Journal of Informetrics 4 (2010) 392–406

Contents lists available at ScienceDirect

Journal of Informetrics

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

Citing-side normalization of journal impact: A robust variant of the Audience Factor

Michel Zitt[a,b,*]

[a] INRA-Lereco, Nantes, France
[b] Observatoire des Sciences et des Techniques (OST), Paris, France

ARTICLE INFO

Article history:
Received 2 February 2010
Received in revised form 10 March 2010
Accepted 11 March 2010

Keywords:
Audience factor
Impact factor
Normalized impact
Citing-side normalization
Source normalization
Journal impact

ABSTRACT

The principle of a new type of impact measure was introduced recently, called the "Audience Factor" (AF). It is a variant of the journal impact factor where emitted citations are weighted inversely to the propensity to cite of the source. In the initial design, propensity was calculated using the average length of bibliography at the source level with two options: a journal-level average or a field-level average. This citing-side normalization controls for propensity to cite, the main determinant of impact factor variability across fields. The AF maintains the variability due to exports–imports of citations across field and to growth differences. It does not account for influence chains, powerful approaches taken in the wake of Pinski–Narin's influence weights. Here we introduce a robust variant of the audience factor, trying to combine the respective advantages of the two options for calculating bibliography lengths: the classification-free scheme when the bibliography length is calculated at the individual journal level, and the robustness and avoidance of ad hoc settings when the bibliography length is averaged at the field level. The variant proposed relies on the relative neighborhood of a citing journal, regarded as its micro-field and assumed to reflect the citation behavior in this area of science. The methodology adopted allows a large range of variation of the neighborhood, reflecting the local citation network, and partly alleviates the "cross-scale" normalization issue. Citing-side normalization is a general principle which may be extended to other citation counts.

1. Introduction

This article introduces a robust variant of the "Audience Factor", the principles of which were delineated in Zitt and Small (2008a). Audience factor is a weighted form of Garfield's Journal Impact Factor (JIF), addressing field-discrepancies by a citing-side normalization, in contrast both with post-facto field-normalization and influence measures.

Garfield's Journal Impact Factors (JIF, Garfield, 1972) released by ISI (now Thomson Reuters) in the "Journal Citation Reports (JCR®)" had a tremendous effect on the scientific community. Technical features, uses and misuses in evaluation context have been discussed in a large literature, from many vantage points (Seglen, 1997; Garfield, 1998; Moed, 2002; Glänzel & Moed, 2002; Bordons, Fernandez, & Gomez, 2004; Nicolaisen, 2007, Chap. 13). Many other variants of JIF were proposed (see for example Rousseau, 2005). A lasting issue is the across-fields inequality due to the field dependence of reference and citation counts (e.g., Murugesan & Moravcsik, 1978). The usual response is the field-normalization of JIF, often based on ISI subject categories (Czapski, 1997; Marshakova-Shaikevich, 1996; Pudovkin & Garfield, 2004; Ramirez,

* Correspondence address: INRA-Lereco, Rue de la Géraudière, BP 71627, 44316 Nantes Cedex 03, France. Tel.: +33 0240675173, fax: +33 0240675074.
E-mail address: zitt@nantes.inra.fr.

1751-1577/$ – see front matter
doi:10.1016/j.joi.2010.03.004

Garcia, & del Rio, 2000; Schubert & Braun, 1996; Sen, 1992; Vinkler, 2002; for a general view on normalization see Radicchi, Fortunato, & Castellano, 2008). Taking fields as a basis makes the process classification-dependent, both in terms of level of field-delineation and classification level.

Next to JIF, a milestone in the impact studies took place with the seminal work of Pinski & Narin's on the "Influence weight" of journals (1976), developing the point of view that all citations are not equal and must be weighted by the prestige (impact) of the source. The treatment of iteration has been recently renewed by the Google-type algorithms (Bergstrom, 2007; Bollen, Rodriguez, & Van de Sompel, 2006; Moya-Anegon, 2007; Palacio-Huerta & Volij, 2004). A few websites now display second-generation influence measures (Bergstrom's eigenfactor[1], Moya-Anegon's SCImago journal rank[2] based on Scopus, Lim's Red Jasper journal rank[3], etc.).

The ratio of citable to citing literature, which determines the inequality of average citation rates among fields, depends on several factors. In a model configuration of isolated fields, the citing propensity of the field depends on the average length of bibliographies; growth of the area; and the obsolescence rate of citations. As soon as fields are not insulated but exchange citations, those transactions, assumed to mirror exchanges of knowledge/information, play an important role in shaping impacts, in favor of generic research areas exporting knowledge/information. We will not develop this background in the present paper.

Garfield addressed the issue long ago (Garfield, 1976), with a special attention to the stability issues in long-term growth of literature. The relation of literature length and impact factor in fields has been noticed recurrently, with analytical (Vinkler, 2004) and empirical studies (on recent data Biglu, 2008). Witnessing the important role of referencing behavior, we called for an exploration of citing-side normalization (e.g. Zitt, Ramanana-Rahary and Bassecoulard, 2005) and proposed a first application at the journal level with the version 0.1 of "Audience Factor" where citations are weighted, on the citing side, by the propensity to cite of the source journal (Zitt & Small, 2008). Technically, the solution picked echoes the studies of fractional counting of citations in other contexts (for example Small & Sweeney, 1985). We mentioned in this article that the principle was general and could be extended to various citation analyses.

Depending on their implementation, "Influence measures" may also account for the citing propensity of sources. However, the control of propensity is combined with the chains of intellectual influence which are the main concern of these measures. In contrast, the purpose of the audience factor, developed at first for exploratory purposes, was to control only for citing propensity and immediacy, leaving the across-areas exchanges fully reflected in the measure. In this citing-side normalization leading to the "audience measure" the normalization factor for emitted citations is the average length of the reference list having a variety of possible definitions. In the first version presented in our 2008 article (say v0.1), two options were offered, either at the individual citing journal level or at the field level by averaging the length on all citing journals in the field.

We describe here a significant improvement of the "Audience Factor", termed v0.2, by finding a trade-off between the two forms initially investigated. Taking individual citing journals as the basis leads to a classification-free process but lacking robustness in limit cases. To avoid giving an excessive weight to journals with very low citing propensity that are often marginal ones for scientific production, such as trade journals, an ad hoc treatment of distribution tails had to be introduced. If instead we take the citing field as the basis, we get a quite robust process but losing a great advantage, the classification-independence of the scheme: the weighting depends on the particular field breakdown and lacks uniqueness in the case of multi-assignment journals. A third way to define the citing context (Source Normalized Impact per Paper – SNIP: Moed, 2009) is based on the pool of citing documents.

In the present work, we use the relative neighborhood of each citing journal as a proxy of the research community on which we assess the propensity to cite. This process combines the advantages of both previous variants (classification-free and robustness), without their shortcomings. The methodology and sources are presented in Section 1, the results in Section 2, before discussion and conclusion.

2. Methodology and sources

2.1. Original audience factor: definitions and properties

The basic definition of the audience measure AF relies on a weighting of emitted citations inversely to the average length of bibliographies in the source, either strictly at the journal level (citations from the journal are weighted in inverse proportion of the bibliography length in the journal: v0.1a) or at the field level (citations from the journal are weighted in inverse proportion of the bibliography length in the field where the journal belongs: v0.1b). Let us recall the definition of AF v0.1a:

Consider a source journal $i(t)$ as a citing journal from some year t with references to other journals extending back in time as listed in the JCR. Let m_i (t,T) be the average number of references emitted by articles from $i(t)$ in time window $T(t) = [t-1\ldots t-T]$. This window defines the "active" references of $[i]$, other references being ignored. Now weight each

1 http://www.eigenfactor.org.
2 http://www.scimagojr.com.
3 http://www.journal-ranking.com.

emitted reference by a function of m_i (t,T):

$$w_i(t,T) = \frac{m_s(t,T)}{m_i(t,T)}$$

where $m_s(t,T)$ is the average number of active references by articles in all source journals (all science).

