Journal of Informetrics 5 (2011) 181–186

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

A proposal for a First-Citation-Speed-Index

L. Egghe [a,b,*], L. Bornmann [c], R. Guns [b]

[a] Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium
[b] Universiteit Antwerpen (UA), Informatie-en Bibliotheekwetenschap, Stadscampus, Venusstraat 35, B-2000 Antwerpen, Belgium
[c] Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

ARTICLE INFO

Article history:
Received 25 August 2010
Received in revised form 11 October 2010
Accepted 13 October 2010

Keywords:
First-Citation-Speed-Index
FCSI
h-Index
Increasing sequence

ABSTRACT

In this paper, we define a First-Citation-Speed-Index (FCSI) for a set of papers, based on their times of publication and of first citation. The index is based on the definition of a *h*-index for increasing sequences.

We show that the index has several good properties in the sense that the shorter the times are between publication and first citation (in a global manner) the higher the FCSI is.

We present two case studies: a first-citation speed comparison of three journals in the field of psychology and a first-citation speed comparison of accepted and rejected, but published elsewhere manuscripts by the journal *Angewandte Chemie International Edition*. Both case studies indicate that our FCSI satisfies the intuitive feeling of what values a FCSI should have in these cases.

1. Introduction

If we have a set of cited articles (e.g. in a journal) we can determine, for each article, the publication time, denoted t_P and, the time that this article received its first citation (denoted t_C and of course we can assume that $t_C \geq t_P$). How t_P and t_C can be expressed (e.g. in months or years) is not important at this moment, but we will spend some comments on it in the discussion section at the end of this paper.So for all these papers, we can determine

$$t_1 = t_C - t_P, \tag{1}$$

i.e. the time between publication of a paper and first-citation to this paper. This is an important indicator since it expresses how fast (t_1 "small") this paper changes its status of "unused" to "used". Since we have a set of papers we, hence, also have a sequence of t_1-values. As is classical in such cases we can wonder if this sequence can be used to define a kind of "First-Citation-Speed-Index" (FCSI). Of course, we should discuss some desired properties for such a speed index (this will, of course, be done in this paper). The FCSI is dependent on the t_1-values of ever cited articles, logically. It will measure the citation speed of cited articles and does not take into account the articles that are not (yet) cited. This set is also important in the measurement of the citedness of a set of articles but should not be involved in the measurement of the first-citation speed (since there is not a first-citation).

Deriving an indicator from a sequence reminds us of the Hirsch-index or *h*-index (Hirsch, 2005). There, in its classical application, we have a decreasing sequence of numbers of citations to papers of e.g. an author. This is a situation where "large is beautiful" since the higher the number of citations to a paper are, the better and also: the higher the *h*-index, the better. In our case of t_1-values, however, it does not make much sense to arrange them in decreasing order. Here we have

* Corresponding author. Permanent address: Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium.
E-mail addresses: leo.egghe@uhasselt.be (L. Egghe), bornmann@gv.mpg.de (L. Bornmann), raf.guns@ua.ac.be (R. Guns).

1751-1577/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2010.10.006

"small is beautiful" since the smaller the t_1-values, the better. So it is natural to arrange these t_1-values in increasing order. But it is not clear how to apply the well-known definition of the h-index to increasing sequences.

In the next section we will propose a method to define "an h-index for increasing sequences". With this h-index we are then able to define an indicator of "first-citation-speed". Here, again "large is beautiful" since high speed indicates small t_1-values. Details will be given. We give some examples but we also formulate some "logical" good properties that a FCSI should have. For instance, if all t_1-values are 0, the FCSI should attain its maximum. Conversely, if all t_1-values are very high (in the limit, ∞), the FCSI should go to its minimum. Also adding a constant $a > 0$ to all t_1-values should decrease the value of the FCSI. Our indicator satisfies these good properties. Also we can show that multiplying all t_1-values with a constant $a > 1$ leads to a decrease of our FCSI, another good property.

In the next section practical examples of journals in psychology and chemistry are given, illustrating the good properties of our FCSI.

