Modified Collaborative Coefficient: a new measure for quantifying degree of research collaboration

Kiran Savanur[1] R. Srikanth[1,2]
1 Raman Research Institute, Sadashivanagar, Bangalore 560080, India.
2 Poornaprajna Institute of Scientific Research, Bangalore 562110, India.

Abstract

Collaborative coefficient (CC) is a measure of collaboration in research, that reflects both the mean number of authors per paper as well as the proportion of multi-authored papers. Although it lies between the values 0 and 1, and is 0 for a collection of purely single-authored papers, it is not 1 for the case where all papers are maximally authored, i.e., every publication in the collection has all authors in the collection as co-authors. We propose a simple modification of CC, which we call Modified Collaboration Coefficient (or MCC, for short), which improves its performance in this respect.

1 Introduction

Collaboration is an intense form of interaction, that allows for effective communication as well as the sharing of competence and other resources: *Melin* [1]. However, the complex nature of human interaction that takes place between collaborators and the magnitude of their collaboration are not easily captured by quantutative tools. For example, the precise relationship between quantifiable activities (e.g. data analysis) and intanginble contributions (e.g. ideas) and their weightage in the final product of the collaboration (e.g. a research paper) is extremely difficult to determine. Science indicators, however, provide additional quantitative information of a more direct and objective nature to be geographical patterns of cooperation among scientific institutions: *Gupta et al* [2].

To compare the extent of collaboration in two fields (or subfields) or to show the trend towards multiple authorships in a descipline, many studies have used either the mean number of authors per paper, termed the *Collaborative Index* by *Lawani* [3] and/or the proportion of multiple-authored

papers, called *Degree of Collaboration* by *Subramanyam* [4] as a measure of the strenth of collaboration in a discipline. These two measures are shown inadequate by *Ajiferuke et al.* [5] and they derived a single meaure that incorporates some of the merits of both of the above. Collaboration Coefficient as defined by *Ajiferuke et al.*, lies between 0 and 1, with 0 corresponding to single auhored papers. However it is not 1 for the case where all papers are maximally authored, i.e. every publication in the collection has all authors in the collection as co-authors. Let the collection k be the research papers published in a discipline or in a journal during a certain period of interest. In the following, we write
f_j = the number of papers having j authors in collection k;
N = the total number of papers in k. $N = \sum j f_j$; and
A = the total number of authors in collection k.

2 Present measures

One of the early measures of degree of collaboration is Collaborative Index (CI) is given by:

$$CI = \frac{\sum_{j=1}^{A} j f_j}{N} \tag{1}$$

It is a measure of mean number of authors. Although it is easily computable, it is not easily interpretable as a degree, for it has no upper limit moreover, it gives a non-zero weight to single-authored papers, which involve no collaboration.

Degree of Collaboration (DC), a measure of proportion of multiple-authored papers is given by:

$$DC = 1 - \frac{f_1}{N} \tag{2}$$

DC is easy to calculate and easily interpretable as a degree (for it lies between zero and one), gives zero weight to single-authored papers, and always ranks higher a discipline (or period) with a higher percentage of multiple-authored papers. However, DC does not differentiate among levels of multiple authorships.

Collaboration Coefficient (CC) was designed to remove the above shortcomings pertaining to CI and DC. It is given by:

$$CC = 1 - \frac{\sum_{j=1}^{A} (1/j) f_j}{N} \tag{3}$$

It vanishes for a collection of single-authored papers, and distinguishes between single-authored, two-authored, etc., papers. However, CC fails to yeild 1 for maximal collaboration, except when number of authors is infinite. We note that DC also equals to 1 for maximal collaboration.

3 The proposed measure, MCC

The derviation of the new measure is almost the same as that of CC, as given in *Ajiferuke et al.*

Imagine that each paper carries with it a single "credit", this credit being shared among the authors. Thus if a paper has a single author, the author receives one credit; with 2 authors, each receives 1/2 credits and, in general, if we have X authors, each receives $1/X$ credits (this is the same as the idea of fractional productivity defined by *Price* and *Beaver* as the score of an author when he is assinged $1/n$ of a unit for one item for which n authors have been credited.)

Hence the average credit awarded to each author of a random paper is $E[1/X]$, a value that lies between 0 and 1. Since we wish 0 to correspond to single authorship, we define the Modified Collaborative Coefficient (MCC), κ, as:

$$\begin{aligned} \kappa &= \alpha\{1 - E[1/X]\} \\ &= \alpha\left\{1 - \sum (1/j) P(X = j)\right\} \\ &= \alpha\left\{1 - \frac{\sum_{j=1}^{A} (1/j) f_j}{N}\right\} \end{aligned} \tag{4}$$

where α is a normalization constant to be determined. Setting $\alpha = 1$ yields the measure CC. The requirement that $\kappa = 0$ for single authorship does not restrict α.