The definition of the audience factor AFT(t) follows that of impact factor IFT(t), except that weighted cites, instead of original cites, are used in the numerator. Note that the standard Journal Impact Factor IFT(t) is obtained by forcing $w_i = 1$.

$$\mathrm{AFT}_j(t) = \frac{\Sigma i[w_i\, c_{ij}(t,T)]}{a_j(t,T)}$$

where c_{ij} denotes the cites from journal i (emitted year t) to journal j (received by its articles dated within window T) and $a_j(t,T)$ the number of articles of journal j in window T.

The basic properties of AF v0.1a are its independence (1) from the field propensity to cite (2) from the field citation speed (3) from the field-classification schemes. The variant v0.1b of AF where bibliography length is averaged at the field level is dependent from field-classification schemes but is more robust.

2.2. New variant (0.2): weighting scheme based on neighborhood

We investigate here a pathway to keep the best of the two variants, by using as a normalization factor the average length of reference lists in a specific neighborhood $N(i)$ of the citing journal i in the citation network. The hypothesis is that the citation neighborhood of a (citing) journal represents a scientific community with specific citing habits. Considering a neighborhood for each citing journal seems more appealing than defining bibliometric clusters of micro-disciplines, a process vulnerable to border effects and also conveying biases of the particular clustering method. We add the requirement (see further for details) that the neighborhood reflects the properties of the local citation network. There is an abundant literature on neighborhoods in social networks in general, for example recently by Jackson and Rogers (2007).

How to define this neighborhood of a (citing) journal i? Many studies have addressed the proximity of journals since Narin (1976), for example for journals classification purposes Cozzens and Leydesdorff (1993), Bassecoulard and Zitt (1999), before the current boom of science mapping. In a classical perspective (see for example Small, 1995) one may think of a set of journals citing the journal i; a set of journals cited by the journal i; or a set created by co-citation or coupling linkage. The first option may retrieve too small a number of neighbors, especially if the journal i has a low impact. The set of journals cited by i are more likely to represent a reliable neighborhood, recognized as an intellectual base of i – for example trade journals are not likely to be heavily represented in this set. In the following, "neighborhood" of a journal refers to this dissymmetrical notion of "cited neighborhood" (outflows of citations from i).

The strength of linkage is here defined as $P(ij)$ = cit(ij)/sqrt[$c(i.)c(.j)$] where cit(ij) denotes the citations from i to j, $c(i.)$ the total citations emitted by i, $c(.j)$ the total citations received by j. This index is built after the model of the Ochiai index[4], but it does not obey the axiom of symmetry, since it is oriented in the direction of the citation flow: in general $P(ij) \neq P(ji)$. We use the measure only for the purpose of ranking the neighbors. The 1-neighborhood(i) is defined here as the set of $n \leq N$ journals j closest to the journal i on this criterion, N is a preset threshold, which is the only parameter of the process. Lower-rank neighbors are ignored. N is not necessarily reached, depending of actual citation outflows of i.

Let us now consider the neighborhood of a particular journal j belonging to the 1-neighborhood(i): the set of $n \leq N$ journals k connected to the journal j. The 2-neighborhood(i) is defined by the union, for all j, of all 1-neighborhood(j). A given journal k may appears several times (up to N) in the 2-neighborhood of i. The frequency of multiples is a function of the local density of the network. In a super-dense configuration, the local clustering coefficient is high and the set of neighbors of every j strongly overlap. As a result, in terms of distinct journals, the cardinal number of 2-neighborhood is equal or hardly superior to the cardinal number of the 1-neighborhood. In contrast, in an extremely open configuration, each journal j may have a different environment, and the 2-neighborhood may have a cardinal number close to N^2.

Hence we propose to take the 2-neighborhood as a basis for normalization, with full account of duplicates. If the journal k appears only once in the 2-neighborhood(i), it receives a weight of $1/T$ if T is the total number of neighbors (*with* duplicates). If the journal k duplicates in x neighbor sets of j journals (say x sets), it receives a weight of x/T. In other words, x is the inflow degree of k for the level-2 links. The maximum value of x is N. Note that, if only through self-citations, the 2-neighborhood often contains the 1-neighborhood in real configurations. Now we calculate a central value of the bibliography lengths for the 2-neighborhood, and this value is assigned to the journal i. Among several possible choices of the central value we picked the median calculated on the distribution keeping duplicates, in order to advantage 2-neighbors appearing twice or more. The measure is relatively robust towards atypical neighbors, such as trade journals with scarce lists, and journals of reviews at the other end.

---

[4] For two overlapping sets, $\mathrm{Ochiai}(i,j) = \mathrm{cooc}(ij)/\sqrt{(n(i).n(j))}$ (Ochiai, 1957); the counterpart in an additive form is the Jaccard index. In case of citations transactions, adaptations can be made either to reflect bilateral flows (Bassecoulard & Zitt, 1999) or unilateral flows, as done here.

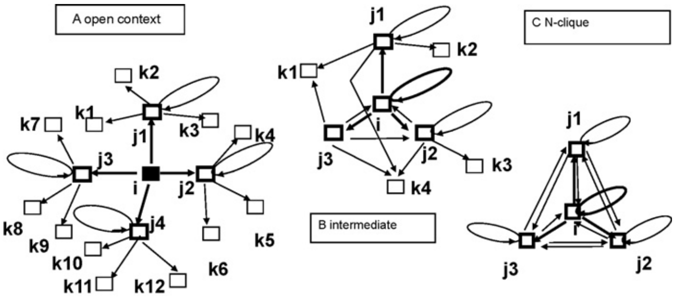

Fig. 1. Neighborhood structure: examples. 1-neighborhood: thick arrows, boxes with thick borders. 2-neighborhood: thin arrows, boxes with thin borders.

This weighting of duplicates, along with the variability of the list of neighbors, typically in the range N to N^2, reflects a rationale of area delineation: in a close system, a quasi-clique configuration, it makes sense to consider a narrow area for normalization. In an open system, a multidisciplinary journal for example, such a restriction would be pointless and a larger context should be accepted. Fig. 1 shows the two extreme and an intermediary configurations for $N=4$, each journal with N neighbors, including possibly itself. Extreme cases yield equal-weight schemes. The weighting scheme corresponding to Fig. 1 for the calculation of the references list length used for the journal i is shown Table 1.

The process also neutralizes a limited number of residual journals (less than 12 in our experiment) with poor neighborhoods. Neutralized journals do not participate in determining the length of reference lists used for the area of any journal i.

In our implementation, the neighborhood N is addressed in terms of number of journals rather than number of papers. Firstly, the granularity is the journal level – and not individual citing articles. Secondly, no additional weight scheme for the size of neighbor journals is considered, but this option could be easily changed. The journal level granularity limits fluctuations, but a source of noise may arise from multidisciplinary or large-scope journals, depending on how they are treated in the database (with or without sections), but in most cases these effects are neutralized by the robust design of weighted medians.

