Measuring co-authors' contribution to an article's visibility

Reference:
Egghe Leo, Guns Raf, Rousseau Ronald.- Measuring co-authors' contribution to an article's visibility
Scientometrics: an international journal for all quantitative aspects of the science of science and science policy- ISSN 0138-9130 - 95:1(2013), p. 55-67
Full text (Publisher's DOI): https://doi.org/10.1007/S11192-012-0832-4
To cite this reference: https://hdl.handle.net/10067/1082680151162165141

Measuring co-authors' contribution to an article's visibility

Leo Egghe[1,2], Raf Guns[2] and Ronald Rousseau [2,3,4]

[1] Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan,
B-3590 Diepenbeek, Belgium
E-mail: leo.egghe@uhasselt.be

[2] Universiteit Antwerpen (UA), IBW, Stadscampus, Venusstraat 35,
B-2000 Antwerpen, Belgium
E-mail: raf.guns@ua.ac.be

[3] KHBO (Association KU Leuven), Fac. of Engineering Technology, Zeedijk 101,
B-8400 Oostende, Belgium
E-mail: ronald.rousseau@khbo.be

[4] KU Leuven, Dept. Mathematics,
Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium

Abstract

The visibility of an article depends to a large extent on its authors. We study the question how each co-author's relative contribution to the visibility of the article can be determined and quantified using an indicator, referring to such an indicator as a CAV-indicator. A two-step procedure is elaborated, whereby one first chooses an indicator (e.g. total number of citations, h-index…) and subsequently one of two possible approaches. The case where the indicator is an h-type index is elaborated in a Lotkaian framework. Different examples illustrate the procedure and the choices involved in determining a CAV-indicator.

Keywords: visibility indicators ; power law model

Introduction

The question "What makes an article influential?" has been studied by Van Dalen and Henkens (2001) and seven years later by Haslam et al. (2009). These two teams consider subfields of the social sciences in order to differentiate with the, more codified, natural sciences. Haslam et al. (2009) found that some factors, such as institutional prestige, did not predict citation impact, whereas other factors did. These factors include eminence of the first author and seniority of a later (second, third…) author. Van Dalen and Henkens operationalize the reputation of an article as the reputation of the author with the best reputation, where an author's reputation is determined by career citation counts. This variable turns out to have a significant effect on the number of received citations of new articles.

In summary, the visibility of a paper is to a large extent determined by the paper's author(s). Moreover, in most research fields the majority of publications has more than one author. These observations raise the question if and how one can determine the relative contribution of a paper's co-authors to its visibility. As a first approximation, one might choose to determine each author's relative contribution to the contents of the article itself, for instance according to a co-authorship scoring system like the one proposed by Hunt (1991). This approach is, however, certainly not always adequate. Consider, for instance, the case of an article authored by a PhD student and his/her advisor. While the student may have contributed the bulk of the article's content, the advisor is most likely better known and therefore contributes more to the article's visibility.

Are there ways to measure a co-author's contribution to the visibility of an article? Maybe it is possible to come up with an acceptable estimate and propose an indicator for an author's contribution to an article's visibility. We will call such an indicator a CAV-indicator (where CAV stands for Contribution to an Article's Visibility). In this contribution we provide several proposals for such a CAV-indicator and study some in more detail within the power law model.

Literature review

According to survey results, over one-third of Chinese physicists considered that the prestige of an author increased the chance that they would cite his article (Liu, 1993). Skilton (2009) studies if the so-called human capital of teams in the natural sciences can predict citation frequency. He operationalizes the notion of human capital by reputation, measured by the h-index (Hirsch, 2005). He then points out that teams that include one or more scientists of reputation are more likely to produce interesting research and consequently be cited more. All this is of course related to 'interestingness' as the, or at least an important, reason for citing (Liu, 2011), to the Matthew effect (Merton, 1968), and to the relation between reputation and being the default choice (Mahbuba and Rousseau, 2011). The Matthew effect is confirmed by Tol (2009) in a study of the 100 most prolific economists: the number of citations to an author's publications is positively affected by that author's previous number of citations. Likewise, in a study of ecological articles, Leimu and Koricheva (2005) find that "social factors, such as the professional standing of the cited author, play a significant role in citation decisions".

