The *f* Index: Quantifying the Impact of Coterminal Citations on Scientists' Ranking

Dimitrios Katsaros, Leonidas Akritidis, and Panayiotis Bozanis
Department of Computer & Communication Engineering, University of Thessaly, Volos, Greece.
E-mail: dkatsar@inf.uth.gr; leoakr@uth.gr; pbozanis@inf.uth.gr

Designing fair and unbiased metrics to measure the "level of excellence" of a scientist is a very significant task because they recently also have been taken into account when deciding faculty promotions, when allocating funds, and so on. Despite criticism that such scientometric evaluators are confronted with, they do have their merits, and efforts should be spent to arm them with robustness and resistance to manipulation. This article aims at initiating the study of the *coterminal citations*—their existence and implications—and presents them as a generalization of self-citations and of co-citation; it also shows how they can be used to capture any manipulation attempts against scientometric indicators, and finally presents a new index, the *f* index, that takes into account the coterminal citations. The utility of the new index is validated using the academic production of a number of esteemed computer scientists. The results confirm that the new index can discriminate those individuals whose work penetrates many scientific communities.

Quantifying an Individual's Scientific Merit

The evaluation of the scientific work through scientometric indicators has long attracted significant scientific interest, but recently has become of ground practical and scientific importance. An increasing number of academic institutions are using such indicators to decide faculty promotions, and automated methodologies have been developed to calculate such indicators (Ren & Taylor, 2007). In addition, funding agencies use them to allocate funds, and recently, some governments have considered the consistent use of such metrics for funding distribution. For instance, the Australian government has established the *Research Quality Framework* as an important feature in the fabric of research in Australia, and the United Kingdom government has established the *Research Assessment Exercise* to produce quality profiles for each submission of research activity made by department/institution (http://www.rae.ac.uk).

Received July 9, 2008; revised December 24, 2008; accepted December 24, 2008

 • Published online 4 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21040

The use of such indicators to characterize a scientist's merit is controversial since this assessment is a complex social and scientific process that is difficult to narrow into a single scientometric indicator. In his recent article, David Parnas (2007) described some possible negative consequences to the scientific progress that could be caused by the "publish or perish" marathon run by all scientists and proposed not taking into account the scientometric indicators. Adler, Ewing, and Taylor (2008) depicted several shortcomings of the metrics currently in use; that is, the Impact Factor and the *h* index. The following phrase, attributed to Albert Einstein, could be representative of the opponents of the scientometric indicators: "Not everything that can be counted counts, and not everything that counts can be counted."

Indeed, neither arguments nor applied methodology currently exist to decide which indicators are correct or incorrect, although the expressive and descriptive power of numbers (i.e., scientometric indicators) cannot unthinkingly be ignored (Evidence Report, 2007). As Lord Kelvin stated:

> When you can measure what you are speaking about, and express it in numbers, you know something about it. But when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind.

In the present article, we argue that instead of devaluing the scientometric indicators, we should strive to develop a "correct/complete set" of them and, most importantly, to use them in the right way. Furthermore, this article studies an aspect of the scientometic indicators that has not been investigated in the past and proposes a new, robust scientometric indicator.

The Notion of Coterminal Citations

Traditionally, the impact of a scholar is measured by the number of authored papers and/or the number of citations. The early metrics are based on some form of (arithmetics upon) the total number of authored papers, the total number of citations, the average number of citations per paper, and so on. Due to the power-law distribution followed by these

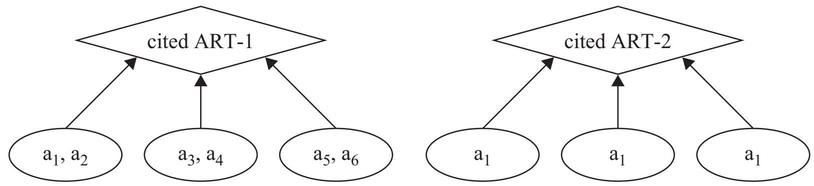

FIG. 1. Citing extremes: No overlap at all (Left); Full overlap (Right).

metrics, they present one or more of the following drawbacks (also see Hirsch, 2005): (a) They do not measure the impact of papers, (b) they are affected by a small number of "big hits" articles, and (c) they have difficulty setting administrative parameters.