Finally the central value of bibliography length in the neighborhood $x_i(t,T)$ is substituted to the journal reference list in this variant of the audience factor. Weight $w_i(t, T)$ is rewritten:

$$w_i(t, T) = k(t, T).\frac{x_s(t, T)}{x_i\,(t, T)}$$

where $k(t,T)$ is a correcting factor equalizing the total flow of citations in the IF and the AF cases. The constraint we wish to respect is the scale correspondence between AF and IF calculated on the same data, typically the equality between the weighted averages of AF and IF at the level of all science.Fine tuning may be necessary depending on the fine-grain delineation of citing and cited sets, for example non-significant sources.

Table 1
Weighting scheme of Fig. 1 based on the number of inflows level 2.

[Journal]	(A) Open context	(B) Intermediate	(C) Clique
I	$0/T$[a]	$2/T$	$3/T$
$j1$	$1/T$	$1/T$	$3/T$
$j2$	$1/T$	$2/T$	$3/T$
$j3$	$1/T$	$0/T$	$3/T$
$j4$	$1/T$	–	–
$k1$	$1/T$	$2/T$	–
$k2$	$1/T$	$1/T$	–
$k3$	$1/T$	$1/T$	–
$k4$	$1/T$	$3/T$	–
$k5$	$1/T$	–	–
$k6$	$1/T$	–	–
$k7$	$1/T$	–	–
$k8$	$1/T$	–	–
$k9$	$1/T$	–	–
$k10$	$1/T$	–	–
$k11$	$1/T$	–	–
$k12$	$1/T$	–	–
Total neighbors level 2 (with duplicates) or total number of inflows level 2: T	16	12	12
Distinct neighbors level 2	16	7	4
Distinct neighbors level 1	4 (not i)	4	4

[a] Zero weight since i belongs to its own 1-neighborhood but not to its own 2-neighborhood (no inflow). Same configuration for $j3$ in column B.

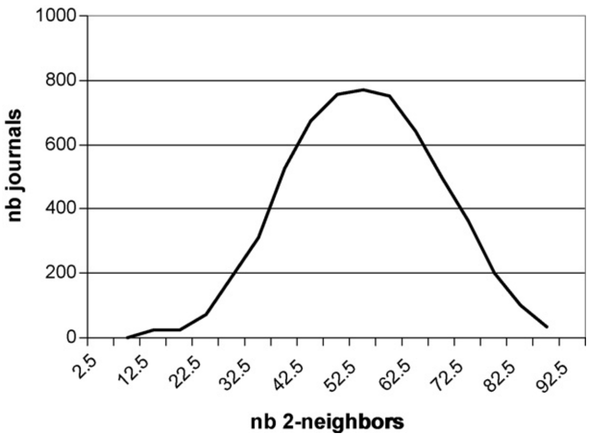

Fig. 2. Distribution of journals by 2-neighborhood size for $N = 10$. Class [2.5] and beyond: 8 journals; class [92.5] and beyond: 11.

The distribution of the number of neighbors by citing journal (Fig. 2) is quite close to a normal distribution of mean 52.1 and StdDev 14.2.

Other statistics of the distribution ($n = 5967$) are median = 52; mode = 47; StdDev = 14.2; min = 1; max = 94. Note that the granularity for $N = 10$ is slightly coarser in average than ISI subject categories, and a specific neighborhood for each journal allows quite large variations.

2.3. *Current data sources and limitations*

The particular JCR file for 2006, used with Thomson Reuters' permission, contains the transaction from ca. 8500 "source journals" (emitting citations) towards ca. 6200 journals, assigned to SCI expanded (some of them with a second affiliation to SSCI-A&HCI). For those 6200, the files also contain full information on bibliography length. For the remaining journals, ca. 2300, among them many journals from SSCI/A&HCI we do not have the latter information, necessary for a precise calculation of the audience factor. The latter set of 2300 journals contains many titles that are assigned uniquely to fields in social science and humanities (SSCI-A&HCI), many sources in computer science, and miscellaneous literature. The neighborhood of citing sources could be calculated on ca. 6000 journals (5967).

The current dataset creates limitations in the calculation of the audience factor for SCI journals, especially for those journals in the main set which receive many citations from the extension of the source set (the 2300 journals). This relates to for example journals bi-affiliated to SCI and SSCI-A&HCI, such as *Econometrica* or *Scientometrics*. Citations from the extension may be counted, but the information for calculating an exact weight of those extra-sources is missing. For each cited journal A, the audience factor is then calculated by extrapolating to citing weight of those journals (citing A) with complete information, to other sources (citing A) with partial information. Hence we applied to the sources the average value found for other sources of citations to the journal whose we wish to calculate the AF. The proportion of extra-sources can be too large for getting a precise value of AF (case of *Econometrica*) or acceptable (case of *Scientometrics*).

Another limitation comes from the aggregation of tail data in the JCR. For example, in the table by cited journal, a "*all others*" category aggregates, for each cited journal, all citations from citing journals with low contribution. The contents of the category in terms of individual journals are not known. Those data are also treated as missing values estimated by the same extrapolation process as above. The third type of data replaced by the extrapolated values concern citing journals with too scarce a neighborhood.

Using a rule of thumb, all journals where the ratio of all sources to completely known sources is inferior to a threshold 2.6 were considered significant enough. This problem is related to the incompleteness of our working file that will be remedied in future implementations. However, this stresses the fact that the required information on bibliographies should be available on sources. It raises, more generally, the issue of database coverage and delineation of citing and cited sets. Chiefly because of these limitations of the working file, the figures presented here should still be considered as experimental.

Another limitation of the results is not due to data, but to the particular choice of taking a 5-year citation window to ensure reliability. Journals not present over the period were ignored. Future implementations will offer several windows. We found in previous experiments that, as expected, the correction brought by AF 2 years over IF 2 years is stronger than on the 5 years window.

Table 2A
AF top rank journals (1–50).

Journal (by AF)	AF	SRC	RK
Reviews of Modern Physics	47.10	***	1
New England Journal of Medicine	35.59	***	2
CA-A Cancer Journal for Clinicians	33.09	***	3
Annual Review of Immunology	30.90	***	4
Materials Science & Engineering R-Reports	27.89	***	5
Science	26.61	***	6
Physiological Reviews	26.42	***	7
Nature Reviews Cancer	25.64	***	8
Nature	24.88	***	9
Chemical Reviews	24.41	***	10
Annual Review of Biochemistry	24.20	***	11
Progress in Materials Science	22.95	***	12
JAMA-Journal of the American Medical Association	21.67	***	13
ACM Computing Surveys	21.57	*	14
Nature Reviews Molecular Cell Biology	20.62	***	15
Nature Medicine	20.55	***	16
Annual Review of Neuroscience	20.36	***	17
Nature Reviews Immunology	19.75	***	18
Lancet	19.04	***	19
Advances in Physics	18.80	***	20
Nature Reviews Neuroscience	18.80	***	21
Surface Science Reports	18.61	***	22
Endocrine Reviews	18.57	***	23
Pharmacological Reviews	17.51	***	24
Nature Genetics	17.14	***	25
Annual Review of Fluid Mechanics	17.01	***	26
Trends in Ecology & Evolution	16.64	***	27
Progress in Polymer Science	16.61	***	28
Nature Biotechnology	16.54	***	29
Cell	16.33	***	30
Nature Reviews Genetics	16.25	***	31
Behavioral and Brain Sciences	16.17	**	32
Annual Review of Cell and Developmental Biology	15.92	***	33
Annual Review of Ecology Evolution and Systematics	15.72	***	34
SIAM Review	15.48	**	35
Physics Reports-Review Section of Physics Letters	15.46	***	36
Nature Immunology	15.36	***	37
Accounts of Chemical Research	15.07	***	38
Annual Review of Pharmacology and Toxicology	14.58	***	39
Annual Review of Psychology	14.10	*	40
Reviews of Geophysics	14.02	***	41
Clinical Microbiology Reviews	13.86	***	42
Mis Quarterly	13.76	**	43
Microbiology and Molecular Biology Reviews	13.56	***	44
Annual Review of Materials Research	13.56	***	45
Annual Review of Plant Biology	13.25	***	46
Annual Review of Entomology	13.18	***	47
Annual Review of Astronomy and Astrophysics	13.12	***	48
Annual Review of Microbiology	12.60	***	49
IEEE Transactions on Evolutionary Computation	12.55	**	50

ns journals not represented. Ratio SRC of all sources and complete sources: ***if SRC < 0.2; **if 0.2 ≦ SRC < 1.9; *if 1.9 ≦ SRC < 2.6; ns if SRC ≦ 2.6.