Skilton (2009) finds that articles co-authored by diverse teams and by authors of high quality are cited more frequently than other articles. Yet, he also finds that social capital factors such as institutional prestige and national affiliation (for instance having a USA affiliation) have no effect. Best results, in terms of received citations, go to articles whose co-authors exhibit greater disciplinary diversity. His detailed analysis leads to the detection of a flow of effects from intellectual capital to team processes to article quality and then to received citations, rather than a direct halo effect.

Linton et al. (2011) examine the relation between academic reputation and institutions' rank according to Quacquarelli Symonds and the Webometrics Ranking of World Universities. Among several results they find that there is a statistically significant relation between the quantity of research by the most prolific author in a field and the institution ranking; a similar relation holds for the number of citations received by the most prolific author and also for their h-index.

Wang et al. (2011) use data mining techniques to discover typical features of highly cited papers. They identify 11 distinctive features of highly cited papers, including the h-index of the first author and the highest h-index of all authors. The authors argue that two mechanisms are at play: a paper's intrinsic quality on the one hand and its visibility on the other. A paper by a well-known author has more visibility and thereby a greater chance of becoming highly cited.

If a well-known author attracts more citations, this may (over time) also yield benefits for his/her co-authors; they receive more citations only because they have a well-known co-author. To correct for this effect, Hirsch (2010) introduces the $\bar{h}$ ('hbar') index as a modification of the original h-index: "for two equally good papers authored by a young researcher, the one with better known senior coauthors is likely to garner more citations simply because of name recognition, thus it is reasonable that $\bar{h}$ counterbalances this effect."

Based on this literature review we conclude that characteristics of authors – such as the number of published articles or highly-cited articles, being a doctoral student or principal investigator, the language used – influence the visibility of articles (along with other factors, such as an article's inherent quality). In the following sections we propose a framework to determine the relative contribution of co-authors to an article's visibility. Since there is no unambiguous way to do this, many alternatives can be considered. In this contribution we do not attempt to find the 'best' alternative.

Methods

Consider an article written by more than one scientist and assume that one wants to determine a number representing how each co-author contributes to the visibility of this article. As a first approximation, the contribution to visibility could be estimated as equal to the contribution to the article's content. In some cases this information may be provided in the article itself, albeit in a descriptive way. Indeed, some journals such as PLoS ONE, even have guidelines requesting that this information be provided. However, since this information is often lacking (as is usually the case in our field), we consider alternative ways of determining each co-author's contribution to the visibility of the article. One reasonable approach to go beyond pure contribution to content – but certainly not the only one – is to assume that an author's contribution to visibility depends on his/her past performance.

We outline a two-step procedure to obtain a CAV-indicator of a co-author. In a first step an indicator must be chosen. In a second step one of two possible approaches is chosen. Both steps are explained in the following sections.

First step: determining an indicator

The aim of the first step is to choose an indicator. A list of some possible choices is given in Table 1.

Indicators are calculated for a set of articles. Such sets, including possibly a singleton set, will play an essential role. Using a clearly defined procedure one determines an indicator score associated with this set of articles. At this point this procedure and the resulting score can be almost anything, hence we provide a list of examples. Clearly many other cases can easily be added to Table 1.

Table 1. Examples of sets of articles, procedures and calculated scores

	Set of articles	Procedure	Indicator
1	One particular article, namely the article under consideration	Counting the authors	Number of authors
2	One particular article, namely the article under consideration	Using an accepted (Hunt-type) scheme to determine a scientist's contribution	Contribution score
3	Articles co-authored by scientist S during a given period	Counting (total counting or fractional counting)	Counting score. In case of total counting this yields the number of articles in the set
4	Articles co-authored by scientist S during a given period	Using an accepted scheme (e.g. Hunt, 1991) to determine a scientist's contribution	Average or total contribution score
5	Articles co-authored by scientist S during a given period	Determining citations received during a given period from a given database and counting these	Number of received citations
6	Articles co-authored by scientist S during a given period	Determining non self-citations received during a given period from a given database and counting these	Number of received non self-citations
7	Articles co-authored by scientist S during a given period	Determining citations received during a given period from a given database and applying the procedure to determine the h-index	h-index (Hirsch, 2005)
8	Articles co-authored by scientist S during a given period	Determining citations received during a given period from a given database and applying the procedure to determine the g-index, or the R-index	g-index (Egghe, 2006a, 2006b) or R-index (Jin et al., 2007)