Hirsch (2005) attempted to collectively overcome all these disadvantages and proposed the *h* index. The *h* index was really path-breaking, and inspired several research efforts to cure its various deficiencies (e.g., its aging-ignorant behavior) (Sidiropoulos, Katsaros, & Manolopoulos, 2007).

Nevertheless, there is a latent weakness in all scientometric indicators that have been developed thus far, either those for ranking individuals or those for ranking publication fora, and the *h* index is yet another victim of this complication. The inadequacy of the indicators stems from the existence of what we term here—for the first time in the literature—the *coterminal citations*.

With a retrospective look, we see that one of the main technical motivations for the introduction of the *h* index was that the metrics used until then (i.e., total, average, max, min, median citation count) were very vulnerable to self-citations, which in general are conceived as a form of "manipulation." In his original article, Hirsch (2005) specifically mentioned the robustness of the *h* index with respect to self-citations and indirectly argued that the *h* index can hardly be manipulated. Indeed, the *h* index is more robust than are traditional metrics, but it is not immune to them (Schreiber, 2007). Actually, none of the existing indicators is robust to self-citations. In general, the issue of self-citations has been examined in many studies (e.g., Hellsten, Lambiotte, & Scharnhorst, 2007; van Raan, 2008), and the usual practice is to ignore them when performing scientometric evaluations since in many cases they may account for a significant part of a scientist's reputation (Fowler & Aksnes, 2007) and sometimes are used to support promotional strategies (Hyland, 2003).

At this point, we argue that there is nothing wrong with self-citations; in many cases, they can effectively describe the "authoritativeness" of an article (Lawani, 1982), such as in the cases that the self-cited author is a pioneer in his or her field who keeps steadily advancing the field in a step-by-step publishing fashion until other scientists gradually discover and follow his or her ideas. Regardless of the reason that they are being made, self-citations work as a driving force in strengthening the impact of the research (van Raan, 2008).

In the sequel, we will exhibit that the problem is much more complex and goes beyond self-citations; it involves the essential meaning of a citation. Consider, for instance, the citing patterns in Figure 1. ART-1 is cited by three other papers (the ovals), and these citing articles have been authored by (strictly) discrete sets of authors: $\{a_1, a_2\}$, $\{a_3, a_4\}$, and $\{a_5, a_6\}$, respectively. On the other hand, ART-2 is cited by three other papers which all have been authored by the same author: $\{a_1\}$. Note that we make no specific mention about the identity of the authors of ART-1 or ART-2 with respect to the identity of the authors $\{a_i\}$; some of the authors of the citing papers may coincide with those of the cited articles. Our problem treatment is more generic than are self-citations.

While we have no problem accepting that ART-1 has received three citations, we feel that ART-2 has received no more than one citation because, for instance, the heavy influence of ART-2 to author a_1 combined with the large productivity of this author. Nevertheless, considering that authors a_1 to a_6 all have read (Have they?) ART-1 and that only one author has read ART-2, it seems that the former article has a larger impact upon the scientific thinking. On one hand, we could argue that the contents of ART-2 are so sophisticated and advanced that only a few scholars, if any, could even grasp some of the article's ideas. On the other hand, how long could such a situation persist? If ART-2 is a significant contribution, then it would get, after some time, its "right" position in the citation network, even if the scientific subcommunity to which it belongs is substantially smaller that the subcommunity of ART-1.

The situation is even more complicated if we consider the citation pattern appearing in Figure 2 where there exist overlapping sets of authors in the citing papers. For instance, Author a_3 is a coauthor in all three citing papers.

This pattern of citation, where some author has (co)authored multiple papers citing another paper, is in the spirit of what is termed in this article the *coterminal citations*. Coterminal citations can be considered as a generalization of what is widely known as cocitation, and their introduction attempts to capture the "inflationary" trends in scholarly communication which are reflected by coauthorship and "exaggerate" citing (Cronin, 2001, 2003; Cronin, Shaw, & Barre, 2003; Person, Glanzel, & Dannell, 2004).