3. Results

A few results based on the first version of AF, v0.1a, reported by Zitt and Small (2008, 2008b), were as follows:

- AF efficiently removed the inter-field variance effect, with a clear upgrading, for example, of the position of mathematics journals.
- The correlation with IF[5] is fairly strong, due to the within-field inequality which represents a high proportion of the total variance. As effects of imports–exports of knowledge are kept in AF (as in IF, in contrasted with usual FNIF: field-normalized impact factor), fundamental biology for example, positioning as an exporter of knowledge to medicine journals, keeps high levels of visibility, although corrected downward by AF.

[5] AF and IF are based on 5-year window. The IF and AF are calculated in parallel with the same options, and the IF used here may slightly differ from the standard Thomson Reuters version, now released on a 5 years basis.

- In order by decreasing variability, we found IF, AF, FNIF. The Impact Factor fully reflects across-field and within-field variances, the FNIF retains only the within-field part of variance; the Audience Factor stands somewhere between. In AF a large part of across-field variance (but the import–export effect) is removed along with across-area variance. Within-field variance is retained by and large.

3.1. Overview

To a large extent we use the same framework as the previous one (2008) for the presentation of results. We keep in the result tables only those journals receiving their citations mainly from the central set of journals – as to the examples above, this excludes *Econometrica* but not *Scientometrics* and also quite a few journals in computer science. The results are not strikingly different from the previous ones, but slight differences can be found.

Table 2 presents the top ranked journals on AF. The table is limited to journals with a good proportion of sources with sufficient information. A few non-significant journals would take place in the top rank, such as *ACM Transactions on Information Systems* and *VLDB Journal* in computer science, and *Psychological Review* and *Econometrica* in Social Sciences and Humanities.

Table 2B
AF top rank journals (51–100).

Journal (by AF)	AF	SRC	RK
Annual Review of Biomedical Engineering	12.44	**	51
Annals of Internal Medicine	12.33	***	52
Annual Review of Physical Chemistry	12.26	***	53
Archives of General Psychiatry	12.11	**	54
Reports on Progress in Physics	12.08	***	55
Psychological Review	11.79	*	56
Annual Review of Physiology	11.65	***	57
Annual Review of Biophysics and Biomolecular Structure	11.52	***	58
Chemical Society Reviews	11.46	***	59
Journal of the National Cancer Institute	11.37	***	60
Bulletin of the American Mathematical Society	11.20	**	61
Nature Neuroscience	11.05	***	62
Annual Review of Genetics	10.74	***	63
Nature Cell Biology	10.60	***	64
Journal of the American Mathematical Society	10.53	**	65
Annual Review of Nutrition	10.51	**	66
Nano Letters	10.50	***	67
Journal of Clinical Investigation	10.48	***	68
Progress in Energy and Combustion Science	10.23	***	69
Journal of the ACM	10.19	*	70
IEEE Transactions on Pattern Analysis and Machine Intelligence	10.13	*	71
Catalysis Reviews: Science and Engineering	10.07	***	72
FEMS Microbiology Reviews	9.98	***	73
Advanced Materials	9.79	***	74
Progress in Lipid Research	9.76	**	75
Trends in Neurosciences	9.73	***	76
Progress in Retinal and Eye Research	9.66	***	77
Annual Review of Phytopathology	9.49	***	78
Gastroenterology	9.43	***	79
Journal of Clinical Oncology	9.42	***	80
Mass Spectrometry Reviews	9.41	***	81
Progress in Neurobiology	9.38	***	82
Circulation	9.38	***	83
Neuron	9.38	***	84
American Journal of Human Genetics	9.37	***	85
Journal of Machine Learning Research	9.29	*	86
International Journal of Computer Vision	9.19	*	87
Genes & Development	9.02	***	88
Machine Learning	9.01	**	89
Advanced Drug Delivery Reviews	8.89	***	90
Current Opinion in Cell Biology	8.87	***	91
International Materials Reviews	8.86	**	92
Progress in Quantum Electronics	8.86	**	93
Trends in Cognitive Sciences	8.75	**	94
Trends in Biochemical Sciences	8.67	***	95
Journal of Experimental Medicine	8.63	***	96
Molecular Cell	8.61	***	97
Annual Review of Nuclear And Particle Science	8.49	***	98
Immunity	8.47	***	99
IEEE Signal Processing Magazine	8.40	**	100

ns journals not represented. Ratio SRC of all sources and complete sources: *** if SRC < 0.2; ** if 0.2 ≦ SRC< 1.9; *if 1.9 ≦ SRC < 2.6; ns if SRC ≧ 2.6.

Table 3
General statistics—original values.

$n = 5290$	Mean	StdDev	Skewness	Q1	Median	Q3
IF	2.05	2.95	6.85	0.68	1.31	2.37
AF	2.24	2.40	5.99	1.00	1.76	2.71
IF (weighted)	2.70	3.15	5.36	1.01	1.99	3.32
AF (weighted)	2.70	2.43	5.32	1.39	2.28	3.44

Weight: frequency of journal's articles in the citation window.

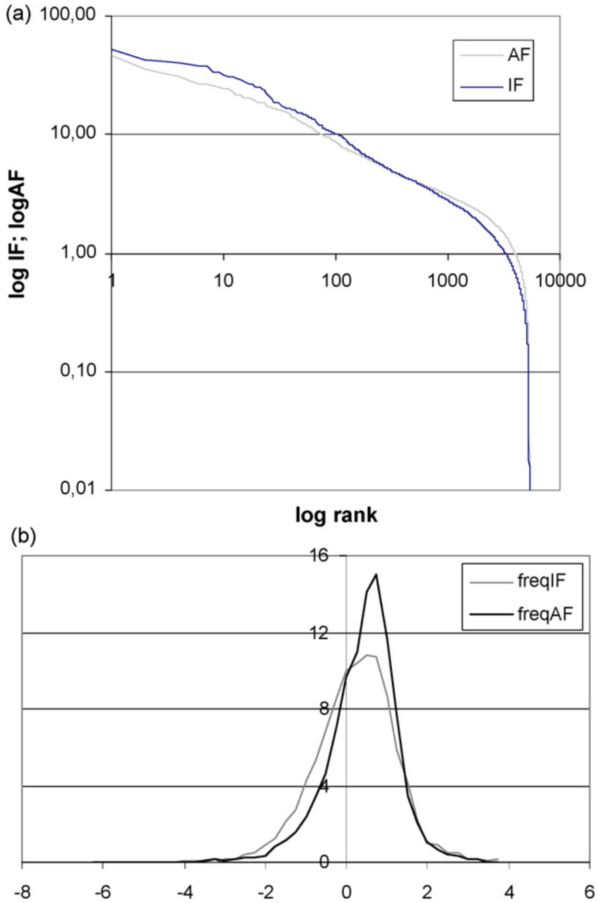

Fig. 3. (A) Distribution of IF and AF (value-rank; double-log scale) items: journals (non weighted). (B) Distribution of IF and AF (log values).

