \mathrm{CCS}(1,N)=\dfrac{N+1}{2NF} \\ \mathrm{CCS}(i,N)=\dfrac{i+T}{2iFT} \\ F=\dfrac{1}{2}\left[\dfrac{1}{N}+\dfrac{N+1}{T}+\sum_{j=1}^{N}\dfrac{1}{j}\right], T=\dfrac{100}{H} \end{cases}$ (5)	$\dfrac{(i+1)(i+T)}{i(i+T+1)}$, T=0−100
L&F	Liu and Fang (2012)	$\dfrac{i^{-q}}{\sum_{j=1}^{N} j^{-q}}$ (6)	$\dfrac{(i+1)^q}{i^q}$
Network-based approach	Kim and Diesner (2014)	$\begin{cases} v_1^N=v+v_t\sum_{j=1}^{N-1}\dfrac{1}{N-j} \\ v_1^N=(v-v_t)+v_t\sum_{j=1}^{N-1}\dfrac{1}{N-j} \\ v_N^N=(v-v_t), v_t=dv, 0<d<1 \end{cases}$ (7)	$\dfrac{v_i^N}{v_{i+1}^N}$ This ratio depends on d
Arithmetic (α)	Abbas (2010)	$\dfrac{1}{N}+\dfrac{N-2i+1}{2}\alpha,$ $0\le\alpha\le 1, \alpha<\dfrac{2}{N(N-1)}, N\ge 2$ (8)	$\dfrac{2+N(N-2i+1)\alpha}{2+N(N-2i-1)\alpha}$
T&G	Trueba and Guerrero (2004)	$\begin{cases} \text{T\&G}(1,N)=\dfrac{2N+1}{N(N+1)}\dfrac{2}{3}(1-f)+c_1f \\ \text{T\&G}(2,N)=\dfrac{2}{N(N+1)}\dfrac{2}{3}(1-f)+c_2f \\ \text{T\&G}(i,N)=\dfrac{2N-i+2}{N(N+1)}\dfrac{2}{3}(1-f), \\ \qquad i=3,4,\ldots,N-1 \\ \text{T\&G}(N,N)=\dfrac{N+2}{N(N+1)}\dfrac{2}{3}(1-f)+c_3f \\ c_1+c_2+c_3=1,\ c_1\ge c_2\ge c_3 \end{cases}$ (9)	$\dfrac{2N-i+2}{2N-i+1}$ This ratio is from the basic formula without the weights of f and c. By Eq. (9), the ratio is changeable depending on f and c

the last c_3f ($c_1+c_2+c_3=1$). With the introduction of the new weights, Credit(1)/Credit(2) depends on both f and c_i.

Abbas (2010) modified arithmetic counting by introducing a weight α as a preference for the individual coauthor ($\alpha<2/(N(N-1))$, $N\ge2$). For example, when $\alpha=1/6$ and $N=3$, it is the common arithmetic counting. The intention of Abbas was to use a weight to give preference to the individual. As α is fixed for obtaining the credit for every coauthor, this approach has limited applications.

From the reviews of Waltman (2015), Xu *et al.*

(2016), etc., there is no credit-assignment schema using both approaches of flexible formula with flexible parameter(s) and flexible weight preference for an individual coauthor. Developing a new schema with these two mechanisms is the main purpose of this study. In this paper, we will introduce F&O counting and its application by comparing it with the existing indicators.

3 First and Others counting for evaluating credits of coauthors

This section presents the basic definitions, theorems, and properties of F&O counting to show the new profiles and advantages compared to others.

3.1 Definitions of F&O and parameters

In the studies concerning Google Scholar and other large academic databases, to describe the heterogeneous network consisting of literature, authors, and all relationships between them, an effective calculation method should be applied for the quantitative analysis of coauthor contributions. When analyzing the relationship between the patent and coinventor, Du *et al.* (2015) used the 1/N model to calculate the coinventor's rank. Weigang *et al.* (2015) proposed a concept of micro scholar social networks (MSSNs) to establish smaller groups within a related field for evaluating the impact of academic literature and scholars. A simple formula was defined to calculate the contribution credits of the coauthors in MSSN. Based on these studies, F&O counting is given by the following definitions.

Definition 1 First and Others counting, F&O(i, N), is defined to estimate the contributions of the ith (i=1, 2, …, N) coauthor of a publication:

$$\mathrm{F\&O}(i,N)=\begin{cases}1-\sum\limits_{j=2}^{N}\mathrm{F\&O}(j,N), & i=1,\\ \dfrac{\beta}{i+(N-\alpha)}, & i=2,3,\ldots,N,\ \alpha\le N,\end{cases} \tag{10}$$

where 0≤F&O(i, N)≤1 (i=1, 2, ..., N) and α and β are the selected values (see details in Definitions 2 and 3). Similar to harmonic counting and A-index, Eq. (10) presents a formula of the SDC approach, notated as F&O or S/N F&O counting. To better understand Eq. (10), Table 2 illustrates the normative credit contributions by F&O for up to six authors when α=1.5 and β=1, and Fig. 1 shows the credit distributions for N=2, 3, …, 10 with α=2 and β=1.

Table 2 Credits by F&O for up to six authors (α=1.5, β=1)

Number of authors (N)	F&O(i, N) i=1	i=2	i=3	i=4	i=5	i=6
1	1.000					
2	0.600	0.400				
3	0.492	0.286	0.222			
4	0.442	0.222	0.182	0.154		
5	0.413	0.182	0.154	0.133	0.118	
6	0.394	0.154	0.133	0.118	0.105	0.095

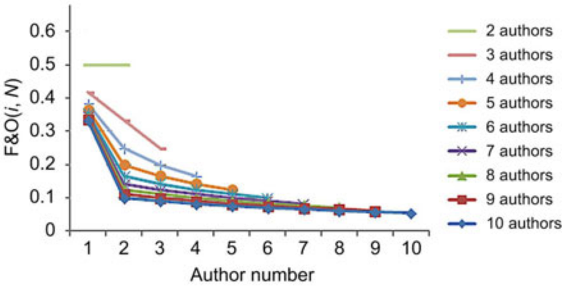

Fig. 1 Distribution of credits for authors by F&O (α=2 and β=1) (References to color refer to the online version of this figure)

Definition 2 α is defined as a tuning parameter in F&O counting, which makes the F&O formula flexible and adjusts the linear distribution of the credits of the coauthors.

The value of α is used to adjust F&O counting to present the percentage of the credit of the first author over others. Theoretically speaking, α can be less than or equal to N. It is suggested α≤(N+2)/2: for 3≤N≤4, α≤(N+1)/2; for 5≤N≤8, α≤N/2; for 9≤N≤14, α≤(N−1)/2; and so on. When α decreases, the percentage of the credit of the first author over others increases. For α=2, β=1, and N>2, the credits obtained by using Eq. (10) from the second to the last authors are 1/N, 1/(N+1), …, 1/(2N−2). These patterns are important for understanding the credit distribution.