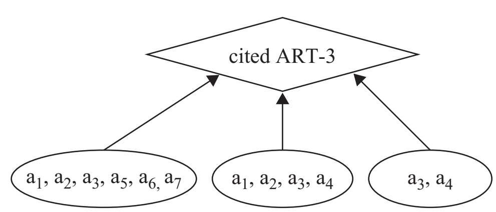

FIG. 2. Citing articles with author overlap.

DOI: 10.1002/asi

Apparently, no prior work exists on dealing with coterminal citations; the closest relevant works include techniques to filter self-citations or weigh multi-author self-citations (Schreiber, 2007; Schubert, Glanzel, & Thijs, 2006) and an early scheme to count cardinalities of citing authors (Lehrl, Kinzel, & Fischer, 1988). Our target is to develop a metric of scientific excellence for individuals that will not be affected by the existence of coterminal citations (i.e., that it will "appropriately" weigh them). We firmly believe that the exclusion of self-citations is not a fair action; neither is any form of ad hoc normalization. Each and every citation has its value—the problem is to quantify this value. The notion of coterminal citations leads naturally to the process of the discovery of their patterns of existence and of their "controlled discount."

The *f* index

We consider the citing example shown in Figure 2 where an article, say A, is cited by three other articles, and let us define the quantity nca^{A} to be equal to the number of articles citing Article A. We define the series of sets $F_i^{\mathrm{A}} = \{a_j\text{: author } a_j \text{ appears in exactly } i \text{ articles citing } A\}$. For Article ART-3, we have that $F_1^{\mathrm{A}} = \{\mathrm{a}_5, \mathrm{a}_6, \mathrm{a}_7\}$\$, $F_2^{\mathrm{A}} = \{\mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_4\}$, $F_3^{\mathrm{A}} = \{\mathrm{a}_3\}$.

Then, we define $f_{\mathrm{i}}^{\mathrm{A}}$ to be equal to the ratio of the cardinality of $F_{\mathrm{i}}^{\mathrm{A}}$ to the total number of distinct authors citing article A; that is, $f_{\mathrm{i}}^{\mathrm{A}} = \frac{|F_i^{\mathrm{A}}|}{total\_number\_distinct\_authors}$. These quantities constitute the coordinates of a nca^{A}-dimensional vector f^{A}, which is equal to $f^{\mathrm{A}} = \{f_1^{\mathrm{A}}, f_2^{\mathrm{A}}, f_3^{\mathrm{A}}, \ldots, f_{nca^{\mathrm{A}}}^{\mathrm{A}}\}$. The coordinates of this vector define a probability mass since $\sum_{i=1}^{nca^{\mathrm{A}}} f_{\mathrm{i}}^{\mathrm{A}} = 1$. For the earlier example of the cited article ART-3, we have that $f^{\text{ART-3}} = \left\{\frac{3}{7}, \frac{3}{7}, \frac{1}{7}\right\}$. Similarly, for the cited article ART-1, we have that $f^{\text{ART-1}} = \left\{\frac{6}{6}, \frac{0}{6}, \frac{0}{6}\right\}$, and for ART-2, we have that $f^{\text{ART-2}} = \left\{\frac{0}{1}, \frac{0}{1}, \frac{1}{1}\right\}$.

Thus, we have converted a scalar quantity (i.e., number of citations that an article has received) into a vector quantity (i.e., f^{A}) which represents the penetration of A's ideas—and consequently of its author(s)—to the scientific community; the more people use a scholar's work, the greater the impact. In general, these vectors are sparse with a lot of zeroes after the first coordinates. The sparsity of the vector reduces for the cited articles that have only a few citations. Naturally, for successful scholars, we would prefer the probability mass to be concentrated to the first coordinates, which would mean that new scientists consistently become aware of and use the article's ideas.