The complete table, also limited to journals with sufficient information, is available from a website.[6] (Table 3).

Few changes are noticed when comparing to the previous version on the top of the list. The overall correlation index of v0.1 and v0.2 AF on the whole list is >0.975 (Pearson) and >0.955 (Spearman). Fig. 3 shows the distributions, both in value-rank (double-log scale) and frequency distribution on transformed (log) values. Most empirical studies stress the skewness of impact factor distributions at the field and global level.

3.2. Overall correlation of AF with IF and FNIF

In order to achieve a precise comparison with AF, the IF is not the standard Thomson Reuters, for comparison sake was recalculated an IF using our available data. Pearson correlations both on original and transformed values (log), and rank correlations (Spearman ranks), exhibit fairly high levels, but lower than in the version 0.1.Determination coefficient is on log values, for example, R-Sq = 0.82 instead of 0.91. Table 4A shows the correlation between IF and AF, with journals

6 http://www.obs-ost.fr/en/know-how/etudes-en-ligne/travaux-2010/note-de-presentation-le-facteur-d-audience-des-periodiques-scientifiques/mesure-experimentale-du-facteur-d-audience-des-periodiques-scientifiques.html.

Table 4
Overall correlations.

4A: Journals Corr. index	n = 5290	IF	AF	
Pearson R	IF	1	0.906[a]	
	AF	0.911	1	
Spearman rank	IF	1	0.860	
	AF	–	1	
4B: Journal-subject category combinations Corr. index	**n = 8438**	**IF**	**NIF**	**AF**
Pearson R	IF	1	0.84[a]	0.90[a]
	FNIF	0.82	1	0.93[a]
	AF	0.89	0.92	1
Spearman rank	IF	1	0.78	0.85
	FNIF		1	0.90
	AF			1

[a] On log values; below the diagonal: original values.

as individuals. In order to compare with FNIF, we need the assignments of journals to subject categories. Many journals are multi-assigned, so we can either pick a single assignment (for example the most favorable relative impact), or keep all assignments. In the latter option, chosen here, the items are now journal-subject categories (pair-wise combinations) ($n = 8438$; Table 4B). Only results with the full range of journals are shown, since reducing to "significant" ones only bring very small changes.

3.3. *Field effects*

The results in the present version generally confirm the findings of v0.1.

3.3.1. *Variability*

Data are based on journal-categories pairs. A lower variance of AF, compared to IF, apparent from Table 3, is expected, due to two phenomena:

- The partial removal of across-field dispersion – the removal is partial since AF keeps the generic effects ("imports–exports") already present in IF. In contrast, for FNIF (on a typical tuning) variability is reduced to the within-field variability of IF.
- A slight reduction of within-field variability. Working on flexible areas, AF reduces the disparities by neutralizing part of the across-areas variability within fields. Total variability of AF is now lower than the within-field counterpart in IF, equivalent to FNIF.

When looking at the results both on direct and log values (Fig. 4), the latter also interesting given the AF/IF skew distribution of Fig. 3, we observe that AF reduces to a large extent across-field variability (keeping part of it since the exchanges of citation are kept), and also part of within-field variability so that the total dispersion is lower than the within-field variability of IF (typically kept within FNIF). The reduction of data to "significant" journals does not globally alter the results.

Another view on the field effect of AF is provided by looking at the distributions at a more aggregate level, the disciplinary level. Fig. 5 provides a general view of AF and IF average values in different "large disciplines". It shows how major deviations of IF values, favorable to biosciences and detrimental especially to *Mathematics*, *Engineering* and at a lesser extent to *Physics* are strongly reduced by AF. *Multidisciplinary* should be regarded with caution given its heterogeneity, and further implementations journal-level would gain to consider separate quasi-sections of journals in this particular case.

Figs. 6 and 7 offer a more detailed view. For legibility, a rank scale of (non cumulative) "quintiles" is used, as in the previous article. By convention, the first quintile – containing the 20% most cited journals – is on the left-hand side. Several types of presentation are shown. Figures of type A show the distribution of quantiles, defined for all science, in each particular discipline, responding to the question: how is the literature of a particular discipline distributed across global quintiles (horizontal vision)? Figures of type B show the distribution of disciplines in each particular quantile of overall visibility, responding to the question: when taking the most visible class (for example) of journals ranked by AF (or by IF), how is it distributed among disciplines (vertical vision)? As expected the most visible quintile contains many biology journals and few mathematical journals if we consider IF. But with AF, the composition is much more balanced. Figures type C shares the vision of figures B but a count of articles is used instead of a count of journals.

Again results are quite similar to the previous implementation, but an improvement of results is recorded in the reduction by AF of across-disciplines discrepancies. Because of the typical positive correlation between the visibility and the size of journals, the first classes appear much heavier when publication numbers are used rather than journal numbers (not shown).

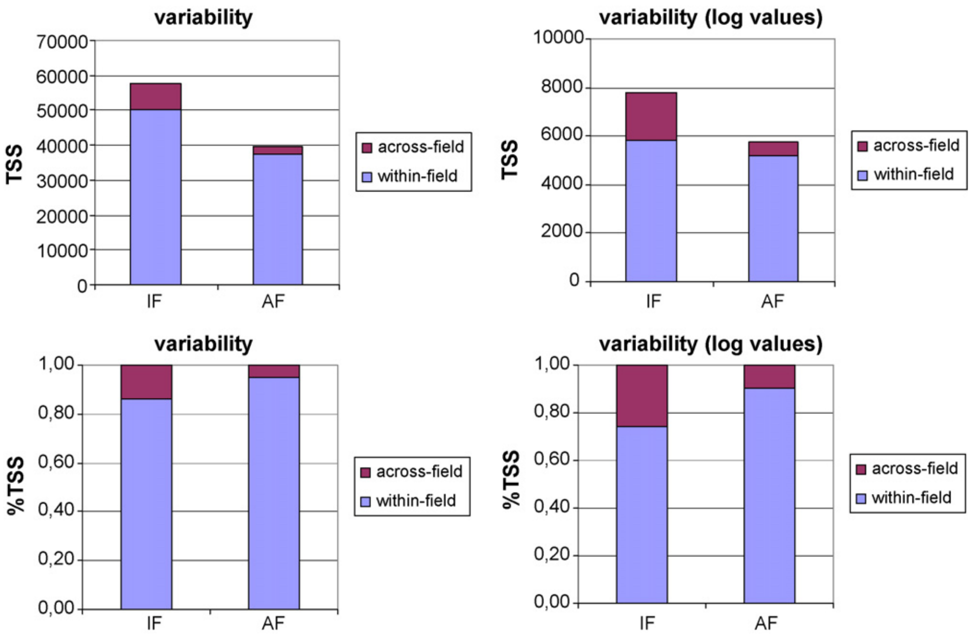

Fig. 4. Across-field and within-field reduction of variability by AF items: journals (non weighted). Up-left: total sum of squares, on original values. Bottom-left: corresponding percentages. Up-right: total sum of squares, on log values. Bottom-right: corresponding percentages

Examination of tables comparing AF and IF (not shown) show that rank differences between the two are maximum for mathematics and computer science journals (scoring better in AF) and biology journals (scoring better in IF). With AF, journals in mathematics can now be found in the top ranks (<100):

SIAM Review, *Bulletin of the American Mathematical Society*, *Journal of the American Mathematical Society*. The short list also contains several computer science journals *ACM Computing Surveys*, *IEEE Transactions on Evolutionary Computation*, *Journal of the ACM*, *Journal of Machine Learning Research*, *International Journal of Computer Vision* (A few candidates among journals with insufficient information have been mentioned above). None of these journals, except *ACM Computing Surveys*, figures in the corresponding top-list of IF. All gain, in comparison with IF, much stronger positions, especially for *Bulletin of the American Mathematical Society* and *Journal of the American Mathematical Society*, which gain more than 400 and 600 positions respectively. Within mathematics, the correlation of AF and IF remains close to the "all fields" value.