9	Articles co-authored by scientist S during a given period	Determining citations received during a given period from a given database and calculating the average number of citations received	Average number of received citations

Second step: two different approaches

In the second step, one chooses an approach to turn the indicator(s) of the first step into a score that indicates a co-author's contribution to article visibility.

First approach

With each collaborator one associates a set of articles as suggested in the second column of Table 1 and a corresponding score is determined. The assumed contribution of each collaborator to the article's visibility (his CAV-indicator value) is then determined as the weighted average of each author's score. Let us consider an example where the h-index is used as indicator (case 7 in Table 1). Now, if four scientists have collaborated on an article and these scientists have an h-index of respectively, 10, 20, 5 and 5 then their CAV-indicator scores are assumed to be: 0.25, 0.5, 0.125 and 0.125 respectively. Using the number of co-authors of the article under consideration as indicator (leading to a simple fractional score, case 1 in Table 1), each collaborator receives a score of 0.25. In the exceptional case that all co-authors have an indicator score of zero then each receives the same contribution score equal to $1/n$, where n is the number of co-authors.

Second approach

First a set of articles is associated with each collaborator and a score according to a chosen procedure is determined.

Next one forms the union of these authors' articles (those used in the first step). This union is the set to which the same procedure is applied, leading to a meta-score. A collaborator's CAV-indicator is then defined as the ratio of his or her score divided by the meta-score.

If the indicator is an h-, g- or R-index then an author's contribution is at most one, but the sum of all contributions is usually larger than 1. We give a simple example. Suppose we have three authors with an h-index of respectively 10, 5 and 3 and their meta-h-index is 15. Their scores then become: 0.66, 0.33 and 0.2. The sum of all scores is 1.2. Only when the meta-score equals the sum of individual scores do we obtain a sum of contributions equal to 1. If the number of authors of the article under consideration is used as basic indicator, the second approach leads to the same visibility score as the first one.

If the average number of citations (case 9 in Table 1) is used as indicator then one collaborator's contribution to the article's visibility may already be higher than one, and hence the sum of all contributions may certainly be larger than one.

Since the relative ratios of the two approaches are the same, rescaling the results of the second approach would yield the same result as the first approach. Still, the two approaches are not the same under all circumstances. This is explained further on.

A model for the case of h-type indices in a Lotkaian framework

In this theoretical section, we assume that Lotka's law holds for the distribution of citation numbers for articles. We do not claim that this is always the case for empirical data but we use this distribution for modeling purposes. Here the density of publications with citation density *j* is given by

$$f(j) = \frac{B}{j^{\alpha}} \tag{1}$$

with $B > 0$, $\alpha > 1$, $j \geq 1$.

Although we discuss several indicators within this model, special attention will be paid to the h-index and h-type indices.

First approach

We first consider the first method with the h-index as indicator. In the Lotkaian model the h-index of co-author *j* can be written as (Egghe & Rousseau, 2006):

$$h_j = T_j^{1/\alpha_j} \tag{1}$$

where T_j denotes the number of articles written by author *j* and α_j (> 1) is the Lotka-exponent for co-author *j*'s citation distribution. Then co-author *j*'s contribution score, assuming a total of *n* co-authors, is:

$$c_j = \frac{T_j^{1/\alpha_j}}{\sum_{k=1}^{n} T_k^{1/\alpha_k}} \tag{2}$$

If all α_j are assumed to be equal (to α) then the contribution becomes:

$$c_j = \frac{T_j^{1/\alpha}}{\sum_{k=1}^{n} T_k^{1/\alpha}} \tag{3}$$

As an example, within this model, we consider $n = 2$, assume that $\alpha = 2$, and that author 1 has written twice as much articles as author 2: $T_1 = 2T_2$. Then these authors' relative CAV-indicators are:

$$c_1 = \frac{\sqrt{2}}{\sqrt{2}+1} \approx 0.586 \quad \text{and } c_2 = \frac{1}{\sqrt{2}+1} \approx 0.414 \tag{4}$$

In this theoretical example the first author contributes more to the visibility, but certainly not the double amount. This fact is related to the robustness of the h-index.

