Journal of Informetrics 7 (2013) 313–317

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

Family-tree of bibliometric indices

Marek Kosmulski *

Lublin University of Technology, Lublin, Poland

ARTICLE INFO

Article history:
Received 7 November 2012
Received in revised form 6 December 2012
Accepted 11 December 2012
Available online 9 January 2013

Keywords:
Successful paper
h-Index
Citation analysis

ABSTRACT

Standard bibliometric indices were re-defined using a generalized concept of "successful paper". A family-tree based upon the new definitions provides new insights into the relationships between the standard indices, and empty boxes in the family-tree may inspire design of new indices.

1. Introduction

Let us consider a citation record of all N papers of an individual sorted from the most cited (c_1 citations) to the least cited (c_N citations), without tied ranks, that is, each paper has its unique rank ($1, 2, \ldots, N$), also when several papers have the same number of citations. Many bibliometric indices based on the analysis of such a citation record have been considered, and a few of those indices are widely used in assessment of the scientific output of an individual, also for practical purposes. The indices have their names and acronyms designed by their authors, but those names do not tell much about the relationships between the indices. Particular indices are highly correlated, that is, the effect of the choice of the index on the rank of an individual in a set of individuals is rather insignificant. Thus the information provided by multiple indices is redundant (Bornmann, Mutz, Hug, & Daniel, 2011). However, a high correlation of the scores does not imply generic similarity of the indices. A family-tree of bibliometric indices based a generalized concept of "successful paper" is presented and discussed in this paper.

2. Successful papers

Kosmulski (2011) argued that the scientific output of an individual could be assessed by the number of his/her successful papers, while the other (unsuccessful) papers could be ignored. Only one definition of a successful paper was considered. The concept of successful paper can be interpreted as a transformation of the ordinal scale (papers ranked as discussed in Section 1) to a nominal scale (a paper can be successful or not) in the sense discussed by Stevens (1946).

Franceschini, Galette, Maisano, and Mastrogiacomo (2012) generalized the concept of a successful paper, and considered several different definitions. The concept of a successful paper can be further generalized, and many standard bibliometric indicators can be re-defined as numbers of successful papers, total numbers of citations of successful papers or average numbers of citations of successful papers. The new definitions are summarized in the family-tree (Table 1, column 1). Each definition of a successful paper produces at least three bibliometric indices representing the number of successful papers

* Tel.: +48 81 5384355.
E-mail addresses: mkosmuls@abo.fi, mkosmuls@hektor.umcs.lublin.pl

1751-1577/$ – see front matter
http://dx.doi.org/10.1016/j.joi.2012.12.002

Table 1
A family tree of bibliometric indices. Existing indices (or expressions dependent on the existing indices) are filled in the corresponding cells. The empty cells are indices, which have not been considered in the literature. N/A, not applicable (e.g., division by zero).

Definition of successful paper All papers	# successful papers Number of papers (N)	# citations of all successful papers Number of citations	Citations per one successful paper Citations per paper
$c_i \geq i$	h (Hirsch, 2005)	R^2 (Jin, Liang, Rousseau, & Egghe, 2007)	A (Jin et al., 2007)
$c_i \geq C \times i$ $C = 10$	w (Wu, 2010)		
$c_i \geq i^n$ $n = 2, 3, \ldots$	$h(n)$ (Kosmulski, 2006a)		
Top $\sqrt{N}$ papers	$\sqrt{N}$ (rounded to the nearest integer)	100 π (Vinkler, 2010)	
$\Sigma c_j \geq i^2$ ($j \leq i$)	g (Egghe, 2006)	$\approx g^2$	$\approx g$
$(\Pi c_j)^{1/i} \geq i$ ($j \leq i$)	t (Tol, 2009)		
$c_i \geq X$ $X = 500$	TP (Fu et al., 2012)		
Top Y papers $Y = 1$	Y	c_1	c_1
Top papers, which have together $\geq Z$ citations	N/A for $\Sigma c_j < Z$ 1 for $c_1 > Z$	$\geq Z$ (or N/A)	N/A for $\Sigma c_j < Z$ c_1 for $c_1 > Z$
$c_i >$ #cited items	NSP (Kosmulski, 2011)		
$c_i >$ median number of citations received by articles in the same journal and in the same year	*success*-Index (Franceschini et al., 2012)		

(column 2), the total number of citations of the successful papers (column 3), and the number of citations per a successful paper (column 4), respectively.