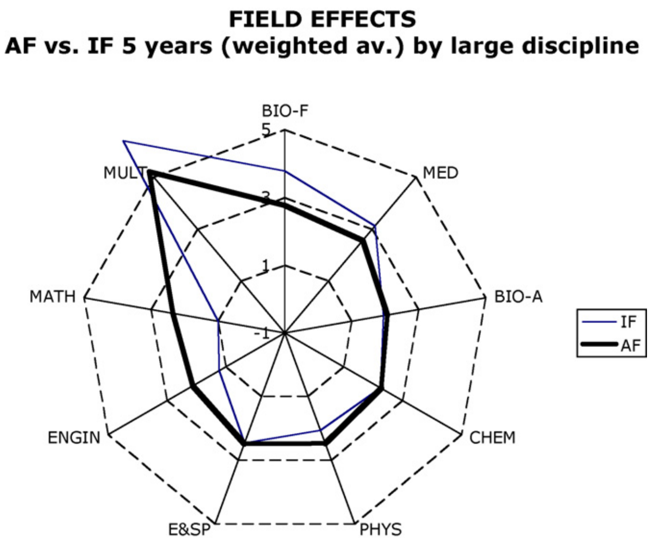

Fig. 5. Average AF and IF by large discipline (weighted by journal size).

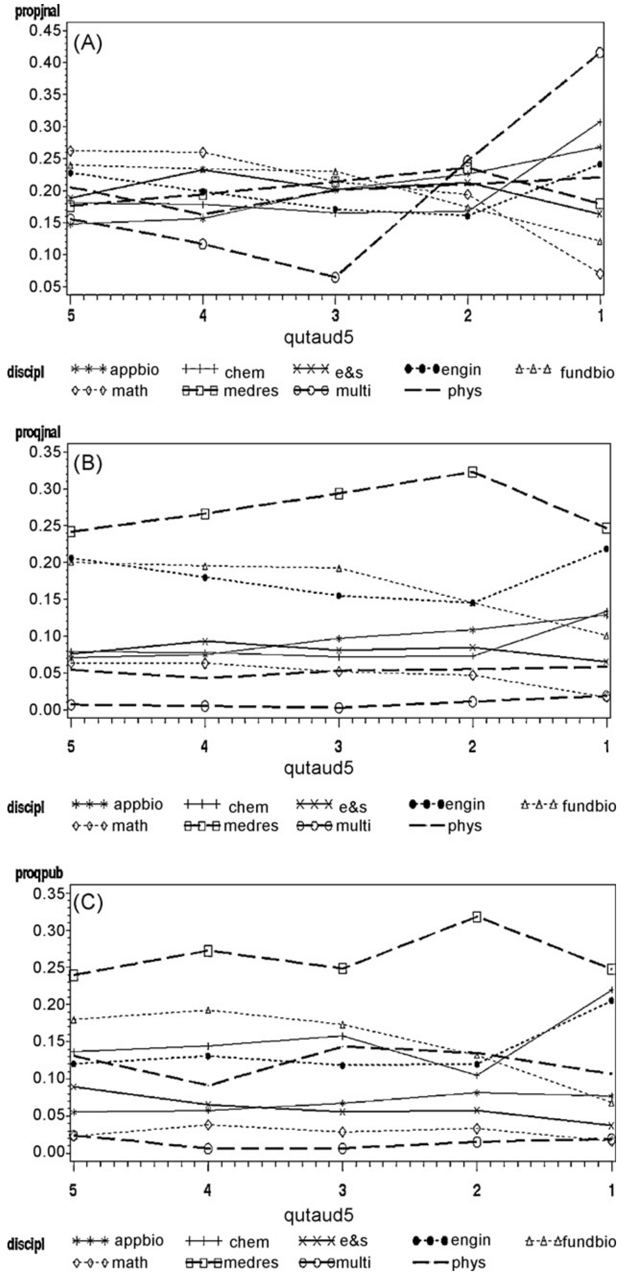

Fig. 6. Distribution of journals across disciplines by level of AF. (A) In each discipline: proportion of the discipline's journals in each 'quintile'. (B) In each global quintile: proportion of each discipline (nb of journals). (C) Like B, for publications.

4. Discussion–Conclusion

In the present version of AF, the artefacts of the former journal level weighting are drastically reduced. The main setting is *N*, the maximum neighborhood at level one. The 2-level neighborhood ensures in almost all cases a robust calculation of bibliography lengths in the scientific area and then of citation weights.

4.1. Various types of normalization

The difference between AF and usual field-normalized IF (classically the ratio IF to the field average) is that the latter (a) normalizes for citation exchanges amongst-fields together with the other sources of impact discrepancy, (b) is typically classification-dependent, both in terms of delineation and of level of fields/areas. A variety of choices are opened:

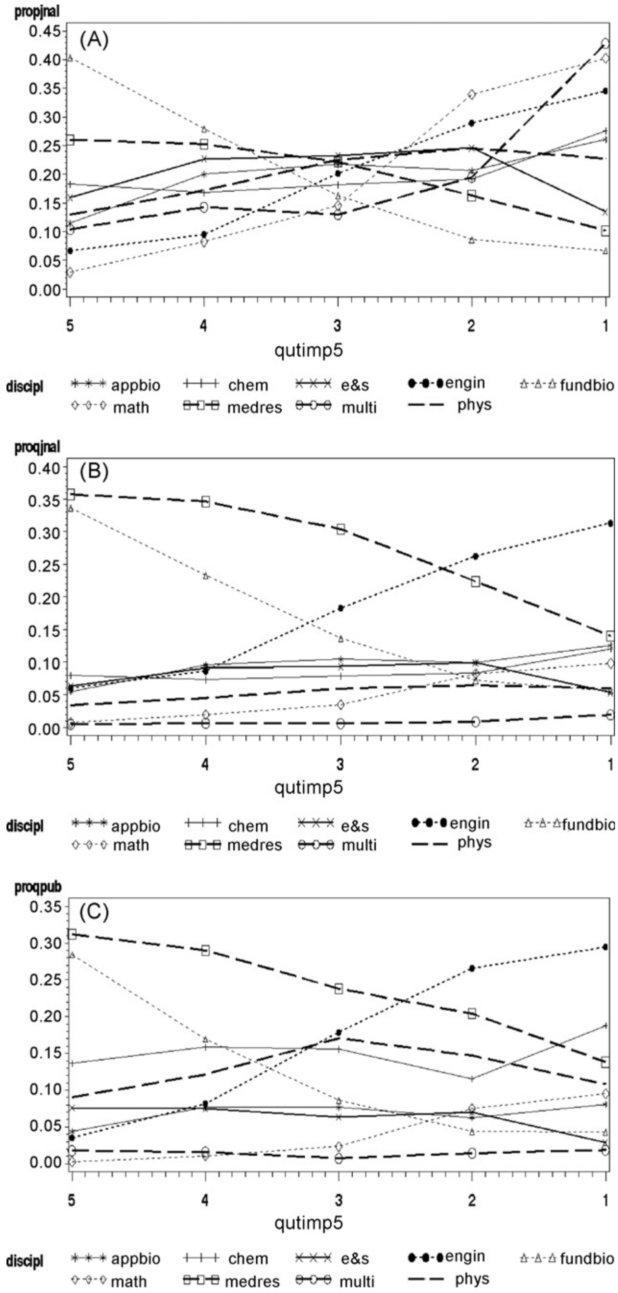

Fig. 7. Distribution of journals across disciplines by level of IF. (A) In each discipline: proportion of the discipline's journals in each 'quintile'. (B) In each global quintile: proportion of each discipline (nb of journals). (C) Like B, for publications.