If all N articles involve all the A authors, then $E[1/X] = 1/A$. If we want κ to satisfy the requirement that $\kappa = 1$ for maximal collaboration, then we must set

$$\alpha = \left(1 - \frac{1}{A}\right)^{-1} = \frac{A}{A-1} \tag{5}$$

We thus obtain from Eqs (4) and (5) the final expression for MCC, which is:

$$
\begin{aligned}
\kappa &= \left(1 - \frac{1}{A}\right)^{-1} \{1 - E[1/X]\} \\
&= A\frac{\{1 - \sum(1/j)P(X = j)\}}{A - 1} \\
\kappa &= \frac{A}{A-1}\left\{1 - \frac{\sum_{j=1}^{A}(1/j)f_j}{N}\right\} \qquad (6)
\end{aligned}
$$

The above equation is not defined for the trivial case when $A = 1$, which is not a problem since collaboration is meaningless unless at least two authors are available. CC approaches MCC only when $A \longrightarrow \infty$, but is otherwise strictly less than MCC by the factor $\left(1 - \frac{1}{A}\right)$.

Table 1: Distribution of authorships for Library and Information Science Abstracts (reproduced from Ref. [5], except for last column)

Number of authors	1961	1966	1971	1976	1981	1986	1991
1	783	1021	1968	2771	3697	4971	0
	(94.11)	(94.28)	(86.35)	(87.06)	(83.47)	(82.88)	
2	43	48	232	312	559	786	0
	(5.17)	(4.43)	(10.18)	(9.80)	(12.62)	(13.10)	
3	6	10	54	65	123	2.83	0
	(0.72)	(0.92)	(2.37)	(2.04)	(2.78)	(170)	
4	-	3	15	23	33	36	0
	-	(0.28)	(0.66)	(0.72)	(0.75)	(0.60)	
5	-	1	8	6	8	17	0
	-	(0.09)	(0.35)	(0.19)	(0.18)	(0.28)	
6	-	-	1	5	5	10	0
	-	-	(0.04)	(0.16)	(0.11)	(0.17)	
7	-	-	0	1	4	3	0
	-	-	(0.00)	(0.03)	(0.09)	(0.05)	
8	-	-	1	-	-	-	0
	-	-	(0.04)	-	-	-	
9	-	-	-	-	-	-	0
	-	-	-	-	-	-	
10	-	-	-	-	-	-	7
	-	-	-	-	-	-	
Total	832	1083	2279	3183	4429	5998	7

4 Examples

MCC for distribution of authorships for 1966 in Table 1 is calculated thus:

$$
\begin{aligned}
\kappa &= \frac{A}{A-1}\left\{1 - \frac{\sum_{j=1}^{A}(1/j)f_j}{N}\right\} \\
&= \left(\frac{1083}{1083-1}\right)\left(1 - \left\{\frac{(1\times 1021) + (1/2\times 48) + (1/3)\times 10) + (1/4\times 3) + (1/5)\times 1)}{1083}\right\}\right) \\
&= 1.0009\left(1 - \frac{1021 + 24 + 3.333 + 1075 + 0.2}{1083}\right) \\
&= 1.0009\,(1 - 1049.283/1083)
\end{aligned}
$$

$$
\begin{aligned}
&= 1.0009\,(1 - 0.9689) \\
&= 1.0009 \times 0.0312 \\
&\simeq 0.0311
\end{aligned} \tag{7}
$$

Similarly, values of MCC for 1961, 1971, 1976, 1981, 1986 and 1991 are calculated and displayed along with the corresponding values of CI, DC and CC in Table 2.

Table 2: Measures of collaboration obtained using Eqs. 1, 2, 3 and 6. Note that MCC alone attains 1 for maximal collaboration.

Year	CI	DC	CC	MCC
1961	1.0660	0.0590	0.0306	0.0306
1966	1.0748	0.0570	0.0311	0.0311
1971	1.1880	0.1365	0.0752	0.0752
1976	1.1778	0.1294	0.0711	0.0711
1981	1.2224	0.1653	0.0904	0.0904
1986	1.2356	0.1712	0.0938	0.0938
1991	10	1	0.857	1

5 Application to some probability distributions

It is sometimes convenient if a relationship can be established between a measure of strength or inequality and a theoretical distribution which fits the observed distribution of a social phenomenon. In most cases, then, the measure can be estimated from the parameters of the distribution.

While there has been no generally accepted model for the distribution of authorships a few have been suggested: *Price and Beaver*[6] suggested the Poisson distribution while *Goffman and Waren*[7] suggested the geometric distribution. The MCC along with the other two measures is given below for these and two other commonly used probability distributions, the binomial and the negative binomial.

If the distribution variable is unbounded, then $A = \infty$, so that $\alpha = 1$ from Eq 5. In this case, MCC reduces to CC. This is the case in the following two distributions.