The rankings of scientists based upon bibliometric indicators can be interpreted in terms of ordinal scales in the sense discussed by Stevens (1946). Order-preserving transformations (Stevens, 1946) leave the scale form invariant. Therefore the results of arithmetic operations like multiplication by a constant number or extraction of a square root are considered as equivalent to the original indicator, and the bibliometric indicators derived from the total number of citations of successful papers by order-preserving transformations are all in column 3. Several new columns might be added to Table 1, e.g., representing median or geometric or harmonic average of the numbers of citations of successful papers. Yet, such types of bibliometric indicators have not attracted much attention of the researchers, and they may only be considered as suggestions for further research.

The names of existing indicators are summarized in the corresponding boxes of Table 1. Empty boxes indicate potential indicators, which have not been explicitly considered. A few boxes describe the relationships between indicators and/or parameters used to define those indicators. Numerical values of all indices for a (authentic and rather typical) citation record from Table 2 are reported in Table 3.

When all papers are formally considered successful (such a definition contradicts an intuitive meaning of a successful paper), the present approach leads to the number of published papers, the total number of citations, and the number of citations per paper, which are well-known bibliometric indicators. For example in the WoS database, these numbers are displayed automatically in a “citation report” mode.

The Hirsch’ (2005) approach corresponds to a successful paper defined as belonging to the h-core ($c_i \geq i$). The number of papers in the h-core is the h-index, the total number of citations in the h-core is represented by the R-index, and the number of citations per paper in the h-core is the A-index (Jin at al.). Actually the R-index is defined as the square root of total number of citations on the h-core. Therefore R^2 rather that R is reported in Table 1. For clarity’s sake, both R and R^2 are reported in Table 3. For any set of indices presented in one row of Table 3 we have:

$$\text{column} \quad 2 \times \text{column} \quad 4 = \text{column} \quad 3.$$

This is because a number in column 4 is defined as the quotient of the numbers columns 3 and 2 (the average is the sum divided by the number of elements, by definition). Therefore $h \times A = R^2$. To the best knowledge of the present author, this relationship between three well-known indices escaped the attention of scientists. Hirsch’ (2005) approach is probably the only non-trivial definition of successful paper, for which all three types of indices considered in Table 1 have been defined and discussed.

Wu (2010) re-defined successful papers in a more general way than in the discussed above Hirsch’ approach. The h-core can be considered as a special case of the Wu-core ($c_i \geq C \times i$) for $C = 1$. The C-values other than 1 produce new indices, which are different from h, A and R. Wu (2010) elaborated the case of $C = 10$, which awards authors of very few highly-cited papers over those who have published numerous moderately cited papers. The number of so defined successful papers (for $C = 10$) has been considered as a bibliometric indicator, but the total number of citations and the number of citations per paper have not. The average number of citations in the Wu core ($C > 1$) is a rather controversial indicator. For example the citation

Table 2
Typical citation record of a series of papers of a scientist.

Rank	Citations	Rank	Citations	Rank	Citations	Rank	Citations
1	142	41	16	81	5	121	1
2	137	42	16	82	5	122	1
3	132	43	15	83	5	123	1
4	117	44	14	84	5	124	1
5	110	45	14	85	4	125	1
6	70	46	14	86	4	126	1
7	67	47	13	87	4	127	0
8	62	48	13	88	4	128	0
9	61	49	12	89	4	129	0
10	54	50	12	90	4	130	0
11	51	51	12	91	4	131	0
12	42	52	12	92	4	132	0
13	42	53	11	93	4	133	0
14	42	54	11	94	4	134	0
15	40	55	11	95	4	135	0
16	40	56	11	96	4	136	0
17	38	57	11	97	4	137	0
18	38	58	10	98	4	138	0
19	37	59	9	99	3	139	0
20	36	60	9	100	3	140	0
21	35	61	8	101	3	141	0
22	35	62	8	102	3	142	0
23	35	63	8	103	3	143	0
24	34	64	8	104	3	144	0
25	33	65	8	105	3	145	0
26	28	66	8	106	3	146	0
27	27	67	8	107	2	147	0
28	27	68	8	108	2	148	0
29	25	69	7	109	2	149	0
30	24	70	7	110	2	150	0
31	23	71	7	111	2	151	0
32	22	72	7	112	2	152	0
33	21	73	6	113	2	153	0
34	21	74	6	114	2	154	0
35	21	75	6	115	2	155	0
36	20	76	6	116	2	156	0
37	20	77	6	117	2		
38	18	78	6	118	1		
39	17	79	6	119	1		
40	16	80	5	120	1		