As the probability mass gets concentrated on the coordinates near the end of f^{A}, the "audience" gets narrower, and in some cases, it may imply bad practices (see Parnas, 2007) such as *publishing pacts* (i.e., citation exchange), *clique building* (i.e., researchers form small groups that use jargon to discuss a narrow topic even though it is broad enough to support the existence of a conference/journal and then publish papers "from the clique for the clique"), and practices which lead to papers with *minimum publishable increment* (i.e., after completion of a substantial study, many researchers divide the results to produce as many publishable papers as possible that share a large fraction of citations to the same papers).

Although working with vectors is complicated; we can exploit a "weighting" vector, say s, to convert vector f into a scalar value through a dot-product operation (i.e., $\hat{f} = f \bullet s$). For the moment, we will use the plainest vector defined as $s_1 = \{nca, nca - 1, \ldots, 1\}$; other choices will be presented in the sequel. Thus, for the example article ART-3, we compute a new decimal number characterizing its significance, and this number is equal to $N_f^{\mathrm{A}} = f^{\mathrm{A}} \bullet s_1 = \frac{3}{7} * 3 + \frac{3}{7} * 2 + \frac{1}{7} * 1 = \frac{16}{7} \Rightarrow N_f^{\mathrm{A}} \approx 2.28$.

The f *index*

Now, we can define the proposed f index in a spirit completely analogous to that of h index. To compute the f index of an author, we calculate the quantities $N_f^{\mathrm{A_i}}$ for each one of his or her authored articles $\mathrm{A_i}$ and rank them in nonincreasing order. The point where the rank becomes larger than the respective $N_f^{\mathrm{A_i}}$ in the sorted sequence defines the f index value for that author. The name for that new index comes from the fact that it is fractional citation counting scheme.

The Weighting Vector

Earlier, we used the most simple weighting vector; different such vectors can disclose different facts about the importance of the cited article. Apart from s_1, we also propose a couple of easy-to-conceive versions of the weighting vector. The vector $s_2 = \{nca, 0, \ldots 0\}$ lies at the other extreme of the spectrum with respect to s_1. Finally, if we suppose that the last nonzero coordinate of f^{A} is f_k^{A}, then we have a third version of the weighting version defined as $s_3 = \left\{nca, nca - \frac{nca}{k}, nca - \frac{2*nca}{k}, \ldots, 1\right\}$. For each one of these weighting vectors, we define the respective f index as f_{s_1}, f_{s_2}, and f_{s_3}. None of these three versions of the weighting vector, and consequently of the respective indexes, can be considered superior to the other two. They present merits and deficiencies in different cases. For instance, the f_{s_1} index does not make any difference for large h index values; for scientists with an h index smaller than 15, the obtained f_{s_1} index can be as much as 50% of the respective h index, which can be partially explained by the fact that lower performance (in terms of number of publications) scholars have larger number of self-citations—an explanation which is consistent with the findings of van Raan (2008).

Validation

The validation of the usefulness of the proposed indexes is not an easy task, given our intention not to harm the reputation of any mentioned scientist. We selected as input data to apply our ideas a number of computer scientists with a high h index (http://www.cs.ucla.edu/~palsberg/h-number.html)

DOI: 10.1002/asi

TABLE 1. Computer scientists' ranking (r) based on h index.