What happens if the g-index or the R-index is chosen as indicator? In this model and assuming that all α_j-values are equal and larger than 2, the contribution scores are the same as those shown in equations (3). Indeed, Egghe (2006b) has shown that, under these assumptions

$$g = \left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{(\alpha-1)}{\alpha}} h \tag{5}$$

For the R-index (Jin et al., 2007) we have:

$$R = \sqrt{\frac{\alpha-1}{\alpha-2}}\, h \tag{6}$$

However, when using the number of received citations as indicator we find another weighting system. Indeed, Egghe (2005, p. 115) has shown that, still in the Lotkaian model, and assuming again that all α_j-values are equal and larger than 2, that the average production is:

$$\mu = \frac{\alpha-1}{\alpha-2} \tag{7}$$

Hence, as the average production is here the number of citations, denoted as *Cit*, divided by the number of articles (*T*), we have:

$$Cit = \frac{\alpha-1}{\alpha-2}\, T \tag{8}$$

Then c_j becomes:

$$c_j = \frac{Cit_j}{\sum_{k=1}^{n} Cit_k} = \frac{T_j}{\sum_{k=1}^{n} T_k} \tag{9}$$

Equation (9) shows that in the Lotkaian case it does not matter if one uses the number of articles or the number of citations as indicator. This is rather remarkable as it is generally assumed that numbers of citations are more skewed than numbers of publications (Stephan & Levin, 1991; Rousseau, 1992). Such a skewness is an application of the general principle that there is more inequality in use than in availability (Rousseau, 1992).

Hence, using the average as an indicator leads to each author receiving the same contribution score.

$$c_j = \frac{\alpha-1}{\alpha-2} \frac{1}{\sum_{k=1}^{n} \frac{\alpha-1}{\alpha-2}} = \frac{1}{n} \tag{10}$$

We finally observe that, if all α are equal, in the limiting cases $\alpha \to 1$ and $\alpha \to \infty$ formula (3) becomes:

$$c_j = \frac{T_j}{\sum_{k=1}^{n} T_k} \qquad \text{and} \qquad c_j = \frac{1}{n} \tag{11}$$

These are the cases that the number of articles or the number of authors of the current article are used as indicator.

Second approach

In this approach we consider a "meta-author" with $\sum_{k=1}^{n} T_k$ articles, where we assume that co-authors have not yet collaborated before. Applying the formula for the h-index in a Lotkaian system we obtain:

$$h_M = \left(\sum_{k=1}^{n} T_k\right)^{1/\alpha} \tag{12}$$

Using equation (1) this expression can be rewritten as:

$$h_M = \left(\sum_{k=1}^{n} h_k^{\alpha}\right)^{1/\alpha} \tag{13}$$

where we have used, for simplicity, the same alpha-value for each author and for the meta-author. Now we want to compare the impact of each co-author with the impact of this meta-author. The contribution score of co-author j for the second approach is denoted by v_j. We have:

$$v_j = \frac{h_j}{h_M} = \frac{h_j}{\left(\sum_{k=1}^{n} h_k^{\alpha}\right)^{1/\alpha}} \tag{14}$$

or, by (1):

$$v_j = \left(\frac{T_j}{\left(\sum_{k=1}^{n} T_k\right)}\right)^{1/\alpha} \tag{15}$$

We note that, since $\alpha > 1$, Jensen's inequality (Hardy, Littlewood, Pólya, Theorem 19) shows that for each $j = 1, \ldots, n$: $v_j > c_j$.