record of an individual presented in Table 2 produces an average number of citations in the Wu core ($C = 10$) equal to 118 (Table 3). Three additional citations of the paper #7 raise the Wu index from 6 to 7, but the average number of citations in the Wu-core drops to 111, and it takes 51 additional citations of papers #1 to 7 to bring the average number of citations in the Wu core back to its original value. Thus a scientist can be punished for additional citations of his/her paper by the others.

Kosmulski (2006a) re-defined successful papers in a more general way than in the discussed above Hirsch' approach. The h-core can be considered as a special case of the $h(n)$-core ($c_i \geq i^n$) for $n = 1$. The n-values other than 1 produce new indices, which are different from h, A and R. The number of so defined successful papers (for $n = 2$ and 3) has been considered as a bibliometric indicator, but the total number of citations and the number of citations per paper have not. The total number

Table 3
Values of bibliometric indices defined in Table 1 calculated for the example from Table 2. Ordered by "threshold".

Definition of successful paper	# items	# citations	Citations per item
All papers	156	2376	15.23
Franceschini et al. (2012)	66	2039	30.89
Egghe (2006)	43	1907	44.35
NSP Kosmulski (2011)	40	1516	37.9
Tol (2009)	39	1844	47.28
Hirsch (2005)	27	1585 ($R = 39.81$)	58.7
Vinkler (2010)	13	1087 ($\pi = 10.87$)	83.62
Top papers, which have together ≥1000 citations	11	1003	91.18
$h(2)$ Kosmulski (2006a)	7	775	110.71
Wu (2010)	6	708	118
Top 1 paper	1	142	142
Fu et al. (2012)	0	0	N/A

of citations in the $h(2)$ or $h(3)$ core can be considered as an indicator, which awards authors of very few highly-cited papers over those who have published numerous moderately cited papers. The average number of citations in the $h(n)$ core ($n > 1$) is a rather controversial indicator for the same reason as discussed above for the average number of citations in the Wu core ($C > 1$).

The definitions by Wu and Kosmulski can be combined by defining successful papers as the W–K core ($c_i \geq C \times i^n$), which reduces to the Wu-core for $n = 1$, to the $h(n)$-core for $C = 1$, and to the h-core for $n = C = 1$. Indices based on a combined Wu–Kosmulski approach in non-trivial cases ($C \neq 1$ and $n \neq 1$) have not attracted much attention, and will not be discussed here.

In Vinkler's (2010) approach the top $\sqrt{N}$ papers are defined as successful (elite papers). The number of papers in the Vinkler's core is unequivocally defined by the number of papers, so $\sqrt{N}$ does not provide any useful information on top of that already provided by N. The π-index (Vinkler) is the total number of citations of successful papers multiplied by 0.01. The π-index and the total number of citations of successful papers (top $\sqrt{N}$ papers) are equivalent in the sense discussed above. The average number of citations of papers in the Vinkler's core can be considered as a new bibliometric indicator. It can increase or decrease in time, namely a scientist can be punished for additional publications (which originally have very few citations) when the increase in the number of papers is not accompanied by additional citations of the elite papers. Scientists can avoid such a punishment by controlling their productivities (keeping them within reasonable limits).

