A novel way to quantify co-author contributions

Cagan H. Sekercioglu

Fifty years ago, McConnell argued that "for anything short of a monograph, the use of more than three authors is not justifiable"[1]. Not surprisingly, this publication was never cited, co-author numbers kept rising, and some academics now think that "multiple authorship endangers the author credit system"[2]. In the latest survey, more than 100 scientific papers published in 2006 had over 500 co-authors, and a physics paper broke the all-time record with 2512 co-authors[3]. With research groups growing larger and larger[4], the trend towards more co-authors is unlikely to end soon. Add to this the growing interest in the quantification, standardization, and automation of measuring scientific impact in the form of various citation metrics, such as the *h* index[5-7], and the growing debate on potential biases[8, 9] and unethical behavior[4, 10], the time is ripe for a standardized method to quantify co-author contributions to multi-author papers.

Any citation metric that treats all co-authors' contributions equally is inherently unfair. However, academic search engines such as Google Scholar, Scopus, and Web of Science automatically calculate citations, *h*-indices, and rankings without regard to author rank. Also considering that an increasing number of institutions (often deluged by hundreds of applicants for each position and without time to pore over individual papers) use these citation metrics to narrow the applicant pool or even hire faculty, standardized quantification of co-authors' contributions becomes critical.

Using author rank as a weighting factor is a simple, intuitive, transparent, and objective approach. Author rank may sometimes be misleading, for example when authors are ranked alphabetically or when the last author is given more credit than the second author[11]. However, such exceptions should be noted in a paper regardless and can easily be incorporated into an author contribution index, as shown below. Furthermore, rank-based standardization of co-author contributions will also motivate co-authors to clarify and rank each person's percent contribution, something that is highly desirable anyway[12].

The "single author equivalent", or the *S* index, uses author rank as a proxy for the contribution of a co-author to a paper as follows:

$$S_{\mathrm{k}} = 1/(k * H_n)$$

where k = author rank, n = number of authors, and

$$H_n = \sum_{k=1}^{n} \frac{1}{k}.$$

The k^{th} ranked co-author is considered to contribute $1/k$ as much as the first author. Co-authors' relative contributions (*S* values) always sum to one, regardless of the author number or how authors are ranked. Author rank can be different from author order, provided that this is declared in the paper. Multiple authors can have the same rank as long as this is stated and is reflected in the calculations.

For example, if the author order in a paper is A,B,C,D,E,F and the respective author rank is 1,2,2,3,4,1:

H_n = 1/1 + 1/2 + 1/2 + 1/3 + 1/4 + 1/1 = 3.583

S_A = 1/(1*3.583) = 0.279, S_B = 1/(2*3.583) = 0.140, S_C = 0.140, S_D = 0.093, S_E = 0.070, and S_F = 0.279

Percent contribution = 100 * S_k

Because *S* is dimensionless, it can be used to calculate each author's share of the citations, to calibrate an author's *h* index, and also to estimate how many single-authored publications all the papers of a scientist would approximate (S_{Total}). Hence, the single author equivalent. Some may argue that this is impossible to calculate precisely. However, as long as the same methods are applied consistently, one can compare the output of different scientists weighted by author rank. Although small differences between S_{Total} values (and corresponding citations and *h* indices) of two scientists mean little, a substantial difference, say 50% or more, is likely to indicate significantly more productivity and/or impact, even when the scientist with the greater S_{Total} has fewer papers. For example:

Scientist A has 10 papers with author ranks of 1,1,2,2,4,5,6,9,15,22.
Scientist B has 10 papers with author ranks of 1,2,4,4,5,7,8,8,9,11.

For clarity of comparison, let's assume all these papers were published during the same year and that the author's rank on each paper equals the number of authors on that paper.

	Scientist A				Scientist B			
Paper	**Rank**	**Citations**	***S***	**Weighted Citations**	**Rank**	**Citations**	***S***	**Weighted Citations**
1	1	64	1	64	1	2	1	2
2	1	44	1	44	2	1	0.333	0.33
3	2	56	0.333	18.65	4	9	0.120	1.08
4	2	35	0.333	11.66	4	8	0.120	0.96
5	4	32	0.167	5.34	5	15	0.082	1.22
6	5	5	0.082	0.41	7	12	0.055	0.66
7	6	3	0.068	0.20	8	55	0.046	2.53
8	9	0	0.039	0	8	60	0.046	2.76
9	15	1	0.020	0.02	9	130	0.039	5.11
10	22	2	0.012	0.02	11	230	0.030	6.92
Total	**67**	**242**	**3.054**	**144.3**	**59**	**522**	**1.871**	**23.57**
	Apparent *h* = 5		**Weighted *h* = 5**		**Apparent *h* = 8**		**Weighted *h* = 2**	

A cursory look suggests that scientist B, who has a higher average author rank, has 2.2 times more citations and a noticeably higher *h* index than scientist A. However, when author rank is considered for each paper, the *S* values are calculated, and then used to weight the citations, the picture is reversed. In fact, A has 63% more “single author equivalent” papers as B. More co-authors on a paper also increase the likelihood of future self-citation, a bias that is offset by distributing the citations among co-authors.

Comparing two people can be hard enough. Considering that many committees, with limited time to choose from dozens or hundreds of applicants, cannot evaluate each applicant’s individual papers and often rely on numbers reported by the applicants, the usefulness of such standardization becomes clear. In addition, quantifying co-authors’ contributions will encourage a healthy dialogue about the meaning of co-authorship and author rank[2, 4, 11], will promote better consideration of author rank in assessing scientific impact, and will lead to improved ways to measure and report co-author contributions, ideally by using percentages.

Most scientists are not happy about their life’s work being reduced to an index or two. Unfortunately, the rapid rise in the numbers of academics, papers, co-authors, and applicants, combined with a worldwide interest in quantifying scientific impact, mean that such indices are here to stay. As scientists, we should play an active role in making sure that these indices are relevant, rigorous, and most importantly, fair.

References

1. McConnell, D. Too many authors. Science 128, 1157-1158 (1958).
2. Greene, M. The demise of the lone author. Nature 450, 1165-1165 (2007).
3. King, C. Multiauthor papers redux: a new peek at new peaks. Science Watch 18, 1-2 (2007). URL: http://www.sciencewatch.com/nov-dec2007/sw_nov-dec2007_page1.htm
4. Lawrence, P. A. The mismeasurement of science. Current Biology 17, R583-R585 (2007).
5. Hirsch, J. E. Does the h index have predictive power? Proceedings of the National Academy of Sciences of the United States of America 104, 19193-19198 (2007).
6. Ball, P. Achievement index climbs the ranks. Nature 448, 737-737 (2007).
7. Lehmann, S., Jackson, A. D. & Lautrup, B. E. Measures for measures. Nature 444, 1003-1004 (2006).
8. Kelly, C. D. & Jennions, M. D. H-index: age and sex make it unreliable. Nature 449, 403-403 (2007).
9. Mishra, D. C. Citations: rankings weigh against developing nations. Nature 451, 244-244 (2008).
10. Todd, P. A. & Ladle, R. J. Citations: poor practices by authors reduce their value. Nature 451, 244-244 (2008).
11. Laurance, W. F. Second thoughts on who goes where in author lists. Nature 442, 26-26 (2006).
12. Anonymous. Who is accountable? Nature 450, 1-1 (2007).