Definition 3 β is defined as a weight of F&O counting and used to adjust the preference credit for an individual author i (i=2, 3, …, N, N≥2). To identify the different contributions of some special coauthors, a weight vector is defined as **BETA**=[β_2, β_3, …, β_N].

Usually, β is a constant with a constraint β=1. In a special case, different β_i's give different preferences to coauthor i in credit calculation. Considering the importance of a corresponding author in the ith position, β_i can be used in scenarios of straight counting for the corresponding author to adjust the credit of the ith author in Eq. (10). The philosophy of the counting method is to use the simple formula with a few parameters, and the weight vector **BETA** should be applied for special cases only, to give preference for the common first or corresponding authors.

3.2 Basic theorems related to F&O counting

Theorem 1 Considering the sequence-determines-credit (SDC) situation and that no coauthors claim preference contribution, F&O counting follows ranking preference and credit normalization.

Proof If no coauthor claims the preference contribution, β=1 will be selected as a suitable value such that the credits of coauthors are distributed following a decreasing sequence. By Definition 1, considering $N\geq2$ and $\alpha\leq N$, we have the following:

for i=2,

$$\text{F\&O}(2,N)=\frac{\beta}{i+N-\alpha}=\frac{1}{2+N-\alpha}\leq\frac{1}{N};$$

for i=3,

$$\begin{aligned}\text{F\&O}(3,N)&=\frac{\beta}{i+N-\alpha}=\frac{1}{3+N-\alpha}\\&\leq\frac{1}{2+N-\alpha}=\text{F\&O}(2,N)\leq\frac{1}{N};\end{aligned}$$

…

for i=N−1,

$$\begin{aligned}\text{F\&O}(N-1,N)&=\frac{\beta}{i+N-\alpha}=\frac{1}{2N-\alpha-1}\\&\leq\frac{1}{2N-\alpha-2}=\text{F\&O}(N-2,N)\leq\frac{1}{N};\end{aligned}$$

and for i=N,

$$\begin{aligned}\text{F\&O}(N,N)&=\frac{\beta}{i+N-\alpha}=\frac{1}{N+N-\alpha}\\&=\frac{1}{2N-\alpha}\leq\frac{1}{2N-\alpha-1}=\text{F\&O}(N-1,N)\leq\frac{1}{N}.\end{aligned}$$

As $\sum_{i=2}^{N}\text{F\&O}(i,N)<(N-1)/N$, F&O(1, N)=1−$\sum_{i=2}^{N}\text{F\&O}(i,N)\geq1/N$ and F&O(1, N)≥F&O(2, N) ≥⋯≥F&O(N, N)>0. It is proved that F&O counting follows ranking preference and credit normalization.

Theorem 1 shows that F&O counting follows the first two axioms of Stallings *et al.* (2013). The ranking preference axiom notes that if no coauthors claim equal contribution, the contribution credits of coauthors are ranked according to the order of the authors. The credit normalization axiom states that the sum of the contribution credits of all coauthors should be 100%. The maximum entropy axiom emphasizes that the credit vector of the coauthors for a publication is uniformly distributed in the domain defined by the first two axioms (see details in Stallings *et al.* (2013)).

Most existing credit-assignment schemas are presented as a unique formula. Some schemas have separate equations, such as those of Lukovits and Vinkler (1995), Trueba and Guerrero (2004), Abbas (2010), and Kim and Diesner (2014), but there are a few in-depth discussions of the inflection point in the credit distribution to separate the formula into two or more parts. With this separation, before the inflection point, the schema follows a formula (a function of credit distribution); after and including this point, the schema follows another formula (another function of the distribution). Theorem 2 and its proof are constituted.

Theorem 2 If no coauthors claim preference contribution, the credit of the second author by F&O counting acts as an inflection point in the credit distribution according to the change of α. Specifically, when α=2, there is F&O(2, N)=1/N for this point.

Proof For i=2, 3, ..., N (N>2) and β=1, according to Eq. (10), let Δ_1=F&O(1, N)−F&O(2, N). Then F&O(1, N)=F&O(2, N)+Δ_1;

Δ_2=F&O(2, N)−F&O(2, N)=0;

Δ_3=F&O(2, N)−F&O(3, N), and then F&O(3, N) =F&O(2, N)−Δ_3;

…

Δ_N=F&O(2, N)−F&O(N, N), and then F&O(N, N)=F&O(2, N)−Δ_N.

As F&O(1, N)=1−F&O(2, N)−F&O(3, N)−⋯−F&O(N, N), there is 1−(N−1)F&O(2, N)+Δ_3+⋯+Δ_N= F&O(2, N)+Δ_1, and Δ_1=1−N·F&O(2, N)+Δ_3+⋯+Δ_N. With the change of α, there are three possible results:

1. When α<2, the following is then true:

$\mathrm{F\&O}(2, N)<1/N$, $\varDelta_1>\varDelta_3+\cdots+\varDelta_N$, i.e., $\mathrm{F\&O}(1, N)-\mathrm{F\&O}(2, N)>\sum_{i=2}^{N}(\mathrm{F\&O}(2,N)-\mathrm{F\&Q}(i,N))$. It means that the credit difference between the first and second authors is larger than the sum of the credit differences between the second and others (Fig. 2a).

2. When $\alpha=2$, the following is then true: $\mathrm{F\&O}(2, N)=1/N$, $\varDelta_1=\varDelta_3+\cdots+\varDelta_N$, i.e., $\mathrm{F\&O}(1, N)-\mathrm{F\&O}(2,N)=\sum_{i=2}^{N}(\mathrm{F\&O}(2,N)-\mathrm{F\&Q}(i,N))$. The credit difference between the first and second authors equals the sum of the credit differences between the second and others (Fig. 2b). There are also $\varDelta_1>\varDelta_3>\varDelta_4>\varDelta_5>\cdots$.

3. When $\alpha>2$, the following is then true: $\mathrm{F\&O}(2, N)>1/N$, $\varDelta_1<\varDelta_3+\cdots+\varDelta_N$, i.e., $\mathrm{F\&O}(1, N)-\mathrm{F\&O}(2, N)<\sum_{i=2}^{N}(\mathrm{F\&O}(2,N)-\mathrm{F\&Q}(i,N))$. The credit difference between the first and second authors is less than the sum of the credit differences between the second and others (Fig. 2c).