A first attempt for defining such a FCSI was given in Bornmann and Daniel (2010). Avoiding the problem of defining a h-index for increasing sequences, they apply the h-index (for decreasing sequences) on the data of time between the present time and the time of the first-citation. They argue that the larger this number is, the more the first-citation time is to the past, hence the earlier the first-citation is received. The present paper tries to improve this method by also taking into account the time of publication since it is essential in calculating first-citation speed.

Mean citation speed in general (not only involving the first citation) and not in the context of the h-index, was studied in Schubert and Glänzel (1986).

The paper closes with some concluding remarks and some open problems (incl. the problem of how to deal with time-units in FCSIs).

2. Definition of the First-Citation-Speed-Index (FCSI) and its good properties

We assume that we have a set of papers (e.g. in a journal) and for each of them we have determined the time t_P of publication and the time t_C at which the first citation is received and $t_1 = t_C - t_P$.

Let

$$t_m = \max t_1 \tag{2}$$

where the maximum is over all papers. Since small values of t_1 are important (cf. the Introduction) we will define a h-index for increasing sequences as follows. Calculate $t_m - t_1 + 1$ for each paper and put them in decreasing order (we have added 1 so that the case of one paper or the case of papers with equal t_1 yields 1 and not 0). On this decreasing sequence, the classical h-index can be calculated. We denote it by h.

We then define the First-Citation-Speed-Index (FCSI), denoted F, as

$$F = \frac{1}{t_m - h + 1} \tag{3}$$

First we look at some simple examples which show that we are on the right track.

1. Set A has 3 articles with t_1-values: 1,2,3. Hence $t_m = 3$ and the $t_m - t_1 + 1$ values are 3,2,1, hence $h = 2$ and hence

$$F = \frac{1}{3 - 2 + 1} = \frac{1}{2}$$

2. Set B has 3 articles with t_1-values: 1,2,4. Now $t_m = 4$ and the $t_m - t_1 + 1$ values are 4,3,1, hence $h = 2$ and hence

$$F = \frac{1}{4 - 2 + 1} = \frac{1}{3}$$

It is logical that the FCSI of B is smaller than that of A.

3. Set C has 3 articles with t_1-values: 1,2,6. Now $t_m = 6$ and the $t_m - t_1 + 1$ values are 6,5,1, hence $h = 2$ and

$$F = \frac{1}{6 - 2 + 1} = \frac{1}{5}$$

It is logical that this value of C is smaller than the ones of A and B.

Of course, as is also the case with the h-index, there is some robustness.

4. Set D has 3 articles with t_1-values: 1,3,4. Now $t_m = 4$ and the $t_m - t_1 + 1$ values are 4,2,1, hence $h = 2$ and

$$F = \frac{1}{4-2+1} = \frac{1}{3},$$

the same as for B although the speed in D is a bit slower (a strictly higher value of F for case D would be a bad property which is not the case here).

Let us now consider some "logical" requirements for a FCSI and we will check if F satisfies these.

1. Let us have N papers all with $t_1 = 0$ (in practise, many papers have this as follows from Bornmann and Daniel (2010) but it is also the case for the data of e.g. Egghe (see Egghe, L., in Web of Science, Thomson Reuters)). Now $t_m = 0$ and $t_m - t_1 + 1 = 1$ for all papers, hence $h = 1$ (as we can see clearly here, h is not a good speed measure). Now $t_m - h + 1 = 0$ and hence $F = +\infty$, the highest possible value, as it should in this case since all t_1-values are 0.

   Note: if we want to avoid $+\infty$ as highest value we can always apply a strictly increasing transformation such as

$$x \to \frac{2}{\pi} Arc\tan(x) \tag{4}$$

   (highest value $+\infty$ is then the highest value 1 while the lowest value 0 remains 0).

2. Let us have N papers with all t_1-values equal and very high (say $\approx +\infty$), the worst case.

   Now $t_m \approx +\infty$, all values $t_m - t_1 + 1 = 1$ and hence $h = 1$.

   Now

$$F = \frac{1}{t_m - h + 1} \approx 0 \tag{5}$$

   (and =0 for $t_1 = +\infty$), the lowest possible value for F.