Geometric

$$\begin{aligned} P(X=j) &= p(1-p)^{j-1}; j=1,2,... \\ E(X) &= \sum_{j=1}^{\infty} jp(1-p)^{j-1} \\ &= 1/p \end{aligned}$$

Where p may be interpreted as the probability of completion of a research work without collaborators.

$$\begin{aligned} E(1/X) &= \sum_{j=1}^{\infty}(1/j)p(1-p)^{j-1} \\ &= -(p/(1-p))\log p \\ Hence, MCC &= 1-E[1/X] = 1+(p(1-p))\log p \end{aligned}$$

Note that $MCC \to 1$ as $p \to 0$ and
$MCC \to 0$ as $p \to 1$

Shifted Poisson

$$\begin{aligned} E[X] &= \sum_{j=1}^{\infty} je^{-y}\lambda^{j-1}/(j-1)! \\ &= \lambda+1 \end{aligned}$$

Where λ may be interpreted as the average number of colleagues consulted by a scholar before the completion of a research work.

$$\begin{aligned} E[1/X] &= \sum_{j=1}^{\infty}(1/j)e^{-y}\lambda^{j-1}/(j-1)! \\ &= (1-e^{-y})/\lambda \\ Hence, MCC &= 1-E[1/X] = 1-(1-e^{-\lambda})/\lambda \end{aligned}$$

Note that $MCC \to 0$, as $\lambda \to 0$ and
$MCC \to 1$ as $\lambda \to \infty$

Similarly, MCC is the same as CC for other distributions, like Lotka distribution, given as:

$$f(k) = \frac{c}{k^{\alpha}} \tag{8}$$

zero-truncated Poisson distribution, given as:

$$h(k) = \frac{\theta^k}{k!(e^{\theta} - 1)} \tag{9}$$

shifted negative binomical, given as:

$$P(X = j) = \binom{v+j-2}{j-1} p^v (1-p)^{j-1}; j = 1, 2, ... \tag{10}$$

and shifted inverse Gaussian-Poisson. where the number of authors is unbounded. However, when A is finite, $MCC > CC$, as illustrated for the following distribution.

Shifted binomial

$$P(X = j) = \binom{n}{j-1} p^{j-1}(1-p)^{n-(j-1)}; j = 1, 2, ..., n+1$$

where p may be interpreted as the probability of a scholar working with another colleague on a research work. n may be assumed as the greatest number of colleagues that it is possible to collaborate with within a field. For example, while it is possible for a scientist to work with as many as 100 colleagues on a research project, it is hardly conceivable for a humanist to collaborate with more than four colleagues: *Ajiferuke (1988).*

$$\begin{aligned} P(X = 1) &= (1-p)^n \\ E(X) &= \sum_{j=1}^{n+1} j\binom{n}{j-1} p^{j-1}(1-p)^{n-(j-1)} \\ &= np + 1 \end{aligned}$$

Here we set A=(n+1). Hence $\alpha > 1$, given by:

$$\alpha = \frac{A}{A-1} = \frac{n+1}{(n+1)-1} = \frac{n+1}{n} \tag{11}$$

$$
\begin{aligned}
E[1/X] &= \sum_{j=1}^{n+1}(1/j)\binom{n}{j-1}p^{j-1}(1-p)^{n-(j-1)} \\
&= \frac{1}{p(n+1)}[1-(1-p)^{n+1}]
\end{aligned}
$$

$$
\begin{aligned}
MCC &= \alpha[1-E[1/X]] \\
&= \frac{(n+1)}{n}\left(1-\frac{1-(1-p)^{n+1}}{p(n+1)}\right)
\end{aligned}
$$

Note that $MCC \to 0$, as $p \to 0$ and
$CC \to n/(n+1)$, but $MCC \to 1$, as $p \to 1$

6 Conclusion

CC is an interesting measure of collaborative strength in a discipline, that has the merit of lying between 0 and 1 (unlike previous measures of collaboration) and tends to 0 as single-authored papers dominate. Both these virtues are inherited by the new measure, MCC. However, unlike CC, which remains strictly less than 1 for finitely many authors, MCC smoothly tends to 1 as the degree of collaboration becomes maximal. This quantitatively captures our intuitive expectation that any quantification of collaborative strength must become 100 % when the collaboration is maximal.

References

[1] G. MELIN, O. PERSSON, Studying research collaboration using co-authorships, *Scientometrics* 36 (1996) 363-377.

[2] B.M. GUPTA, SURESH KUMAR, C.R. KARISIDDAPPA, Collaboration profile of theoretical population genetics speciality, *Scientometrics*, 39 (1997) 293-314.

[3] S. M. LAWANI, Quality, Collaboration and Citations in cancer reseach: A 268 bibliometric study, *Ph.D. Dissertation*, Florida State University, 1980, p395

[4] K. SUBRAMANYAM, Bibliometric studeis of research collaboration: A review, *Journal of Information Science*, 6 (1983) 33-38.

[5] I. AJIFERUKE, Q. BURREL, J. TAGUE, Collaborative coefficient: A single measure of the degree of collaboration in research, *Scientometrics* 14 (1988) 421-433.

[6] D. DE SOLLA PRICE, D. DE B. BEAVER, Collaboration in an invisible college, *American Psychologist*, 21 (1966) 1011.

[7] W. GOFFMAN, K. S. WARREN, A mathematical analysis of a medical literature: Schistosomiasis 1852-1962. In: K. CHESHIRE (Ed.), A Symposium: Information in the Health Sciences. Working to the Future, Cleveland Medical Library Association of the Cleveland Health Sciences Library, Dec. 3-4, 1969.