- If one wishes to normalize for all aspects except the citation exchanges and growth conditions – which implies a higher rating of "knowledge-exporting" areas, then AF should be preferred.
- If one wishes to put all fields on the same average impact level then the field-normalized IF should be picked. Typically, this outcome is classification-dependent, both in terms of delineation and of level of fields. However, a micro-level "cited-side" normalization, based on the neighborhood of cited journals, may be envisioned, to make this process classification free.
- If one wishes to reflect knowledge flows, then "influence measures" should be preferred. They typically incorporate a reinforcement process of visible sources, analogous to a Matthew effect, reinforcement of loops in already skew distributed phenomena. Depending on the implementation, practical problems with dealing with self-citations may be encountered (Bar-Ilan, 2009).

- The interest of a further field-normalization of AF, say on a field basis, is dubious, losing the classification-free property and some overall coherence.

Citing-side normalization can be designed on different methods. It implies two choices:

- Firstly, the design of the community of reference for each citing item, for example: the pool of all articles citing the same cited item (article, journal); the journal of the citing article; a set of journals linked to this journal. This choice is central since the citing behavior is assumed to be reflected at this level. Each options has its particular advantages depending on how it addresses robustness, multidisciplinarity, representativeness of real communities. The choice here is to rely on a journal-level granularity, with a Bradfordian slant – looking at the core – and based on a 2-level setting with a double aim: seeking robustness and, to some extent, trying to address the cross-scale normalization issue, mentioned below. An alternative measure (SNIP mentioned above) relies on a finer granularity, the pool of citing articles, but the higher discriminating power (especially for heterogeneous journals[7]) can be counteracted by lesser robustness in the way communities are defined, especially in generalizations to actor's citation scores. Document-level granularity is an appealing option but, in our view, should exploit neighborhoods in the network to reflect micro-communities in a robust way.
- Secondly, the choice of metrics. Statistics of propensity to cite, given the community of reference, and constrained by available data, include averages, medians, weighted or not. In AFv0.2, the granularity is linked to the use of JCR data, then without access to detailed distributional characteristics at the paper level.

The fact that AF retains in its measure the relative position of fields as "exporters of knowledge" or "importers" does not imply a hierarchy between scientific communities. It is not surprising that a typical academic indicator such as citations from source journals privilege the most basic fields; an application-oriented indicator, such as a citations privileging the patent source, would result in another balance; or a social-oriented indicator, such as citation analysis from normative/prescriptive documents (see other ways in Lewison, 2003), perhaps from newspapers and magazines, would again change the point of view.

4.2. Possible limitations

Some exceedingly small journals, or journals with very few references and hence a too scarce neighborhood, may still produce anomalies. We had to discard a few isolated cases – in contrast to AFv0.1 where the problem was serious. Further analysis could check the case of some other journals in the very low tail of the neighborhood distribution (Fig. 2). Another issue, perhaps, is the fact that in some particular cases, the propensity to cite may be constrained for editorial reasons in citing sources, as for example in the "letters" of Nature at a period. Journals with constraints on the number of references can produce local irregularities, but the principle of aggregation proposed largely reduces this type of problem. Also, the calculation of precise AF needs information on the lists of references of all sources considered, list not limited to those references going to the target set of cited journals (for example JCR journals). Asymmetry of information may need particular settings.

4.3. Audience and size effects

In the case of isolated fields where the size effect may be studied, there is no direct linkage of average impact to size of fields. Range and deviations are another question, as already noticed in the early work of Garfield quoted above. If we compare two fields, one with a small number of journals and the other with a large number of journals, and assume that a single journal were to capture all citations in its field, clearly the impact of this monopolistic journal will be higher in the large field. One also should distinguish between the direct effect and the indirect effect of size. If, for some reason, field size and average length of bibliographies were correlated, an empirical link of impact and field size might be observed. Moreover, the notion of field is fragile: beyond socially recognized definitions for large academic disciplines, "fields" are a mix of observable properties of bibliometric networks – with respect to the choice of linkages and unavoidable artefacts – and researchers' perceptions. The bibliometric delineation of fields raises delicate issues (see for example Zitt & Bassecoulard, 2006) but the fuzzy delineation obtained by superimposition of local neighborhoods of individual citing sources, feeding a particular cited journal, is likely to be fairly realistic.

4.4. Audience and scale effects

Can citing-side normalization spare us from considering scale effects? The instability of relative indicators to the embedded sets used as references (fields, subfields, with largely arbitrary definitions; journals; or embedded research fronts) has been stressed by Zitt et al. (2005), Adams, Gurney and Jackson (2008), Van Raan (2006). For bibliometric variables such as

[7] The problem of multidisciplinary journals is likely to be alleviated by the new assignment rules of Thomson Reuters, where journals are broken down into pseudo-sections assigned to category codes. In the present implementation, only standard data from JCR are used.

the average impact, the structure of science is such that an area at any level is not, generally speaking, a sample of a wider area which includes it. Normalization by local citation habits is more precise than nomenclature-based normalization – for example *ex post* normalization by subject categories. However, this quality per se does not solve the problem of the level of normalization. We have tried to alleviate the issue, by building a neighborhood both variable and robust, as to the citation practices: variable, in order to alleviate the issue of normalization levels; robust, a feature that would be difficult to achieve when using the direct proximity in networks. When giving to a source a weight which depends on the referencing habits of neighbor journals, we face some arbitrary choice in fixing the threshold of neighborhood N. For reasonable values of N, our particular device warrants some stability for small variations of N.[8] The fact that the circle of neighbors of a given citing source is allowed to vary within a wide range, from dense to open areas, makes the calculation of AF weights, in a particular range, independent from scale to a certain extent. The Ochiai-like formula used for citation linkages, brings only a partial normalization by the size of journals, so that large journals are also slightly favored, a feature likely to increase robustness. This independence is obviously not complete and the question of the optimal "zoom" is alleviated in a certain range but not solved.

4.5. Applications of citing-side normalization

A wide range of citation studies can apply outcomes from citing-side normalization, which can be used at the journal level or at the actor level. The principle is that each source of citation has to be normalized by the citing propensity of its particular community. Some adaptation is necessary since the easier application is the "citing year" rationale which prevails for example in impact factor calculation. Applications to "cited year" view – the most usual practice for comparing actors – asks for choices in the way citation windows are managed, but this technical issue can be addressed with appropriate settings. An example of possible application is the family of composite indicators such as the family of *h*-indexes. Citing-side normalization offers an efficient way to normalize the "citation" component, and make in this respect *h*-indexes more comparable across fields. However, it leaves the "publication" component as is, and then only offers a partial normalization solution.