Hence, $\sum_{k=1}^{n} v_k > \sum_{k=1}^{n} c_k = 1$. With this method the total score of an article when the h-index is used as indicator is larger than one. Taking the limits for $\alpha \to 1$ and $\alpha \to \infty$ gives $\sum_{k=1}^{n} v_k = 1$ and $\sum_{k=1}^{n} v_k = n$, respectively.

In view of these results we may say that the first approach is a weighted fractional method (coinciding with the first limiting case for the second approach), while the second approach is a kind of weighted total counting (coinciding with standard total counting in the second limiting case).

The difference $(\sum_{k=1}^{n} v_k) - 1$ can be considered as the added value with respect to the case that the authors are considered as one meta-author. The results in the second approach can be interpreted by stating that the article itself indicates what method of counting should be used, namely somewhere in-between fractional and total counting. This is quite remarkable, but recall that we work within the Lotkaian model.

Using equation (5) we have:

$$g_M = \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{(\alpha-1)}{\alpha}} h_M$$

$$= \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\frac{(\alpha-1)}{\alpha}} \left(\sum_{k=1}^{n} h_k^{\alpha}\right)^{1/\alpha}$$

$$= \left(\sum_{k=1}^{n} \left(\frac{\alpha - 1}{\alpha - 2}\right)^{\alpha-1} h_k^{\alpha}\right)^{1/\alpha}$$

hence: $g_M = (\sum_{k=1}^{n} g_k^{\alpha})^{1/\alpha}$ (16)

Similarly, using equation (6):

$$R_M = \sqrt{\frac{\alpha - 1}{\alpha - 2}}\ h_M$$

$$= \left(\left(\frac{\alpha - 1}{\alpha - 2}\right)^{\alpha/2} \left(\sum_{k=1}^{n} h_k^{\alpha}\right)\right)^{1/\alpha}$$

$R_M = (\sum_{k=1}^{n} R_k^{\alpha})^{1/\alpha}$ (17)

These are analogous forms of (13) for the g- and the R-index. By equations (5) and (6) we see that using the g-index or the R-index leads to the same weighting system as for the h-index. Also using the number of citations leads to the same weighting system (in the Lotkaian model) as using the number of articles:

$$\left(\frac{Cit_j}{\sum_{k=1}^{n} Cit_k}\right) = \left(\frac{\frac{\alpha - 1}{\alpha - 2} T_j}{\frac{\alpha - 1}{\alpha - 2} \sum_{k=1}^{n} T_k}\right) = \left(\frac{T_j}{\sum_{k=1}^{n} T_k}\right)$$

Artificial example

We first consider an artificial example to clarify the difference between the approaches. Suppose that we have an article jointly written by two authors A and B. Both authors have previously authored eight articles with the following citation scores: for A 9 – 5 – 4 – 4 – 1 – 0 – 0 – 0 and for B 10 – 7 – 7 – 6 – 5 – 2 – 1 – 0. We have the following indicator values: $T_A = T_B$ = 8, C_A = 23, C_B = 38, h_A = 4, h_B = 5, g_A = 4, g_B = 6.

Both authors have the same number of publications. Hence, if we use this indicator, their contribution score is 0.5 for both approaches assuming that they have not previously collaborated.

What happens if we choose the total number of citations as an indicator? Again, we obtain the same result for the two approaches. Author A's score now becomes 0.38 and author B's score becomes 0.62, reflecting the latter's higher citation scores.

If the h-index is chosen as an indicator, the two approaches yield different results. Using the first approach, author A gets a score of 0.44, whereas B's score is 0.56. For the second approach, we take the union of both authors' articles, leading to a meta-h-index of 5. In this case, author A's CAV-indicator value becomes 0.8 and B's becomes 1. Although B's citation scores are clearly higher, the difference between the two is actually quite small, due to the robustness of the h-index.

Because the g-index can better discriminate between different author profiles, it also yields a larger difference in the resulting CAV-indicator values. We obtain 0.4 and 0.6 for authors A and B respectively using the first approach. For the second approach, we find a meta-g-index of 7, leading to a contribution score of 0.57 for A and 0.86 for B.