In Egghe's (2006) approach the papers are successful as long as the total number of citations of the top i papers is at least i^2. The g-index (Egghe, 2006) has an unique property, namely in typical citation records of individuals the numbers of successful papers (belonging to the g-core) are equal to or marginally lower than the numbers of citations per item, and the total numbers of citations are equal to or marginally higher than g^2. Then all three indices based on Egghe-type definition of successful papers are nearly equivalent in the sense discussed by Stevens (1946) (at least for typical citation records, cf. Table 3). The Egghe-type definition of successful papers (Table 1) needs additional explanation when $c_1 > N^2$. The present author prefers an approach, in which dummy papers with 0 citations each can be added when necessary.

In Tol's (2009) approach the papers are successful as long as the geometric average of the top i papers is at least i. The number of so defined successful papers (t) has been considered as a bibliometric indicator, but the total number of citations and the number of citations per paper have not.

The idea of considering the papers, which have at least X citations as successful has been widely discussed, e.g., Google Scholar Blog (2011) ($X = 10$), Pilc (2002), Kosmulski (2012) ($X = 100$). It was criticized for the arbitrariness of X (certain value of X may favor or disfavor individuals) (Hirsch, 2005) ($X = 50$). For example the citation record shown in Table 2 would receive the same score for $X = 110$ and for $X = 71$. Therefore an owner of such a citation record would welcome $X = 110$, while scientists who have many papers with 71–109 citations would prefer $X = 71$. The number of so defined successful papers ($X = 500$) has been considered as a bibliometric indicator (Fu, Wang, & Ho, 2012), but the total number of citations and the number of citations per paper have not. Obviously the arbitrariness of X results in limited credibility of the total number of citations and of the number of citations per paper in successful (having at least X citations) papers as bibliometric indicators.

The idea of considering the top Y papers as successful has been widely discussed, e.g., Hirsch (2005) ($Y = 5$), Kosmulski (2012) ($Y = 1$). The number of citations of the most-cited paper (used as a bibliometric indicator) was criticized for its one-sidedness. When the number of successful papers is pre-defined (Y), it does not bring any information about a scientist having at least Y papers. Therefore, the number of so-defined successful papers is not useful as a bibliometric indicator. Also the number of citations per paper for $Y = 1$ (column 4) must not be considered as a new indicator since it is equal to the number of citations of the most-cited paper (column 3).

The above indicators are based solely on the citation records of all papers of an individual. The summary of definitions of successful papers presented in Table 1 is not exhaustive, but it covers the best-known cases. The indices based solely on the citation records of the papers of an individual have a common disadvantage of favoring old scientists over young scientist and giving undue credit to co-authors of multi-author papers.

The top papers, which have together $\geq Z$ citations can be formally considered as successful papers, but indicators based on such a definition have not been seriously considered as measures of scientific output of individuals because of serious shortcomings of such an approach. Arbitrariness of Z (which has to be selected somehow) is the main problem, namely certain value of Z may favor or disfavor individuals. Only $c_1 < Z < \Sigma c_j$ produces non-trivial results (different scientists have different numbers of successful papers). Unlike other indicators discussed in the present study (in which high values indicate successful scientists), low number of papers in the "Z-core" is more desired than a high number.

The top P % (for example top 10% most cited) is another possible definition of a successful paper, which can be discussed in the framework of the present model. With $P = 100/\sqrt{N}$ it is equivalent to the Winkler's definition.

Many other definitions of successful papers can be proposed when data other than numbers of citations of particular papers are available. For example Kosmulski (2011) defined successful papers as those, which have been cited more times than their numbers of references. Franceschini et al. (2012) defined successful papers (among other definitions) as those, which have been cited more times than median number of citations received by articles in the same journal and in the same year. The numbers of so defined successful papers has been considered as bibliometric indicators (NSP and *success*-index, respectively), but the total numbers of citations and the numbers of citations per paper have not.

Practically unlimited number of definitions of successful papers can be proposed based on different types of bibliometric data. Those data may include the number of co-authors, the year of publication, etc. Thus the "success" of a scientific paper can be adjusted for co-authorship, the age of publication (scientist), and the branch of science represented by that scientist.