4.6. To conclude

The audience factor aims at neutralizing a main source of citation inequality across fields, the propensity to cite and to cite rapidly, while keeping another source of inequality onboard, the import–export of citations. This is obtained by violating the dogma of "all emitted citations are created equal", obeyed by JIF. AF does not comply with the dogma, clearly and explicitly, since different weight is attributed to citing references as a function of citation practices in the scientific area. Other methods also violate the dogma in a less transparent way: field-normalized IF and influence measures. The latter are powerful tools to track intellectual influence and give another point of view, with also disputable points. AF, with more limited ambition, efficiently reduces the gap between "low citers" such a mathematics and "high citers" such as biosciences. Moreover, within a certain range, the present version allows for large variations in the set of journals used as a proxy of communities, alleviating the fundamental issue of how large reference sets should be used for normalization purposes. AF was presented here as a journal-level indicator, parallel to the impact factor, and also uses a journal-level proxy for delineating research communities used for measuring the propensity to cite. Journal-level indicators such as JIF, AF or influence measures are appealing to characterize media of scientific communication, they are not appropriate for scientific evaluation, as shown in an abundant literature. Beyond journal-level measures, the weighting scheme by citing propensity of sources is a quite general principle, which can be applied, *mutatis mutandis* to any measures of citation, by ruling the granularity on the citing side (how to define the citing community) as well as on the cited side (impact of journals, individual or aggregate actors, institutions).

Acknowledgements

The author is most indebted to Henry Small (Thomson Reuters) for the many and fruitful exchanges since their co-work in 2007–2008 that led to the first version of AF, and to two referees for helpful suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.joi.2010.03.004.

References

Adams, J., Gurney, K., & Jackson, L. (2008). Calibrating the zoom: A test of Zitt's hypothesis. *Scientometrics*, *75*(1), 81–95.

[8] The choice of $N = 10$, however, can be challenged on the ground that a community peaking at 100 journals may appear too large. Some tests are necessary in this respect.

Bar-Ilan, J. (2009). A closer look at the sources of informetric research. *Cybermetrics*, *13*(1), paper 4 (from keynote address, 12th ISSI International Conference, 14–17 July, Rio de Janeiro) and pers. comm.

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals. *Scientometrics*, *44*(3), 323–345.

Bergstrom, C. T. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. College & Research Libraries News, 68(May(5)). www.eigenfactor.org Last accessed 01.02. 2010 www.eigenfactor.org/methods.pdf.

Biglu, M. (2008). The influence of references per paper in the SCI to impact factors and the Matthew effect. *Scientometrics*, *74*(3), 453–470.

Bollen, J., Rodriguez, M. A., & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687.

Bordons, M., Fernandez, M. T., & Gomez, I. (2004). Advantages and limitations in the use of Impact Factor measures for the assessment of research performance. *Scientometrics*, *53*(2), 195–206.

Cozzens, S. E., & Leydesdorff, L. (1993). Journal systems as macro-indicators of structural change in the sciences. In A. F. J. van Raan, R. E. de Bruin, H. F. Moed, A. J. Nederhof, & R. W. J. Tijssen (Eds.), *Science and technology in a policy context* (pp. 219–233). Leiden: DSWO/Leiden University.

Czapski, G. (1997). The use of deciles of the citation impact to evaluate different fields of research in Israel. *Scientometrics*, *40*(3), 437–443.

Garfield, E. (1972). Citation as a tool in journal evaluation. *Science*, *178*, 471–479.

Garfield, E. (1976). Is the ratio between number of citations and publications cited a true constant? *Current Contents*, 6, Feb 9, Reprinted in Essays of an Information Scientist, Vol. 2, 1974–1976, 419–421.

Garfield, E. (1998). Random thoughts on citationology, its theory and practice. *Scientometrics*, *43*, 69–76.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(2), 171–193.

Jackson, M. O., & Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, *97*(3), 890–915.

Lewison, G. (2003). Beyond outputs: New measures of biomedical research impact. *Aslib Proceedings*, *55*(1/2), 32–42.

Marshakova-Shaikevich, I. (1996). The standard impact factor as an evaluation tool of science fields and scientific journals. *Scientometrics*, *35*(2), 283–290.

Moed, H. F. (2002). The impact-factor debate: The ISI's uses and limits. *Nature*, *415*, 731–732.

Moed, H. F. (2009). Measuring contextual citation impact of scientific journals. arxiv.org/pdf/0911.2632 Last accessed 02.02.2010. Forthcoming in Journal of Informetrics.

Moya-Anegon, F. (2007). SCImago. SJR—SCImago Journal & Country Rank. http://www.scimagojr.com Last accessed 01.02.2010.

Murugesan, P., & Moravcsik, M. J. (1978). Variation of the nature of citation measures with journal and scientific specialties. *Journal of the American Society for Information Science*, *29*, 141–155.

Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. National Science Foundation. Contract NSF C-627

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*, 609–641.

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in Japan and its neighboring regions. *Bulletin of the Japanese Society of Scientific Fisheries*, *22*, 526–530.

Palacio-Huerta, I., & Volij, O. (2004). The measurement of intellectual influence. *Econometrica*, *72*(3), 963–977.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory with application to the literature of physics. *Information processing and management*, *12*, 297–312.

Pudovkin, A. I., & Garfield, E. (2004). Rank-normalized impact factor: A way to compare journal performance across subject categories. In *Proceedings of the 67th annual meeting of the American society for information science & technology, Vol. 41* (pp. 507–515).

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Science of the United States of America*, *105*(45), 17268–17272.

Ramirez, A. M., Garcia, E. O., & del Rio, J. A. (2000). Renormalized impact factor. *Scientometrics*, *47*(1), 3–9.

Rousseau, R. (2005). Median and percentile impact factors: A set of new indicators. *Scientometrics*, *63*(3), 431–441.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*(7079), 498–502.

Schubert, A., & Braun, T. (1996). Cross-field normalization of scientometric indicators. *Scientometrics*, *0*(0), 1–14.

Sen, B. K. (1992). Normalized impact factor. *Journal of Documentation*, *48*(3), 318–325.

Small, H. (1995). Navigating the citation network. In *Proceeding of the 58th annual meeting of the American society for information science, Vol. 32* (pp. 118–126).

Small, H., & Sweeney, E. (1985). Clustering the Science Citation Index using co-citations I—A comparison of methods. *Scientometrics*, *7*(3-6), 391–409.

Van Raan, A. F. J. (2006). Performance-related differences of bibliometric statistical properties of research groups: Cumulative advantages and hierarchically layered networks Source. *Journal of the American Society for Information Science and Technology*, *57*(14), 1919–1935.

Vinkler, P. (2002). Subfield problems in applying the Garfield (Impact) Factors in practice. *Scientometrics*, *53*(2), 267–279.

Vinkler, P. (2004). Characterization of the impact of sets of scientific papers: The Garfield (impact) factor. *Journal of the American Society for Information Science and Technology*, *55*(5), 431–435.

Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing and Management*, *42*(6), 1513–1531.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, *vol 63*(2), 373–401.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology (JASIST)*, *vol 59*(11), 1856–1860.

Zitt, M. & Small, H., (2008b). Field normalization of impact factors: a citing side approach, 10th International Conference on Science and Technology Indicators, 17–20 Sept., Vienna, Austria.