Let us suppose now that A and B have one previous article in common: the article with 5 citations, second in A's list and fifth in B's list. This does not affect the first approach. For the second approach, however, it can affect the meta-indicator and thereby the individual authors' CAV-indicator. One can see that the results for the following indicators now change:

- number of publications: $v_A = v_B =$ 8/15
- number of citations: $v_A \approx$ 0.41, $v_B \approx$ 0.68
- g-index: $v_A \approx$ 0.67, $v_B = 1$

Empirical examples

Because it is in practice difficult to determine a person's h-index at some previous point in time, we will use known data for two of the current authors. We take the following article as an example:

Jin, BH., Rousseau, R. (2001). An introduction to the barycentre method with an application to China's mean centre of publication. *Libri*, 51(4), 225-233.

In the following discussion, we restrict ourselves to articles (and citations) that are indexed in Thomson Reuters' Web of Science (all databases).

At the time of publication of the aforementioned article, Jin had 3 publications whereas Rousseau had 87. As one might expect, their h-indices were quite different as well: Jin's h-index was equal to 1, whereas Rousseau's was 8 (Rousseau & Jin, 2008). Applying the first approach and using the h-index as indicator, we find that Jin's contribution to the article's visibility is 0.11, while Rousseau's contribution is 0.89. For the second approach, we need the two authors' meta-h-index. In this case, the meta-score is 8 (Jin's article contributing to her h-index had 5 citations and is therefore not part of the h-core). This method yields a contribution score of 0.125 for Jin and 1 for Rousseau. Note that the sum of contributions again exceeds one.

Another example with two authors is formed by the following article:

Egghe, L., Waltman, L. (2011). Relations between the shape of a size-frequency distribution and the shape of a rank-frequency distribution. Information Processing & Management, 47(2), 238-245.

The number of citations of the two authors (excluding 2012) are 1883 for Egghe and 171 for Waltman. If we choose total number of citations as an indicator, the contribution scores become 0.92 for Egghe and 0.08 for Waltman.

When one or several doctoral students publish their first article in collaboration with their thesis supervisor and an indicator is chosen which is based on earlier publications (cases 3-8[1] in Table 1), then their CAV-indicator values are zero, while their supervisor's is one. This was the case for RR's first publication with another doctoral student and his supervisor Prof. Van Daele:

Rousseau, R., Van Daele, A., Vanheeswijck, L. (1976). On the commutation theorem for tensor products of Von Neumann algebras. *Proceedings of the American Mathematical Society*, 61(1), 179-180.

Of course, if visibility is determined based on the number of co-authors, then each CAV-indicator value is 1/3.

As a last empirical example, we consider the following article:

Liu, YX., Rafols, I., Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, 68(1), 31–44.

---

[1] Case 9: using the average number of citations is undefined if an author has no articles

Interestingly, this article is authored by three researchers of different seniority as reflected by their numbers of publications (citations): Liu: 6 (46); Rafols: 27 (535); Rousseau: 186 (1973). Their h-indices are, respectively, 4, 13, and 24. Applying the first approach, we find the following CAV-indicator values:

- number of publications: Liu: 0.03, Rafols:0.12, Rousseau: 0.85
- number of citations: Liu: 0.02, Rafols: 0.21, Rousseau: 0.77
- average number of citations: Liu: 0.20, Rafols: 0.52, Rousseau: 0.28
- h-index: Liu: 0.10, Rafols: 0.32, Rousseau: 0.59

This again illustrates the well-known fact that the choice of indicator is crucial.

For the second approach, we first determine the meta-indicators for the union of the three authors. These are: 214 (number of publications), 2509 (number of citations), 11.72 (average number of citations), and 26 (h-index). Hence, applying the second approach, we find the following CAV-indicator values:

- number of publications: Liu: 0.03, Rafols: 0.13, Rousseau: 0.87
- number of citations: Liu: 0.02, Rafols: 0.21, Rousseau: 0.79
- average number of citations: Liu: 0.65, Rafols: 1.70, Rousseau: 0.90
- h-index: Liu: 0.15, Rafols: 0.50, Rousseau: 0.92

Contribution to a number of articles

One author may of course contribute to several papers. By adding up the scores obtained for individual papers, we can determine one author's contribution to the visibility of a group of papers (e.g. his/her entire oeuvre). Doing so can also lead to different relative results for the first and second approach. We show this in the following simple example with two authors A and B, using the h-index as indicator. We look at two papers for A and two for B.