3. Compare two cases

3.1. N papers with the same t_1-values

   Here $t_m = t_1$, and all $t_m - t_1 + 1$ values are 1, hence $h = 1$. So

$$F = \frac{1}{t_m - h + 1} = \frac{1}{t_1} \tag{6}$$

3.2. N papers with the same t_1'-values, namely $t_1' = t_1 + 1$. The same argument as above, with t_1 replaced by t_1' yields

$$F = \frac{1}{t_1 + 1} < \frac{1}{t_1} \tag{7}$$

   a logical fact. More general: if all t_1'-values are equal to $t_1 + a$ ($a > 0$) then

$$F = \frac{1}{t_1 + a}, \tag{8}$$

   decreasing in a, which is a logical fact.

3.3. N papers with the same t_1''-values, namely $t_1'' = at_1$ where $a > 1$. The same argument as above with t_1 replaced by t_1'' yields

$$F = \frac{1}{at_1} < \frac{1}{t_1} \tag{9}$$

   a logical fact.

4. The next case shows a robustness property of F. Let $N - 1$ papers have equal t_1-values t_1 and the Nth paper has t_1-value $t_1' > t_1$. Then $t_m = t_1'$. In table for calculating h we have that the first $N - 1$ papers have $t_m - t_1 + 1 = t_1' - t_1 + 1$ values and the Nth paper has $t_m - t_1' + 1 = 1$ value. If N is sufficiently large we have that $h = t_1' - t_1 + 1$ and

$$F = \frac{1}{t_m - h + 1} = \frac{1}{t_1' - (t_1' - t_1 + 1) + 1}$$

$$F = \frac{1}{t_1}$$

Table 1
First citation data for *Discourse & Society*, *Psychological Methods* and *American Journal of Community Psychology*.

Discourse & Society			*Psychological Methods*			*American Journal of Community Psychology*		
Rank	t_1	$t_m - t_1 + 1$	Rank	t_1	$t_m - t_1 + 1$	Rank	t_1	$t_m - t_1 + 1$
1	0	9	1	0	7	1	0	4
2	1	8	2	1	6	2	0	4
3	1	8	3	1	6	3	0	4
4	1	8	4	1	6	4	0	4
5	1	8	5	1	6	5	0	4
6	1	8	6	1	6	6	0	4
7	1	8	7	1	6	7	0	4
8	2	7	8	1	6	8	0	4
9	2	7	9	1	6	9	0	4
10	2	7	10	1	6	10	0	4
11	2	7	11	1	6	11	0	4
12	3	6	12	1	6	12	1	3
13	3	6	13	2	5	13	1	3
14	4	5	14	2	5	14	1	3
15	4	5	15	2	5	15	1	3
16	5	4	16	2	5	16	1	3
17	5	4	17	2	5	17	2	2
18	8	1	18	2	5	18	2	2
			19	3	4	19	2	2
			20	4	3	20	2	2
			21	4	3	21	2	2
			22	4	3	22	2	2
			23	4	3	23	2	2
			24	5	2	24	2	2
			25	6	1	25	2	2
			26	6	1	26	2	2
						27	2	2
						28	2	2
						29	2	2
						30	2	2
						31	2	2
						32	2	2
						33	3	1
						34	3	1
						35	3	1
						36	3	1
						37	3	1

(as if the Nth paper was not there): this is exactly the same robustness as with the h-index where small values do not count. We could have several papers with high t_1'-values as long as the h-index is based on the top articles with t_1'-values, $t_1 < t'$. This robustness of the h-index is considered as a good property!

3. Case study: three journals from subject category 'Psychology'

We will now investigate the FCSI in practice. To this end, we collected publication and citation data for the following three journals from the ISI subject category 'Psychology': *Discourse & Society*, *Psychological Methods*, and *American Journal of Community Psychology*.

All data were collected from Web of Science on May 10, 2010. The citations in the data refer to articles published in these journals in the year 2000, thus allowing for sufficient time for all articles to gain (at least) a first citation. The data were obtained using a query like "SO = (AMERICAN JOURNAL OF COMMUNITY PSYCHOLOGY) AND PY = 2000" – and similarly for the other two journals – and analyzed using the 'Citation Report' feature. The time units here are years; for instance, if $t_1 = 0$, the first citation was received in the year of publication. We note that, overall, there are only two articles – one in *Discourse & Society* and one in *Psychological Methods* – that have no citations and hence, as explained in the introductory section, they are removed from the data set.