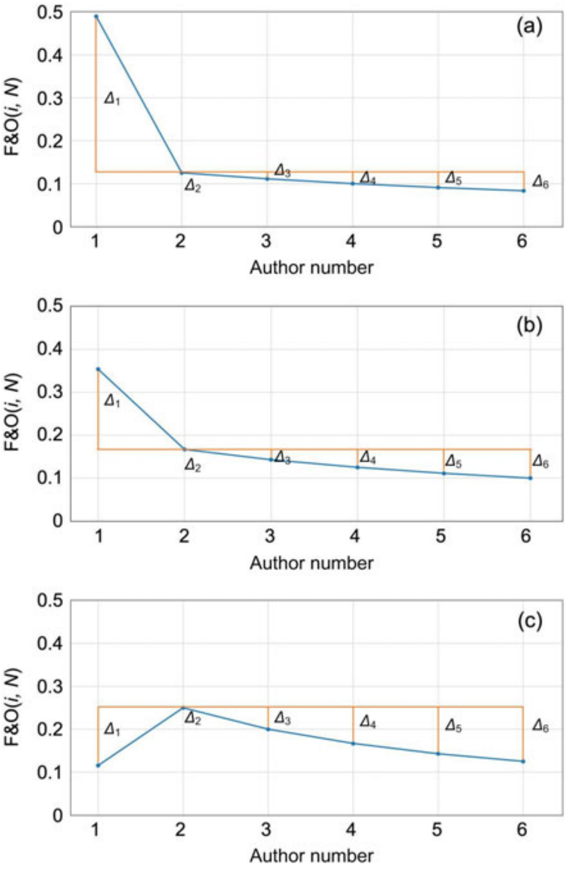

Fig. 2 F&O counting with α=0 (a), α=2 (b), and α=4 (c) (β=1 and N=6)

3.3 Ratio between credits of coauthors

The importance of distinguishing between the credits of the first author and those of others was discussed by Boxenbaum *et al.* (1987) and Ellwein *et al.* (1989). To study this relationship and property in depth, Definition 4 is proposed.

Definition 4 ρ_i is defined as the ratio between the contribution credits of authors i and i+1:

$$\rho_i=\frac{\mathrm{F\&O}(i,N)}{\mathrm{F\&O}(i+1,N)}, \tag{11}$$

where i=1, 2, …, N–1.

According to Eq. (11), ρ_i is not a constant but depends on i, N, α, β_i, and β_{i+1}. This shows the flexible property of F&O counting. In a special case, the ratio of credits between the first and second authors ρ_1 can be used to present an important feature of F&O counting. The credit of the second author becomes an inflection point (see Fig. 2 in cases where $N\geq4$). This is the basic difference between the proposed counting and existing schemas. In case of the ratio of credits between the first and second authors, there is

$$\begin{aligned}\rho_1&=\frac{\mathrm{F\&O}(1,N)}{\mathrm{F\&O}(2,N)}=\frac{\mathrm{F\&O}(2,N)+\varDelta_1}{\mathrm{F\&O}(2,N)}\\&=1+\frac{\varDelta_1}{\mathrm{F\&O}(2,N)},\end{aligned} \tag{12}$$

$$\begin{aligned}\varDelta_1&=\mathrm{F\&O}(1,N)-\mathrm{F\&O}(2,N)\\&=1-N\cdot\mathrm{F\&O}(2,N)+\varDelta_3+\varDelta_4+\cdots+\varDelta_N,\end{aligned} \tag{13}$$

where $\varDelta_i$=F&O(2, N)–F&O(i, N) (i=2, 3, ..., N).

ρ_1 is sensitive to α and β. In the case N=2, there is $\rho_1=(4-\alpha-\beta)/\beta$. Consider the following:

1. If $\alpha=(N-1)/2=1/2$ and $\beta=1$, then $\rho_1=5/2$ and F&O(2, N)=0.2857 for the second author and F&O(1, N)=1−F&O(2, N)=0.7143 for the first author.

2. If $\alpha=2.5$ and $\beta=1$, then $\rho_1=0.5<1$, F&O(2, N)=0.6667 for the second author, and F&O(1, N)= 1−F&O(2, N)=0.3333 for the first author. The credit of the first author is less than the credit of the second author. In this sense, observing Fig. 2 (N=6), in Fig. 2a, α=0, ρ_1=3.9168; in Fig. 2b, α=2, ρ_1=2.1262; in Fig. 2c, α=4, ρ_1=0.4620. It indicates that α can be used to adjust ρ_1, and then F&O counting is more flexible and robust in applications.

3.4 Generalization of Coles and harmonic countings

The Coles counting (Cole and Cole, 1973) and harmonic counting have been widely applied in the study of credit distribution for coauthors (Vinkler, 2000; Tscharntke *et al.*, 2007). Remarks 1 and 2 show that these two schemas can be obtained by the modification of α and β in F&O counting defined by Eq. (10).

Remark 1 If no coauthors claim preference contribution, i.e., $\beta=0$, F&O counting behaves as Coles counting and implements straight counting for the first author.

As in Coles counting, F&O(1, N)=1 and F&O(i, N)=0 (i=2, 3, …, N).

Remark 2 If no coauthors claim preference contribution, $\alpha=N$ and $\beta=1/(1+1/2+1/3+\cdots+1/N)$, F&O counting yields credit distribution in the same form as harmonic counting: F&O(i, N)=β/i (i=2, 3, …, N) and $F(1, N)=1-\sum_{i=2}^{N}\text{F\&O}(i,N)$.

Even though there are many formulas that estimate the contribution credits of coauthors, they are a fixed data series. F&O counting defined by Eq. (10) is a formula that illustrates the variation of the credits based on α and β.

4 *U*/*N* model for evaluating the contribution of coauthors

The author rank is evident in the byline of a publication and can be used to imply the contribution of each author. To facilitate the communication between authors and journal editors, the role of corresponding author was created (Buehring *et al.*, 2007; Hu, 2009; Hagen, 2010; Abramo *et al.*, 2012). The basic idea of weighted mean is normally used in the contribution evaluation for authors with different academic roles. This section proposes a *U*/*N* model with two methods, which were adapted from the original FCAE investigated by Trueba and Guerrero (2004), Tscharntke *et al.* (2007), and others.

4.1 Multiplication of ρ_1 to the credit of the corresponding author

If a publication has only one corresponding author and other coauthors are arranged in order, F&O counting can be modified into the *U*/*N* model by the multiplication of ρ_1 (Eq. (12)) to the original credit of the corresponding author.

$$\text{UF\&O}(i,N)=\rho_1\cdot\text{F\&O}(i,N), \tag{14}$$

where the corresponding author is in the ith position (i=2, 3, …, N), and ρ_1 is the ratio between the credits of the first and second authors. Notation of UF&O is used to make a difference to F&O. The credit of the first author will also be modified by Eq. (10), and there is UF&O(1, N)=$1-\sum_{i=2}^{N}\text{UF\&O}(i,N)$.