r	Scientist–h	r	Scientist–h	r	Scientist–h
1	Hector Garcia-Molina–77	17	Oded Goldreich–48	23	Carl Kesselman–42
2	Jiawei Han–66	17	Philip S. Yu–48	24	Olivier Faugeras–41
3	Ian Foster–65	17	Prabhakar Raghavan–48	25	Teuvo Kohonen–40
4	Robert Tarjan–64	17	Leslie Lamport–48	25	Amit Sheth–40
5	Rakesh Agrawal–62	17	Douglas C. Schmidt–48	25	Craig Chambers–40
6	Jennifer Widom–60	18	Michael I. Jordan–47	25	Demetri Terzopoulos–40
6	Scott Shenker–60	18	Donald E. Knuth–47	25	David A. Patterson–40
7	Jeffrey D. Ullman–59	18	Ronald Fagin–47	25	Philip Wadler–40
8	Deborah Estrin–58	18	Micha Sharir–47	25	Jose Meseguer–40
9	David Culler–56	19	H.V. Jagadish–46	25	George Karypis–40
9	Amir Pnueli–56	19	Mihir Bellare–46	26	Geoffrey E. Hinton–39
10	Richard Karp–55	19	Pat Hanrahan–46	26	Stefano Ceri–39
10	Serge Abiteboul–55	19	Garcia Luna Aceves–46	26	Leonard Kleinrock–39
11	David J. DeWitt–54	20	Michael Franklin–45	26	Saul Greenberg–39
11	David E. Goldberg–54	20	Alex Pentland–45	26	Judea Pearl–39
12	Anil K. Jain–53	20	Martin Abadi–45	26	David Dill–39
13	Hari Balakrishnan–53	20	Andrew Zisserman–45	27	Vern Paxson–38
13	Randy H. Katz–52	20	Thomas A. Henzinger–45	27	John A. Stankovic–38
14	Takeo Kanade–52	20	Vipin Kumar–45	27	Krithi Ramamritham–38
14	Rajeev Motwani–51	20	Nancy Lynch–45	27	Ramesh Govindan–38
15	Don Towsley–50	21	Christos Faloutsos–44	27	Jon Kleinberg–38
15	Christos H. Papadimitriou–50	21	Thomas S. Huang–44	28	Al. Sangiovanni-Vincentelli–37
15	Sebastian Thrun–50	21	Sally Floyd–44	28	Edmund M. Clarke–37
15	Jack Dongarra–50	21	Robin Milner–44	29	Herbert Edelsbrunner–36
15	Ken Kennedy–50	21	Won Kim–44	29	Richard Lipton–36
16	Didier Dubois–49	22	M. Frans Kaashoek–43	29	Ronald L. Rivest–36
16	Lixia Zhang–49	22	Kai Li–43	29	Willy Zwaenepoel–36
16	Michael J. Carey–49	22	Monica S. Lam–43	29	Jason Cong–36
16	Michael Stonebraker–49	22	Sushil Jajodia–43	30	Victor Basili–35
16	Moshe Y. Vardi–49	22	Rajeev Alur–43	30	Mario Gerla–35
16	David S. Johnson–49	23	Raghu Ramakrishnan–42	30	Andrew S. Tanenbaum–35
16	Ben Shneiderman–49	23	Barbara Liskov–42	31	Maja Mataric–33
16	W. Bruce Croft–49	23	Tomaso Poggio–42	32	John McCarthy–32
17	Mihalis Yannakakis–48	23	Victor Lesser–42	32	David Haussler–32
17	Miron Livny–48	23	Joseph Goguen–42	33	Stanley Osher–31
17	Luca Cardelli–48	23	Henry Levy–42	33	Tim Finin–31

who are, without question, top-quality researchers. Since the data provided by this URL are not up to date and contain inconsistencies, we first cleaned them and then kept the scientists with an h index larger than 30.

The ranking in nonincreasing h index is illustrated in Table 1 the rankings with the new indicators f_{s_2} and f_{s_3} appear in Table 2 Both indicators cause changes in the ranking provided by the h index. As expected, the values of the f_{s_2} index are significantly different than the respective h index values. Note that these differences (and their size) appear in any position, independently of the value of the h index. If these differences concerned only the scientists with the largest h index, then we could (safely) argue that for someone who has written many papers and that each paper has received a large number of citations, then some overlap citations and some self-citations are unavoidable. This is not the case though, and it seems that there is a deeper, latent explanation.

Seeking this explanation, we calculated the differences in ranking positions for each scientist when ranked with the h index versus when they are ranked with the f_{s_2}. The results are illustrated in Table 3 and 4. The general comment is that the scientists who climb up the largest number of positions are those whose work can "penetrate" (and thus benefit) large "audiences." For instance, the research results by Lixia Zhang and John A. Stankovic, who now work on sensors, are cited in communities such as databases, networking, and communications. Other scientists whose works are used by large audiences are those working on "computer organization" (e.g., M. Frans Kaashoek, Barbara Liskov, Andrew S. Tanenbaum, etc.). Note here that scientists' age has nothing to do with the ranking relocation since both younger researchers (e.g., Lixia Zhang) can climb up positions just like elder scientists can (e.g., Andrew S. Tanenbaum).