With t_1 and t_m defined as in the previous section, Table 1 contains the first citation data for *Discourse & Society*, *Psychological Methods*, and *American Journal of Community Psychology*. It is now straightforward to determine the h-index for each journal on the basis of the third column for each journal. We first look at *Discourse & Society* and find $h = 7$. The largest value for t_1 is 8, hence $t_m - h + 1 = 2$ and $F = 1/2$. Next, we turn to *Psychological Methods*. For this journal, we find that $h = 6$, and hence $t_m - h + 1 = 1$ and $F = 1$.

The *American Journal of Community Psychology* has $h = 4$. Thus, $t_m - h + 1 = 0$ and $F = 1/0 = +\infty$. This is a case somewhat similar to the one discussed in requirement 1.

Table 2
First citation data for accepted or rejected, but published elsewhere manuscripts with $t_1 < 18$.

t_1	$t_m - t_1 + 1$	Accepted manuscripts	Rejected manuscripts
17	61	8	11
16	62	18	22
15	63	3	25
14	64	15	25
13	65	19	21
12	66	22	43
11	67	22	46
10	68	28	45
9	69	52	47
8	70	60	57
7	71	60	63
6	72	74	70
5	73	72	71
4	74	92	65
3	75	66	51
2	76	74	32
1	77	44	24
0	78	86	27

Generally, if the first $t_m + 1$ papers have $t_1 = 0$, then each of these papers has $t_m - t_1 + 1 = t_m + 1$, and $h = t_m + 1$. Hence, in this case $t_m - h + 1 = 0$ and $F = +\infty$. This 'best case scenario' can also be found in practice, as evidenced by Table 1. Indeed, the first 11 ($> t_m + 1 = 4$) papers received their first citation within less than a year, leading to the largest F possible.

These results accord well with our intuition of how these three journals should compare regarding first citation speed. Although *Discourse & Society* has published less articles than *Psychological Methods*, its t_m is larger. The same observation holds when comparing either to *American Journal of Community Psychology*. More importantly, more than one quarter of the articles in the latter journal have been cited within the same year, which clearly exceeds the other two journals. Similarly, a larger fraction of articles from *Psychological Methods* has been cited after one year, compared to *Discourse & Society*.

It is interesting to compare the FCSI to the median of the t_1 values. In principle, this also gives an indication of first citation speed (note, though, that, for the median, smaller values indicate higher speed). In this case study, however, the median for each journal equals 2, which would lead us to conclude that these journals are similar in first citation speed. This is clearly in contrast with the FCSI, which assigns a different score to each one. It seems that the FCSI paints a more correct picture of their existing differences.

The immediacy index (II, Thomson Reuters' Journal Citation Reports) is the number of citations in the year of publication divided by the number of publications. Hence, it can also be considered an indicator of early interest in a given paper. For the three journals from Table 1, we find II values of, respectively, 0.263, 0.185 and 0.405. Just like for the FCSI, the *American Journal of Community Psychology* has the highest value. The ranking of the other two journals by II is, however, different from the ranking by F. Closer examination shows that this is due to the number of papers published in 2000; in fact, both journals receive 5 citations within the same year, but *Psychological Methods* has published more articles, leading to a lower II.

4. Case study: first-citation speed comparison of accepted and rejected, but published elsewhere manuscripts by *Angewandte Chemie International Edition*

For the second case study we used bibliometric data of Bornmann and Daniel (2010) for 1899 manuscripts that were submitted to *Angewandte Chemie International Edition* (AC-IE). What the editors of AC-IE look for most of all is excellence in chemical research. Manuscripts that reviewers deem to be of high quality are selected for publication. Manuscripts that do not meet the high standards are rejected. Of the 1899 manuscripts that were reviewed by the AC-IE in the year 2000, 46% ($n = 878$) were accepted for publication in AC-IE, and 54% ($n = 1021$) were rejected. A search in the literature databases Science Citation Index (Thomson Reuters) and Chemical Abstracts (Chemical Abstracts Services) revealed that of the 1021 rejected manuscripts, 959 (94%) were published later in 136 other (different) journals. For accepted and rejected (but published elsewhere) manuscripts, Bornmann and Daniel (2010) determined – in addition to the number of citations – the number of months since publication and the first time the paper was cited. If $t_1 = 0$, the first citation was received within the publication month. The searches were done using the Web of Science.