Taking an example of N=5 and the last author as the corresponding author, the distribution of the credits for the authors is: UF&O(2, 5)=F&O(2, 5), UF&O(3, 5)=F&O(3, 5), UF&O(4, 5)=F&O(4, 5), UF&O(5, 5)=ρ_1·F&O(5, 5), and UF&O(1, 5)=$1-\sum_{i=2}^{5}\text{UF\&O}(i,5)$ (Table 3). In the case where the credits decrease very quickly in some schemas, Eq. (14) needs to be modified with a larger ρ_i.

4.2 Mean credits of first and corresponding authors

Eq. (15) is proposed to obtain the adjusted credits UF&O(1, N) for the first author and UF&O(i, N) for the corresponding author i ($1<i\leq N$) (Hagen, 2013):

$$\begin{aligned}\text{UF\&O}(1,N)&=\text{UF\&O}(i,N)\\&=\frac{\text{F\&O}(1,N)+\text{F\&O}(2,N)}{2}.\end{aligned}\tag{15}$$

It can be proved further that UF&O(1, N)= UF&O(i, N)>F&O(j, N) ($i\neq j$, $1<j\leq N$). That is to say, the adjusted index for the first author or for the corresponding author is greater than that for any other author.

Take an example of N=5 and the last author as the corresponding author for a better understanding. Table 3 lists the distribution of the credits of the *U*/*N* model: UF&O(1, 5)=UF&O(5, 5)=(F&O(1, 5)+ F&O(2, 5))/2, UF&O(2, 5)=F&O(3, 5), UF&O(3, 5)=F&O(4, 5), and UF&O(4, 5)=F&O(5, 5). Fig. 3 shows the distribution of the UF&O model for up to 10 authors with α=2.

Table 3 Modification of F&O counting to UF&O*

Author number	F&O	UF&O	
		Eq. (14)	Eq. (15)
1	0.5672	0.3034	0.3521
2	0.1370	0.1370	0.1145
3	0.1145	0.1145	0.0973
4	0.0973	0.0973	0.0840
5	0.0840	0.3478	0.3521

* N=5 and the last author as the corresponding author

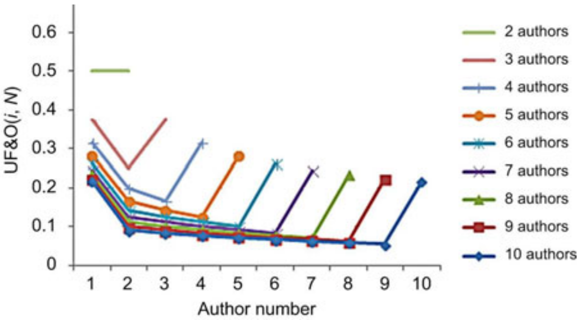

Fig. 3 Credit distribution of UF&O for 2–10 authors (α=2) (References to color refer to the online version of this figure)

5 Comparative analysis based on experimental data

This section presents the analysis of the developed models using experimental data. Lack of fit (lack_of_fit) and the coefficient of determination (R^2) are used as the evaluation criteria as in Hagen (2010; 2013).

5.1 Description of Harvard survey data

Caruso *et al.* (2006) completed a study on the contribution evaluation of the members in collaborative groups at Harvard University and the University of Chicago. The original intention of this research was to investigate the phenomenon of over claiming contribution when the group members identified their individual achievements, as well as the conflict between their self-evaluations and the evaluations made by others.

Because of added difficulties for collecting data from papers with five or more authors, Caruso *et al.* (2006) carried out a random survey based on papers with five authors. This survey received answers from 197 respondents, of which only 1 was from a fifth author and was removed from the study due to its relative sample size. Table 4 lists the distribution of 196 respondents. The total percentage of their self-focused evaluation is 149.54%, in which under the self-focused condition the contribution percentage for the first author is normalized as 39.95%, and for the second author 25.52%. As for the investigation under the other-focused condition, the total contribution percentage is 131.42%. The contribution percentage for the first author is normalized as 42.70% and for the second 24.77%. Vinkler (1993) also discussed the self-evaluation of the contribution of coauthors.

Another aspect of this present study was to ask the first author to provide a general contribution distribution for all four authors, denominated as the 'higher view', and the results were 40%, 25%, 20%, and 15%, respectively. Then the fourth author was also required to complete a general assessment, referred to as lower view, and the results were 35%, 30%, 20%, and 15%, respectively.

Table 4 Harvard survey data for coauthors

Author number	Number of authors	Contribution evaluation			
		Self-focused	Other-focused	Higher view	Lower view
1	66	0.3995	0.4270	0.40	0.35
2	68	0.2552	0.2477	0.25	0.30
3	47	0.2303	0.1927	0.20	0.20
4	15	0.1459	0.1327	0.15	0.15

Table 5 summarizes the distribution of indexes (N=4) for coauthors assuming no statement about author contribution. Several approaches are included, e.g., A-index, geometric, and harmonic. In Table 5, the indexes of the F&O model are calculated from Eq. (10) with α=1, 2 and β=1.

Table 5 Credits of coauthors from different indexes (N=4)

Author number	Credit				
	A-index	Geometric	Harmonic	F&O (α=1)	F&O (α=2)
1	0.521	0.533	0.480	0.493	0.383
2	0.271	0.267	0.240	0.200	0.250
3	0.146	0.133	0.160	0.167	0.200
4	0.063	0.067	0.120	0.143	0.167

5.2 Criterion and results of evaluation

To evaluate the proposed F&O counting and related schemas, lack_of_fit (Browne and Cudeck, 1992; Hagen, 2010; Kim and Diesner, 2014) was used as the criterion:

$$\text{lack\_of\_fit} = \frac{1}{n-1}\sum \frac{O^2 - E^2}{E}, \tag{16}$$

where n is the total number of empirical observations, and O and E are the values of empirical observation (Caruso *et al.*, 2006) and model prediction, respectively.

Table 6 details the lack_of_fits on the contribution indexes for coauthors (Table 5) based on the Harvard survey data (Table 4).