First, we have a paper co-authored by A and B. Both A and B already have a publication record of three papers with the following citation scores:

rank	A	B
1	5	6
2	2	4
3	0	2

Hence we have: h_A = 2; h_B = 2; h_M = 3. According to the first approach, $c_A = c_B = 0.5$. According to the second approach, $v_A = v_B = 0.67$.

Next, A publishes a paper co-authored together with C. The citation record of A and C is the following:

rank	A	C
1	5	10
2	2	10
3	0	10
4		3

We have the following h-indices: h_A = 2; h_C = 3; h_M = 4. According to the first approach, c_A = 0.4 and c_C= 0.6. According to the second approach, v_A = 0.5 and v_C = 0.75.

Finally, B publishes a paper, co-authored with D. The citation record of B and D is the following:

rank	B	D
1	6	3
2	4	3
3	2	3
4		0

We have the following h-indices: h_B = 2; h_D = 3; h_M = 3. According to the first approach, c_B = 0.4 and c_D = 0.6. According to the second approach, v_B = 0.67 and v_D = 1.

We can now determine the total contribution of A and B to the papers they have co-authored. According to the first approach, the total contribution of A is 0.5 + 0.4 = 0.9 and the total contribution of B is 0.5 + 0.4 = 0.9. According to the second approach, A's total contribution is 0.67 + 0.5 = 1.17 and B's total contribution equals 1.33. Whereas the first approach leads to a tie, the second approach ranks B higher than A. This illustrates that the two approaches are not interchangeable.

A retrospective view

In Table 1 articles and citations are considered over a certain period. Normally one assume that this period ends just before the article of which one wants to estimate relative contributions to visibility, is finished or published. Yet, one might consider periods extending beyond the publication date. This would lead to an adaptive form of visibility weights. One could then say that in retrospect the junior author (maybe the doctoral student) was most likely the main contributor to an article that turned out later to be ground-breaking. Surely, if a future Nobel Prize winner contributes then this article gains, in retrospect, a lot of visibility.

As an example we reconsider the Jin-Rousseau paper. The previous results were obtained taking a 2001 perspective. We can, however, also adopt a contemporary perspective and take later (2001–2012) articles and citations into account as well. Now we find an h-index of 6 for Jin and of 24 for Rousseau (April 12, 2012). The first approach leads to a contribution of 0.2 for Jin and 0.8 for Rousseau. We find a meta-h-index of 24. The highest cited article is authored by (among others) both Jin and Rousseau; the rest of the h-core consist of 22 articles by Rousseau and 1 article by Jin. Hence, according to the second approach we have a contribution of 0.25 for Jin and 1 for Rousseau. We can conclude that in this example the unevenness of the two authors' contribution has decreased over the course of the years.

Comments

As we have already explained, an author's contribution to the article's contents is not necessarily the same as his contribution to an article's visibility. This may even lead to the addition of authors who do not fulfill the normal criteria for authorship, only "to help the paper appear more legitimate" (Claxton, 2005). According to the International Committee of Medical Journal Editors (ICMJE, 2010) an author is a person who contributes to each of the following steps:

1) has substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
2) drafting the article or revising it critically for important intellectual content;
3) gives final approval of the version to be published.

Yet, these rules are not very helpful in determining a co-author's contribution. How much is "giving final approval" worth? And what is "a substantial contribution"? Is someone who does not collect data not a real co-author? For this reason schemes have been proposed to score a person's contribution, Hunt's (Hunt, 1991) being one of the oldest and best-known. It was designed to decide if a person deserves the status of co-author or not, but it can equally be used to assign a contributorship score. At the same time it is well-known that sometimes one 'sacred spark' makes the difference between a simple run-of-the-mill article and an article with lasting influence.