Another important question concerns whether the particular area of expertise of a researcher could help him or her acquire a larger reputation. Undoubtedly, the research area plays its role, but it is not the definitive factor. Consider, for instance, the case of data mining, which is a large area and has attracted an even larger number of researchers. We see that George Karypis has earned four positions in the ranking provided by f_{s_2}. If the area of expertise was the only rational explanation for that, then why is Rakesh Agrawal, who founded the field, among the scientists who lost the most number of positions in the ranking provided by f_{s_2}? The answer

DOI: 10.1002/asi

TABLE 2. Computer scientists' ranking (r) based on f_{s2}. The f_{s3} value also is represented.

r	Scientist–f_{s_2}–f_{s_3}	r	Scientist–f_{s_2}–f_{s_3}	r	Scientist–f_{s_2}–f_{s_3}
1	Hector Garcia-Molina–68–74	17	Donald E. Knuth–41–45	21	Geoffrey E. Hinton–37–37
2	Jiawei Han–57–63	17	Philip S. Yu–41–46	22	Teuvo Kohonen–36–39
2	Ian Foster–57–62	18	Miron Livny–40–45	22	Andrew Zisserman–36–41
3	Robert Tarjan–56–61	18	Luca Cardelli–40–46	22	Sushil Jajodia–36–41
4	Scott Shenker–54–59	18	Ronald Fagin–40–45	23	Joseph Goguen–35–40
5	Jennifer Widom–53–58	18	H. V. Jagadish–40 –44	23	Rajeev Alur–35–41
5	Jeffrey D. Ullman–53–55	18	Didier Dubois–40–44	23	Philip Wadler–35–38
6	David Culler–52– 53	18	Alex Pentland–40–43	23	Amit Sheth–35–39
7	Deborah Estrin –51–56	18	Thomas S. Huang–40–42	23	Nancy Lynch–35–42
7	Rakesh Agrawal–51–60	18	Sally Floyd–40–43	23	Leonard Kleinrock–35–38
8	David E. Goldberg–50–52	18	Robin Milner–40–42	23	Vern Paxson –35–37
9	Richard Karp–49–55	18	M. Frans Kaashoek–40–41	23	John A. Stankovic–35–37
10	David J. DeWitt–48 –51	18	Carl Kesselman–40–42	24	Saul Greenberg–34–37
10	Hari Balakrishnan–48 –52	19	Moshe Y. Vardi–39 –46	24	Stefano Ceri–34–37
11	Anil K. Jain–47–50	19	Martin Abadi–39–43	24	Raghu Ramakrishnan–34–40
11	Amir Pnueli –47–52	19	Christos Faloutsos–39–43	24	Krithi Ramamritham–34–38
11	Takeo Kanade–47–50	19	Mihalis Yannakakis–39–46	24	Jon Kleinberg–34–36
12	Randy H. Katz–46 –51	19	Mihir Bellare–39–45	25	Ramesh Govindan–33 –36
12	Lixia Zhang–46–48	19	Oded Goldreich–39–45	25	Edmund M. Clarke–33 –34
13	Don Towsley–45–49	19	Garcia Luna Aceves–39–43	26	Judea Pearl–32 –36
13	Serge Abiteboul–45–52	19	Kai Li–39–41	26	Richard Lipton–32–35
13	David S. Johnson–45–48	19	Barbara Liskov–39–40	26	Ronald L. Rivest–32–34
14	Ken Kennedy–44–49	19	Tomaso Poggio–39–41	26	Victor Basili–32 –35
14	Rajeev Motwani–44–48	19	Henry Levy–39–40	26	Andrew S. Tanenbaum–32–34
14	Sebastian Thrun–44–48	19	Michael Franklin–39–42	26	David Haussler–32–34
14	Ben Shneiderman–44–48	20	Won Kim–38–42	27	Jose Meseguer–31–37
14	Prabhakar Raghavan–44–46	20	Monica S. Lam–38–42	27	David Dill–31–35
15	W. Bruce Croft–43–46	20	Vipin Kumar–38–41	27	Willy Zwaenepoel–31 –34
15	Christos H. Papadimitriou–43–47	21	Victor Lesser–37–41	29	Al. Sangiovanni-Vincentelli–30–34
15	Michael I. Jordan–43–46	21	Thomas A. Henzinger–37–43	28	Mario Gerla–30–33
16	Michael Stonebraker–42 –45	21	Micha Sharir–37–43	29	Herbert Edelsbrunner–29–34
16	Jack Dongarra–42–48	21	Olivier Faugeras–37–40	29	Tim Finin–29–30
16	Leslie Lamport–42–45	21	Craig Chambers–37–40	30	Jason Cong–28–33
16	Douglas C. Schmidt–42–46	21	Demetri Terzopoulos–37 –38	31	Maja Mataric–27–30
16	Michael J. Carey–42–46	21	David A. Patterson–37–39	31	Stanley Osher–27–31
16	Pat Hanrahan–42–44	21	George Karypis–37–38	32	John McCarthy–26–29