Table 6 Lack of fit between observation and prediction

Model	lack_of_fit			
	Self focused	Other focused	Higher view	Lower view
Geometric	0.0532	0.0381	0.0573	0.0682
A index	0.0520	0.0370	0.0566	0.0664
F&O (α=1, β=1)	0.0130	0.0081	0.0121	0.0324
Harmonic	0.0099	0.0047	0.0104	0.0226
Arithmetic	0.0089	0.0073	0.0111	0.0104
F&O (α=2, β=1)	0.0012	0.0041	0.0008	0.0049

When compared with the self-focused data, the results of lack_of_fit from geometric, A-index, and harmonic are 0.0532, 0.0520, and 0.0099, respectively, and F&O (α=2, β=1) obtains the best result at 0.0012. Compared with the other-focused data, the lack_of_fit from geometric, A-index, and harmonic is 0.0381, 0.0370, and 0.0047, respectively, while for F&O (α=2, β=1), it is 0.0041. In both self- and other-focused views, the lack_of_fit results from harmonic and arithmetic countings are better compared to geometric and A-index, but F&O (α=2, β=1) yields the best result in both cases. These scenarios are also observed in both higher and lower views in Table 5.

In terms of correlation analysis (Hagen, 2013), R^2 of self-focused with A-index and F&O (α=2, β=1) is 0.9972 and 0.9978, respectively, indicating that both A-index and F&O (α=2, β=1) have high correlations with the Harvard survey data.

5.3 Comparison with other schemas

Using the data from the research fields of chemistry (Vinkler, 2000), medicine (Wren *et al.*, 2007), and psychology (Maciejovsky *et al.*, 2009), the comparison of the coauthor credits among empirical data and the prediction models is obtained (Table 7). The empirical data is the same as applied in Hagen (2013) and Kim and Diesner (2014). To save space, Table 7 lists only the case of N=5 in both fields of chemistry and medicine and N=4 in the field of psychology. To compare medicine data, all predicted credits are modified in the U/N model using Eq. (15).

The prediction models include two kinds of counting: (1) flexible formula with changing parameters, such as the models in Lukovits and Vinkler (1995) and Liu and Fang (2012), and (2) flexible weight as preference for the individual coauthor, such as the schemas of Trueba and Guerrero (2004) and F&O counting. The results of other schemas such as arithmetic, geometric counting, A-index, and NBA are also reported in Section 7.

For the model of Liu and Fang (2012), when q=0.89, it is the case of the better performance mentioned by Hagen (2013); when q=1, the credits are the same as in the harmonic model. For medicine data, the credits of the authors are distributed following Eq. (15). In both cases of q=0.89 and q=1.0, R^2>0.94 was obtained for the data from three fields. From the results of lack_of_fit, it is observed that Liu and Fang's model (q=0.89) fits the data of chemistry (0.0018) and psychology (0.0024) better than fits the data of medicine (0.0109).

As the formula of Lukovits and Vinkler (1995) has only one variable parameter H, the results of lack_of_fit for cases H=10 and 21 show the same tendency as those for the other schemas. For medicine data, the credits of the authors are distributed following Eq. (15).

The approach of Trueba and Guerrero (2004) belongs to the U/N model category and uses f, c_1, c_2, and c_3 as weights to give the preference for the first (c_1), second (c_2), and last (c_3) authors. In Table 7, as proposed in the original research, f=1/3 was used. In the case where the credits of the first and second authors are with preferences, the model takes c_1=0.5, c_2=0.5, and c_3=0. For medicine data, the credits are distributed as UT&G(1, 5)=T&G(1, 5), UT&G(2, 5)=T&G(3, 5), UT&G(3, 5)=T&G(4, 5), UT&G(4, 5)=T&G(5, 5), UT&G(5, 5)=T&G(2, 5). In the case where the credits of the first, second, and third authors are with preferences, the model takes c_1=0.7, c_2=0.15, and c_3=0.15. In both cases, there is R^2>0.91 with the data from three fields. From the results of lack_of_fit,

Table 7 Comparison of the coauthor credits between the empirical data and prediction models

Subfield	Author number	Empirical	Credit							
			Liu and Fang (2012)		Lukovits and Vinkler (1995)		Trueba and Guerrero (2004)		F&O	
			q=0.89	q=1.0	H=10	H=21	f=1/3, c_1=0.5	f=1/3, c_1=0.7	α=1.5, β=1	α=1.0, β flexible
Chemistry	1	0.40	0.4089	0.4380	0.4162	0.4488	0.3296	0.3963	0.4134	0.4183
	2	0.25	0.2206	0.2190	0.2081	0.2048	0.3148	0.1982	0.1818	0.2500
	3	0.15	0.1538	0.1460	0.1503	0.1425	0.1333	0.1537	0.1538	0.1429
	4	0.10	0.1191	0.1095	0.1214	0.1113	0.1185	0.1333	0.1333	0.1000
	5	0.10	0.0976	0.0876	0.1040	0.0926	0.1037	0.1185	0.1176	0.0889
lack_of_fit			0.0018	0.0026	0.0032	0.0044	0.0084	0.0062	0.0093	0.0006
Medicine	1	0.34	0.3148	0.3285	0.3122	0.3268	0.3296	0.2972	0.2976	0.3341
	2	0.12	0.1538	0.1460	0.1503	0.1425	0.1333	0.1537	0.1538	0.1429
	3	0.08	0.1191	0.1095	0.1214	0.1113	0.1185	0.1333	0.1333	0.1000
	4	0.07	0.0976	0.0876	0.1040	0.0926	0.1037	0.1185	0.1176	0.0889
	5	0.38	0.3148	0.3285	0.3122	0.3268	0.3148	0.2972	0.2976	0.3341
lack_of_fit			0.0109	0.0061	0.0121	0.0068	0.0097	0.0195	0.0192	0.0045
Psychology	1	0.42	0.4531	0.4800	0.4747	0.5048	0.4000	0.4333	0.4421	0.4195
	2	0.24	0.2445	0.2400	0.2278	0.2212	0.2444	0.2278	0.2222	0.2400
	3	0.19	0.1704	0.1600	0.1646	0.1538	0.1556	0.1556	0.1818	0.1833
	4	0.15	0.1319	0.1200	0.1329	0.1202	0.2000	0.1833	0.1538	0.1571
lack_of_fit			0.0024	0.0069	0.0044	0.0106	0.0071	0.0049	0.0010	0.0002
R^2			0.9457	0.9431	0.9232	0.9197	0.9123	0.9109	0.9011	0.9817

it is observed that Trueba and Guerrero's model with c_1=0.5, c_2=0.5, and c_3=0 fits the data of medicine (0.0097) better than that with c_1=0.7, c_2=0.5, and c_3=0.15.

In Table 7, the performance of F&O counting is presented in two cases by changing the flexible parameters: (1) flexible formula with α=1.5 and β=1; (2) flexible weight for an individual coauthor with preference: for N=5, **BETA**=[β_2, β_3, β_4, β_5]=[1.5, 1.0, 0.8, 0.8]; for N=4, **BETA**=[β_2, β_3, β_4]=[1.2, 1.1, 1.1]. In both cases, α was set to 1. In both cases, R^2>0.90 was obtained with the data from three fields, but with flexible weights, R^2=0.9817, which is the best result over all schemas in Table 7. The results of lack_of_fit also show better results for the data of chemistry (0.0006), medicine (0.0045), and psychology (0.0002).