For different values of t_1 and $t_m - t_1 + 1$, Table 2 shows the number of accepted and rejected, but published elsewhere manuscripts. There are, e.g., 8 accepted and 11 rejected, but published elsewhere manuscripts with $t_1 = 17$ and $t_m - t_1 + 1 = 61$. For both manuscript groups, the largest value for the difference between t_c and t_p is 77 ($= t_m$). Similar to the *American Journal of Community Psychology* in the previous section, we have a 'best case scenario' in Table 2 for the accepted manuscripts: 86 manuscripts received their first citation within the publication month (about 10% of all accepted manuscripts). This leads to the largest possible F. For the rejected, but published elsewhere manuscripts we find $h = 76$. The largest value for t_1 is 77 ($= t_m$), hence $t_m - h + 1 = 2$ and $F = 1/2$.

The difference in F between both manuscript groups point to a higher first-citation-speed for accepted than for rejected, but published elsewhere manuscripts. These findings are in accordance to the results of Bornmann and Daniel (2010). They found not only a higher h index (Hirsch, 2005) for accepted than for rejected, but published elsewhere manuscripts, but also a higher speed, defined there as a longer time between first citation and date of search for the first citation. Thus, the FCSI results are convergently valid: they correspond to the results produced by other (speed) indicators and to the qualitative outcome of the AC-IE peer review process.

5. Conclusions and suggestions for further research

We presented a proposal for a First-Citation-Speed-Index (FCSI), hereby introducing an h-index for increasing sequences. Several good "logical" properties are shown and the practical examples underline the good distinctive power of the FCSI. The FCSI takes values between 0 and $+\infty$ but we indicate how the FCSI can be transformed into an FCSI with values in the interval $[0, 1]$ and with the same good properties.

Several problems remain. First of all, one should search for other good FCSIs. As in concentration theory (cf. Egghe, 2005) one has defined several concentration measures with good properties. So we should do the same for FCSIs. All FCSIs should be based on the values $t_1 = t_C - t_P$.

For measuring FCSI, we only use the ever-cited articles. This is logical since first-citation speed can only be measured on ever-cited articles. Yet, the set (or fraction) of non-cited articles is also an important informetric phenomenon. As no indicator is perfect (and so is FCSI and even the h-index (see Egghe, 2010)) we could think of using an indicator on non-cited articles that would complement the FCSI.

Next there is the problem with the time units. Of course, if we keep the same time unit, examples can be compared. The problem with time unit can be formulated in two different ways. Suppose we keep the same time unit but we apply the transformation $t \to at$ where $a > 1$. Then all time periods (such as t_1) are strictly larger and hence the FCSI should decrease. But we can also look at $t \to at$ being a transformation where the time unit is changed (e.g. going from a year to a month). The same time period is then multiplied by $a = 12$ and it is not clear how a FCSI should behave in this context.

This problem is comparable (though not identical) with the principle of scale invariance for concentration measures (also called inequality measures). If a transformation $r \to ar$ means a change of the currency (e.g. from \$ to €) then the inequality should be identical: wealth or poverty is not changed when we change the currency! But if the transformation $r \to ar$ (say for $a > 1$) means that each person's income (or capital) is multiplied by a (keeping the same currency) then it is not at all clear that inequality remains the same. A study of the effect of scale on the h-index is given in Egghe (in press).

The problem of the time units in the framework of FCSIs is left as an open problem.

References

Bornmann, L., & Daniel, H.-D. (2010). The citation speed index: A useful bibliometric indicator to add to the h index. *Journal of Informetrics*, *4*(1), 83–88.

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Oxford, UK: Elsevier.

Egghe, L. (2010). The Hirsch-index and related impact measures. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 65–114). Medford, NJ: Information Today, Inc.

Egghe, L. (in press). A mathematical model of the scale-dependence of the h-index and other h-type indices, Scientometrics.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Schubert, A., & Glänzel, W. (1986). Mean response time—A new indicator of journal citation speed with applications to physics journals. *Czechoslovak Journal of Physics*, *36*(1), 121–125.