TABLE 3. Largest relocations with respect to rank position: Most positions up.

Scientist-*h*	*h* rank	Earned position in f_{s_2}
David Haussler	32	+6
Carl Kesselman	23	+5
Geoffrey E. Hinton	26	+5
Lixia Zhang	16	+4
M. Frans Kaashoek	22	+4
Barbara Liskov	23	+4
Tomaso Poggio	23	+4
Henry Levy	23	+4
Craig Chambers	25	+4
Demetri Terzopoulos	25	+4
David A. Patterson	25	+4
George Karypis	25	+4
Vern Paxson	27	+4
John A. Stankovic	27	+4
Victor Basili	30	+4
Andrew S. Tanenbaum	30	+4
Tim Finin	33	+4

TABLE 4. Largest relocations with respect to rank position: Most positions down.

Scientist–*h*	*h* rank	Lost position in f_{s_2}
Rakesh Agrawal–62	5	−2
Amir Pnueli–56	9	−2
Didier Dubois–49	16	−2
Mihalis Yannakakis–48	17	−2
Oded Goldreich–48	17	−2
Andrew Zisserman–45	20	−2
Jose Meseguer–40	25	−2
Serge Abiteboul–55	10	−3
Moshe Y. Vardi–49	16	−3
Micha Sharir–47	18	−3
Nancy Lynch–45	20	−3

lies in the particularities of the research subfields; George Karypis contributed some very important results useful also in the field of bioinformatics. To strengthen this, consider the case of Jiawei Han. He is a data-mining expert whose work penetrates to communities such as mining, databases, information retrieval, and artificial intelligence, and he is ranked second, based either on *h* index, or on f_{s_2} or on f_{s_3}.

Examining the scholars with the largest losses, we see that scientists who have made groundbreaking contributions and offered some unique results (e.g., Mihalis Yannakakis

DOI: 10.1002/asi

and Moshe Y. Vardi) drop in the ranking provided by the f_{s_2}. This has nothing to do with the theoretical versus the practical sides of computer science; contrast the cases of M. Yannakakis and M. Vardi versus A. Zisserman and R. Agrawal. It is due to the nature of the scientific results that do not “resound” to other communities.

Discussion

We describe, for the first time here, another dimension of publication methodologies—the existence of *coterminal citations*—and set forth an effort to discover such patterns in citation networks. The proposed f index represents a computerized, automated way to assign weigh/value to citations, although it is not alone sufficient to determine the function and value of citation; for instance, their cognitive background also should be taken into consideration (Garfield, 1964).

The astute reader will have realized by now that in our efforts to recognize and weigh coterminal citations, we have in our arsenal the research works dealing with Web link spam (Gyongyi & Garcia-Molina, 2005) (e.g., TrustRank, BadRank, etc.). Unfortunately, the situation is radically difficult in citation networks because they consist of entities richer than the Web pages and the Web links encountered in Web spam. Each node (i.e., a citing article) in a citation network consists of entities (i.e., coauthors) which form a complex overlay network above the article citation network.