6 Comprehensive analysis of credit-assignment schemas

When compared with nine existing credit-assignment schemas (Table 1), F&O counting proposed in this paper presents some unique features and better performance (Tables 6 and 7). Detailed analyses are given as follows:

1. In the case of no statement about author contributions, all the formulas in Table 1 can be used as a simple and effective way to estimate the distribution of coauthor contribution. When compared against the Harvard survey data, which considers the intention of authors, and based on the analyses through lack_of_fit and R^2, harmonic and F&O (α=2 and β=1) yielded the results with better performance (Table 6). By R^2, the A-index also shows better performance.

2. In the case where a simple statement about author contributions is provided, for example, declaring the corresponding author and the author order, F&O can be weighted and appropriately adjusted so that the formula can be changed to the U/N model (see the comparison of the coauthor credits between empirical data and prediction models in Table 7).

3. The theoretical formulas of arithmetic, A-index, geometric, harmonic, and others relate only to the number of authors N and an author's order. For example, in some cases, the first author is considered to contribute more than all other coauthors; in other

scenarios, however, all authors make the same academic contribution. Even though the schemas of Lukovits and Vinkler (1995), Abbas (2010), Liu and Fang (2012), and Kim and Diesner (2014) provide flexible parameters to modify their formulas, they are still limited without preference for individual authors. The approach of Trueba and Guerrero (2004) includes weight preference, but the performance needs to be improved. F&O with tuning parameter α and weight β can be adjusted according to specific application requirements.

Fig. 4 shows the analyses of relationship between F&O distribution and the adjustment coefficient (using N=6 as an example), which indicates that F&O distribution is very sensitive to the adjustment coefficient. The range of the adjustment coefficient α in Fig. 4 is [−100, 2], and the corresponding F&O credits are in [0.0097, 0.354]. When α is within [−2, 2], the contribution index of the other coauthors does not vary significantly. In an extreme case, α=−100, the contribution of the first author is approximately 95.5%, while the contribution of all other authors combined is less than 5%. It is near Coles straight counting. This fact shows that F&O has an extensive applicability because of its flexibility of adjustment of the tuning parameters.

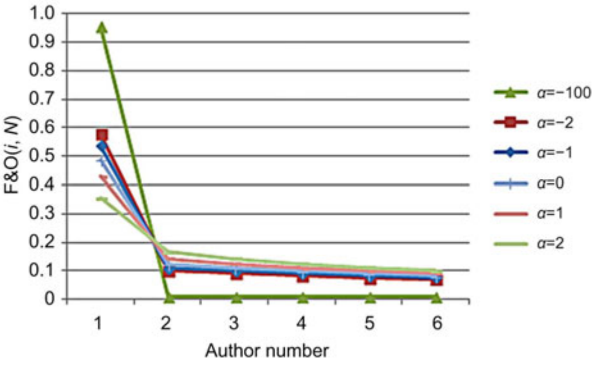

Fig. 4 Credits of F&O for coauthors and varying turning parameter (N=6) (References to color refer to the online version of this figure)

4. The importance of the formulas of SDC is to highlight the contribution of the first author and introduce decreasing contribution values for other coauthors. However, in the formulas for arithmetic, A-index, geometric, harmonic, and others, the ratio of credits between the first author and other authors is fixed or semi-fixed, which is inconsistent with the reality of many publications. Therefore, F&O along with the adjustment coefficients α and β can be used to more accurately describe the relationship between the first author and the coauthors.

5. The results of F&O show that the credit of the second author is an inflection point of the distribution function (Figs. 2 and 4). Table 8 lists the change of ρ_1, the credit ratio between the first and second authors, for publication with two or more authors (N=2–6) and varying α.

Table 8 Changing trends of ρ_1 by F&O with α=[−100, 2] and N from 2 to 6

α	ρ_1				
	N=2	N=3	N=4	N=5	N=6
−100	103.00	103.01	103.03	103.05	103.09
−2	5.00	5.13	5.31	5.53	5.77
−1	4.00	4.14	4.35	4.58	4.84
0	3.00	3.17	3.39	3.65	3.92
1	2.00	2.20	2.45	2.73	3.01
2	1.00	1.25	1.53	1.83	2.13

From Table 8, three noteworthy trends are observed: (1) When α is increased, there is a decreasing trend for ρ_1. Taking a paper with two authors as an example, the ratio changes from 103.00 (α=−100) to 1.0 (α=2). (2) With the increase in the number of coauthors in a publication, the ratio ρ_1 increases for all values of α. (3) In the extreme case of α=−100, the credit of the first author is 103 times that of the second. The ratio becomes insensitive to change in the number of authors, because the contribution of the first author is calculated to be almost 95% according to the approach of F&O. This special case is similar to the application of Coles counting (Cole and Cole, 1973).

In the discussion of the variation of ρ_1, it is useful to apply Eq. (14) to modify the credit of the corresponding author. Abbas (2010) gave more attention to the credit ratio between the first and last authors. These two studies are valuable and complementary in the research of credit-assignment schemas.

7 Conclusions

In studying relevant credit-assignment schemas (e.g., arithmetic, A-index, geometric, harmonic, and others), we cited four models (i.e., N/N, $1/N$, S/N, and

U/*N*) for evaluating coauthor contribution. We proposed a First and Others (F&O) schema and introduced the tuning parameter α and weight β to make the evaluation model more attuned to actual contribution levels. By changing α, the ratio of credits between the first author and others can be adjusted in the F&O approach. By the separation of the calculation in F&O, the credit of the second individual shows an inflection point in the distribution. Additionally, this paper applies nine existing theoretical schemas together with F&O to completely analyze the survey data provided by Caruso *et al.* (2006) and a set of empirical data form chemistry (Vinkler, 2000), medicine (Wren *et al.*, 2007), and psychology (Maciejovsky *et al.*, 2009). The analyses demonstrated the practicality and utility of the proposed new schema. Table 9 summarizes and compares credit-assignment schemas.

Currently, Google Scholar, Scopus, and other major citation databases are using *H*-index based on the *N*/*N* model to evaluate the contribution of coauthors and their academic footprint. The Nature Index uses the 1/*N* model. These methods require further adjustment and improvement for an accurate and general evaluation function, which has an important impact on the global scientific and technical research community.

The main SDC approaches (*S*/*N* models, including arithmetic, A-index, geometric, harmonic, and others proposed in the literature and F&O counting proposed in this paper) take into account the contribution index based on the author rank.