The values of bibliometric indices calculated for a typical publication record (Table 2) are summarized in Table 3. In principle the "most demanding" definitions of successful papers (e.g., Wu and $h(2)$) produce low numbers of successful papers and low total numbers of citations of successful papers, but high numbers of citations per paper, and "less demanding" definitions of successful papers (e.g., Egghe and Tol) produce high numbers of successful papers and high total numbers of citations of successful papers, but low numbers of citations per paper. The order of the definitions from "most demanding" to "less demanding" may vary from one scientist to another, but substantial deviations from the pattern shown in Table 3 are excluded by definitions of particular indices, e.g., $h(2) \le h \le g$ by definition. The anticorrelation between the columns 2 and 4 in Table 3 is disturbed by NSP. Several highly cited papers do not belong to the NSP-core (e.g., paper #2 with 137 citations and 151 references), and several moderately cited papers do belong to the NSP-core (e.g., paper #124 with 1 citation and 0 references), while in most other indicators only the top-cited papers are counted as successful.

3. Indicators not covered in the family tree

Several bibliometric indices could be re-defined using the idea of successful paper and arranged into a family tree. Many other indices, also those based solely on the citation records of the papers of an individual can hardly be re-defined using the idea of successful paper described above. For example in the w index (Wohlin, 2009), the papers having at least $X = 5$ citations are considered as successful, but papers having at least 10 citations are "more successful" than papers having 5–9 citations, papers with at least 20 citations are more successful than those with 10–19 citations, etc. Thus there are several levels of "success". In the multidimensional approach (Garcia-Perez, 2009) the Hirsch' approach is used to define the 1st level of successful papers, but among those papers which are not successful ($c_i < i$), lower categories of successful papers are defined using an analogous procedure. The approach discussed by Wohlin and Garcia-Perez is based on multi-level definition of successful papers (contrary to one-level approach discussed in the present study).

4. Further research

The above definitions of successful papers can be applied to independent citations. A family tree presented in Table 1 can be verbatim copied, only the data (Table 2) are different. Yet exclusion of self citations requires extra work while the citation records including self-citations are readily available.

Sets of successful papers be considered at other aggregation levels than single author (e.g., institution, journal). Hierarchical indices (Kosmulski, 2006b) based on the idea of successful papers (successful scientists, successful research groups, etc.) can also be built.

References

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, *5*(3), 346–359.

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

Franceschini, F., Galette, W., Maisano, D., & Mastrogiacomo, L. (2012). The *success*-index: An alternative approach to the *h*-index for evaluating an individual's research output. *Scientometrics*, *92*(3), 621–641.

Fu, H. Z., Wang, M. H., & Ho, Y. S. (2012). The most frequently cited adsorption research articles in the Science Citation Index (expanded). *Journal of Colloid and Interface Science*, *379*(1), 148–156.

Garcia-Perez, M. A. (2009). A multidimensional extension to Hirsch's h-index. *Scientometrics*, *81*(3), 779–785.

Google Scholar Blog. (2011). http://googlescholar.blogspot.com/2011/11/google-scholar-citations-open-to-all.html

Hirsch, J. E. (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*, 16569–16572.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Kosmulski, M. (2006a). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, *7*, 4–6.

Kosmulski, M. (2006b). (*I – A bibliometric indicator*) *I – wskaźnik bibliometryczny*. Lublin: Forum Akademickie., p. 31 [in Polish]

Kosmulski, M. (2011). Successful papers: A new idea in evaluation of scientific output. *Journal of Informetrics*, *5*(3), 481–485.

Kosmulski, M. (2012). Calibration against a reference set: A quantitative approach to assessment of the methods of assessment of scientific output. *Journal of Informetrics*, *6*(3), 451–456.

Pilc, A. (2002). Królowie nauki, ranking polskich uczonych [Kings of science, a ranking of Polish scientists]. *Wprost*, *14*(1010.) [in Polish].

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Tol, R. S. (2009). The h-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics*, *80*(2), 317–324.

Vinkler, P. (2010). The π_v-index: A new indicator to characterize the impact of journals. *Scientometrics*, *82*, 461–475.

Wohlin, C. (2009). A new index for the citation curve of researchers. *Scientometrics*, *81*(2), 512–533.

Wu, Q. (2010). The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology*, *61*(3), 609–614.