The proposed visibility indices are calculated based on sets of articles attached in some way to a scientist. This methodology is different from Tol's (Tol, 2012) who determines pseudo-Shapley values, power and market value of a scientist with respect to his/her scientific environment. Moreover, we obtain indicators to compare co-authors, while Tol obtains indicators to compare colleagues in the same institute or school.

A reviewer suggested to work out a project to determine a relation between an independent, external measure of contribution to article visibility and some of the CAV-indicators. This would not only confirm the usefulness of our indicators, but would lead to the determination of the 'best' among the many CAV-indicators proposed.

Authorship to this contribution was determined by alphabetical order, but we leave it to the reader to estimate each contributor's 'real' CAV-indicators.

References

Claxton, L.D. (2005). Scientific authorship: Part 2. History, recurring issues, practices, and guidelines. *Mutation Research/Reviews in Mutation Research*, 589(1), 31–45.

Egghe, L. (2005). Power laws in the information production process: Lotkaian Informetrics. Amsterdam: Elsevier.

Egghe, L. (2006a). An improvement of the *H*-index: the *G*-index. *ISSI Newsletter*, 2, 8–9.

Egghe, L. (2006b). Theory and practice of the g-index. *Scientometrics*, 69(1), 131-152.

Egghe, L. & Rousseau, R. (2006). An informetric model for the h-index. *Scientometrics*, 69(1), 121-129.

Hardy, G., Littlewood, J.E. and Pólya, G. (1952). *Inequalities* (second edition). Cambridge University Press.

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J. and Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1), 169-185.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science of the USA*, 102, 16569–16572.

Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometrics, 85(3), 741-754.

Hunt, R. (1991). Trying an authorship index. *Nature,* 352, 187.

ICMJE (2010). *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and editing for Biomedical Publication*. Retrieved April 20, 2012 from http://www.icmje.org/urm_full.pdf

Jin, BH., Liang, LM., Rousseau R. and Egghe, L. (2007). The *R*- and *AR*-indices: complementing the h-index. *Chinese Science Bulletin*, 52 : 855–863.

Leimu, R. and Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology and Evolution*, 20(1), 28-32.

Linton, J.D., Tierney, R. and Walsh, S.T. (2011). Publish or perish: how are research and reputation related? *Serials Review*, 37(4), 244-257.

Liu, MX. (1993). A study of citing motivation of Chinese scientists. *Journal of Information Science*, 19(1), 13-23.

Liu, YX. (2011). The diffusion of scientific ideas in time and indicators for the description of this process (Doctoral thesis). Antwerp University.

Mahbuba, D. and Rousseau, R. (2011). The Matthew effect and a relation with concept symbols and defaults. *Annals of Library and Information Studies*, 58, 335-345.

Merton, R.K. (1968). The Matthew effect in science. *Science,* 159, 56–63.

Rousseau, R. (1992). Concentration and diversity of availability and use in information systems: a positive reinforcement model. *Journal of the American Society for Information Science*, 43(5), 391-395.

Rousseau, R. & Jin, BH. (2008). The age-dependent h-type AR^2-index: Basic properties and a case study. *Journal of the American Society for Information Science and Technology*, 59(14), 2305–2311.

Skilton, P.F. (2009). Does the human capital of teams of natural science authors predict citation frequency? *Scientometrics*, 78(3), 525-542.

Stephan, P.E. and Levin, S.G. (1991). Inequality in scientific performance: adjustment for attribution and journal impact. *Social Studies of Science*, 21, 351-368.

Tol, R.S.J. (2009). The Matthew effect defined and tested for the 100 most prolific economists. *Journal of the American Society for Information Science and Technology*, 60(2), 420-426.

Tol, R.S.J. (2012). Shapley values for assessing research production and impact of schools and scholars. *Scientometrics,* 90(3), 763-780.

Van Dalen, H. P. and Henkens K. (2001). What makes a scientific article influential? The case of demographers. *Scientometrics*, 50(3), 455-482.

Wang, M., Yu, G. and Yu, D. (2011). Mining typical features for highly cited papers. *Scientometrics*, 87(3), 695-706.