As for FCAE approaches (*U*/*N* model), they focus on the special contributions of the first author and the corresponding author. The proposed F&O and UF&O schemas can serve as references for citation databases, as they present improved properties and applicability (Table 9). The main results of the lack_of_fit and coefficient of determination (R^2) in Table 9 are selected from the comparison results with a set of empirical data in Table 7. The results of UF&O are listed in Table 3.

As Waltman (2015) argued, any new citation impact indicator should have a clear added value relative to existing schemas. As is shown in Table 9, F&O counting presents new properties in five aspects: (1) generalization of two existing schemas including straight and harmonic counting, (2) flexibility of the application by introducing both flexible formula and weight preference, (3) innovation by revealing the inflection point of the credit distribution, (4) performance improvement to reduce the over-fitting problem in the study of credit-assignment schemas, and (5) a simple F&O formula applicable for big data from the scholar information database.

Last, as Caruso *et al.* (2006) mentioned, it is difficult to evaluate the cooperation intention and contribution distribution of the team members. Different situations will result in different answers. In the case of big data, there is statistical significance when discussing different models for evaluating the coauthor contribution. As such, readers should not limit their focus to a specific method or index, and should select various methods by the reasoning of the overall evaluation instead.

Table 9 Summary and comparison of bibliometric counting

Model	Counting	Distribution method	lack_of_fit/R^2	Flexible formula	Flexible weight(s)
N/*N*	Whole	One score for each	No fitting	–	–
1/*N*	Fractional	1/*N* for each	0.1121/0.0363	–	–
S/*N*	Arithmetic	Sequence-determines-credit (SDC)	0.0200/0.8440	Yes (α)	–
	Geometric		0.0136/0.9595	–	–
	A index		0.0087/0.9346	–	–
	L&V		0.0068/0.9197	Yes (T)	–
	Harmonic		0.0061/0.9431	–	–
	NBA		0.0058/0.9102	Yes (d)	–
	F&O		0.0045/0.9817	Yes (α)	Yes (β)
U/*N*	T&G	First-corresponding-author-emphasis (FCAE)	0.0097/–	–	Yes (f, c)
	UF&O		0.0069/–	Yes (α)	Yes (β)

To advance this research, the adaptive determination of the values of the tuning parameter α and weight β needs to be studied to establish a reasonable mechanism for using the F&O approach. In the case where an article is developed with several 'first authors' and several 'corresponding authors' and other authors might not be strictly ordered (Hu, 2009), the contributions of the coauthors can be specified by the weight vector $\mathbf{BETA}=[\beta_2, \beta_3, \ldots, \beta_N]$. It may be a topic for further research to balance the counting approach and weight preference for F&O applications.

The connection of F&O as a basic formula to more advanced scientific publication mining technologies will be studied to improve its applicability and reception among the scientific community.

Acknowledgements

The author thanks the platform of SCIENCENET and blogger Shuang-chun WEN, whose blog sparked the interest in in-depth research on this topic. The author also acknowledges support from Márcio SOUZA, Ícaro DANTAS, Tian-cheng LI, Yong-he HAN, Daniel LI, and especially Ms. Ya-qin YAN for assistance with English editing.

References

Abbas, A.M., 2010. Generalized linear weights for sharing credits among multiple authors. arXiv:1012.5477.

Abbas, A.M., 2011. Weighted indices for evaluating the quality of research with multiple authorship. *Scientometrics*, **88**(1):107-131. http://dx.doi.org/10.1007/s11192-011-0389-7

Abramo, G., Cicero, T., D'Angelo, C.A., 2012. How important is choice of the scaling factor in standardizing citations? *J. Informetr.*, **6**(4):645-654. http://dx.doi.org/10.1016/j.joi.2012.07.002

Assimakis, N., Adam, M., 2010. A new author's productivity index: *p*-index. *Scientometrics*, **85**(2):415-427. http://dx.doi.org/10.1007/s11192-010-0255-z

Boas, R.P.Jr., 1964. Mathematical authorship. *Science*, **145**(3629):232. http://dx.doi.org/10.1126/science.145.3629.232

Boxenbaum, H., Pivinski, F., Ruberg, S.J., 1987. Publication rates of pharmaceutical scientists: application of the waring distribution. *Drug Metabol. Rev.*, **18**(4):553-571. http://dx.doi.org/10.3109/03602538708994132

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. *Sociol. Meth. Res.*, **21**(2):230-258. http://dx.doi.org/10.1177/0049124192021002005

Buehring, G.C., Buehring, J.E., Gerard, P.D., 2007. Lost in citation: vanishing visibility of senior authors. *Scientometrics*, **72**(3):459-468. http://dx.doi.org/10.1007/s11192-007-1762-4

Caruso, E., Epley, N., Bazerman, M.H., 2006. The costs and benefits of undoing egocentric responsibility assessments in groups. *J. Person. Soc. Psychol.*, **91**(5):857-871. http://dx.doi.org/10.1037/0022-3514.91.5.857

Cole, J.R., Cole, S., 1973. Social Stratification in Science. University of Chicago Press, Chicago.

Dong, Y., Johnson, R.A., Chawla, N.V., 2015. Will this paper increase your *h*-index? Scientific impact prediction. Proc. 8th ACM Int. Conf. on Web Search and Data Mining, p.149-158. http://dx.doi.org/10.1145/2684822.2685314

Du, J., Tang, X.L., 2013. Perceptions of author order versus contribution among researchers with different professional ranks and the potential of harmonic counts for encouraging ethical co-authorship practices. *Scientometrics*, **96**(1):277-295. http://dx.doi.org/10.1007/s11192-012-0905-4

Du, Y.P., Yao, C.Q., Li, N., 2015. Using heterogeneous patent network features to rank and discover influential inventors. *Front. Inform. Technol. Electron. Eng.*, **16**(7): 568-578. http://dx.doi.org/10.1631/FITEE.1400394

Egghe, L., Rousseau, R., van Hooydonk, G., 2000. Methods for accrediting publications to authors or countries: consequences for evaluation studies. *J. Am. Soc. Inform. Sci. Technol.*, **51**(2):145-157. http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:2<145::AID-ASI6>3.0.CO;2-9

Ellwein, L.B., Khachab, M., Waldman, R.H., 1989. Assessing research productivity: evaluating journal publication across academic departments. *Acad. Med.*, **64**(6):319-325. http://dx.doi.org/10.1097/00001888-198906000-00008

Endersby, J.W., 1996. Collaborative research in the social sciences: multiple authorship and publication credit. *Soc. Sci. Q.*, **77**(2):375-392.