We believe that the detection and weighing of coterminal citations in citation networks is a quite difficult procedure, and the cooperation of the authors is mandatory. Maybe the scientific community should set some rules about citing—rules not only ethical but practical as well (e.g., enforced by the reviewers). For instance, we could qualify each reference in the “References” section of every published article, to describe which citations involve only relevant work (with the qualifier: CONTEXT), which citations refer to earlier work done by the authors of the article (with the qualifier: SELF), which citations refer to works implemented earlier as competing works in the article (with the qualifier: EXTENT), and so on. Apart from these categories, others could be devised as well; whether the citing article’s results contradict or support the results of the cited articles (with the qualifier: OPPOSE or EXTENT) and many other along the same vein. Ideas similar to this was described by Elisabeth Duncan and colleagues in 1981 (project proposal) and by Rauter (2006).

In any case, we believe that scientometric indicators are not a panacea, and we should do further research before applying a set of them to characterize the achievements of a scholar. Indicators do have their significance, but some methodologies, both ethical and practical, should change for reliable and automated measurements of science.

Acknowledgments

The authors thank Christos Kapoutsis for suggesting the term “coterminal citations” and providing insightful comments to earlier versions of the article, and they also thank the anonymous reviewers, whose comments have greatly improved the present article.

References

Adler, R., Ewing, J., & Taylor, P. (2008). Joint committee on quantitative assessment of research: Citation statistics. Australian Mathematical Society Gazette, 35, 166–188.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices. Journal of the American Society for Information Science and Technology, 52, 558–569.

Cronin, B. (2003). Scholarly communication and epistemic cultures. New Review of Academic Librarianship, 9, 1–24.

Cronin, B., Shaw, D., & Barre, K.L. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. Journal of the American Society for Information Science and Technology, 54, 855–871.

Evidence Report. (2007). The use of bibliometrics to measure research quality in the UK higher education system [Report produced for the Research Policy Committee of Universities, United Kingdom, by Evidence Ltd.]. London: Higher Education Funding Council for England.

Fowler, J.H., & Aksnes, D.W. (2007). Does self-citation pay? Scientometrics, 72, 427–437.

Garfield, E. (1964). Can citation indexing be automated? In M.E. Stevens, V.E. Giuliano, & L.B. Heilprin, (Eds.), Proceedings of the Symposium on Statistical Association Methods for Mechanized Documentation (pp. 189–192). Washington, DC: U.S. Government Printing Office.

Gyongyi, Z., & Garcia-Molina, H. (2005). Spam: It’s not just for inboxes anymore. IEEE Computer, 38, 28–34.

Hellsten, I., Lambiotte, R., & Scharnhorst, A. (2007). Self-citations, co-authorships and keywords: A new approach to scientists’ field mobility. Scientometrics, 72, 469–486.

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

Hyland, K. (2003). Self-citation and self-reference: Credibility and promotion in academic publication. Journal of the American Society for Information Science and Technology, 54, 251–259.

Lawani, S.M. (1982). On the heterogeneity and classification of author self-citations. Journal of the American Society for Information Science, 33, 281–284.

Lehrl, S., Kinzel, W., & Fischer, B. (1988). Der Science Impact Factor. In v.H.-D. Daniel & R. Fisch, (Eds.), Evaluation von Forschung (pp. 291–305). Konstanz, Germany: Universitaetsverlag.

Parnas, D.T. (2007). Stop the numbers game: Counting papers slows down the rate of scientific progress. Communications of the ACM, 50, 19–21.

Person, O., Glanzel, W., & Dannell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. Scientometics, 60, 421–432.

Rauter, J. (2006). Zitationsanalyse und Intertextualitaet. Hamburg, Germany: Kovac.

Ren, J., & Taylor, R.N. (2007). Automatic and versatile publications ranking for research institutions and scholars. Communications of the ACM, 50, 81–85.

Schreiber, M. (2007). Self-citation corrections for the Hirsch index. Europhysics Letters, 78(3), article number 30002.1-30002.6.

Schubert, A., Glanzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. Scientometrics, 67, 503–514.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics, 72, 253–280.

van Raan, A.F.J. (2008). Self-citation as an impact-reinforcing mechanism in the science system. Journal of the American Society for Information Science and Technology, 59, 1631–1643.