Gauffriau, M., Larsen, P.O., Maye, I., *et al.*, 2007. Publication, cooperation and productivity measures in scientific research. *Scientometrics*, **73**(2):175-214. http://dx.doi.org/10.1007/s11192-007-1800-2

Hagen, N.T., 2010. Harmonic publication and citation counting: sharing authorship credit equitably—not equally, geometrically or arithmetically. *Scientometrics*, **84**(3): 785-793. http://dx.doi.org/10.1007/s11192-009-0129-4

Hagen, N.T., 2013. Harmonic coauthor credit: a parsimonious quantification of the byline hierarchy. *J. Informetr.*, **7**(4):784-791. http://dx.doi.org/10.1016/j.joi.2013.06.005

He, B., Ding, Y., Yan, E.J., 2012. Mining patterns of author orders in scientific publications. *J. Informetr.*, **6**(3): 359-367. http://dx.doi.org/10.1016/j.joi.2012.01.001

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. *PNAS*, **102**(46):16569-16572. http://dx.doi.org/10.1073/pnas.0507655102

Hodge, S.E., Greenberg, D.A., 1981. Publication credit. *Science*, **213**(4511):950-950. http://dx.doi.org/10.1126/science.213.4511.950

Hu, X., 2009. Loads of special authorship functions: linear growth in the percentage of "equal first authors" and corresponding authors. *J. Am. Soc. Inform. Sci. Technol.*, **60**(11):2378-2381. http://dx.doi.org/10.1002/asi.21164

Huang, M.H., Lin, C.S., Chen, D.Z., 2011. Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact. *J. Am. Soc. Inform. Sci. Technol.*, **62**(12):2427-2436. http://dx.doi.org/10.1002/asi.21625

Kalyane, V.L., Vidyasagar Rao, K., 1995. Quantification of credit for authorship. *ILA Bull.*, **30**(3-4):94-96.

Kim, J., Diesner, J., 2014. A network-based approach to coauthorship credit alloction. *Scientometrics*, **101**(1):587-602. http://dx.doi.org/10.1007/s11192-014-1253-3

Kim, J., Kim, J., 2015. Rethinking the comparison of coauthorship credit allocation schemes. *J. Informetr.*, **9**(3): 667-673. http://dx.doi.org/10.1016/j.joi.2015.07.005

Larsen, P.O., 2008. The state of the art in publication counting. *Scientometrics*, **77**(2):235-251. http://dx.doi.org/10.1007/s11192-007-1991-6

Lin, C.S., Huang, M.H., Chen, D.Z., 2013. The influences of counting methods on university rankings based on paper count and citation count. *J. Informetr.*, **7**(3):611-621. http://dx.doi.org/10.1016/j.joi.2013.03.007

Lindsey, D., 1980. Production and citation measures in the sociology of science: the problem of multiple authorship. *Soc. Stud. Sci.*, **10**(2):145-162. http://dx.doi.org/10.1177/030631278001000202

Liu, X.Z., Fang, H., 2012. Fairly sharing the credit of multi-authored papers and its application in the modification of *h*-index and *g*-index. *Scientometrics*, **91**(1):37-49. http://dx.doi.org/10.1007/s11192-011-0571-y

Lukovits, I., Vinkler, P., 1995. Correct credit distribution: a model for sharing credit among coauthors. *Soc. Ind. Res.*, **36**(1):91-98. http://dx.doi.org/10.1007/BF01079398

Maciejovsky, B., Budescu, D.V., Ariely, D., 2009. The researcher as a consumer of scientific publications: how do name-ordering conventions affect inferences about contribution credits? *Market. Sci.*, **28**(3):589-598. http://dx.doi.org/10.1287/mksc.1080.0406

May, M., Brody, H., 2015. Nature index 2015 global. *Nature*, **522**(7556):S1. http://dx.doi.org/10.1038/522S1a

Narin, F., 1976. Evaluative Bibliometrics: the Use of Publication and Citation Analysis in the Evaluation of Scientific Activity. Computer Horizons, Mountain Lakes, p.206-219.

Price, D.S., 1981. Multiple authorship. *Science*, **212**(4498): 986. http://dx.doi.org/10.1126/science.212.4498.986-a

Stallings, J., Vance, E., Yang, J., *et al.*, 2013. Determining scientific impact using a collaboration index. *PNAS*, **110**(24):9680-9685. http://dx.doi.org/10.1073/pnas.1220184110

Trueba, F.J., Guerrero, H., 2004. A robust formula to credit authors for their publications. *Scientometrics*, **60**(2):181-204. http://dx.doi.org/10.1023/B:SCIE.0000027792.09362.3f

Tscharntke, T., Hochberg, M.E., Rand, T.A., *et al.*, 2007. Author sequence and credit for contributions in multiauthored publications. *PLoS Biol.*, **5**(1):e18. http://dx.doi.org/10.1371/journal.pbio.0050018

van Hooydonk, G., 1997. Fractional counting of multiauthored publications: consequences for the impact of authors. *J. Am. Soc. Inform. Sci. Technol.*, **48**(10):944-945. http://dx.doi.org/10.1002/(SICI)1097-4571(199710)48:10<944::AID-ASI8>3.3.CO;2-K

Vinkler, P., 1993. Research contribution, authorship and team cooperativeness. *Scientometrics*, **26**(1):213-230. http://dx.doi.org/10.1007/BF02016801

Vinkler, P., 2000. Evaluation of the publication activity of research teams by means of scientometric indicators. *Curr. Sci.*, **79**(5):602-612.

Waltman, L., 2012. An empirical analysis of the use of alphabetical authorship in scientific publishing. *J. Informetr.*, **6**(4):700-711. http://dx.doi.org/10.1016/j.joi.2012.07.008

Waltman, L., 2015. A review of the literature on citation impact indicators. arXiv:1507.02099.

Weigang, L., Dantas, I.A., Saleh, A.A., *et al.*, 2015. Influential analysis in micro scholar social networks. Proc. 1st Int. Workshop on Social Influence Analysis, p.22-28.

Wren, J.D., Kozak, K.Z., Johnson, K.R., *et al.*, 2007. The write position—a survey of perceived contributions to papers based on byline position and number of authors. *EMBO Rep.*, **8**(11):988-991. http://dx.doi.org/10.1038/sj.embor.7401095

Xu, J., Ding, Y., Song, M., *et al.*, 2016. Author credit-assignment schemas: a comparison and analysis. *J. Assoc. Inform. Sci. Technol.*, **67**(8):1973-1989. http://dx.doi.org/10.1002/asi.23495

Yang, Y., Shan, C., Zhang, S., 2014. Counting methods and economist ranking based on ESI. *J. Intell.*, **33**(9):76-82 (in Chinese). http://dx.doi.org/10.3969/j.issn.1002-1965.2014.09.014

Zhang, C.T., 2009. A proposal for calculating weighted citations based on author rank. *EMBO Rep.*, **10**(5):416-417. http://dx.doi.org/10.1038/embor.2009.74