Journal of Informetrics 4 (2010) 239–248

Contents lists available at ScienceDirect

Journal of Informetrics

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

Journal influence factors

Massimo Franceschet*

Department of Mathematics and Computer Science, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

ARTICLE INFO

Article history:
Received 26 September 2009
Received in revised form
13 November 2009
Accepted 16 December 2009

Keywords:
Journal influence measures
Impact factor
Eigenfactor metrics
Cross-field variability

ABSTRACT

We performed a thorough comparison of four main indicators of journal influence, namely 2-year impact factor, 5-year impact factor, eigenfactor and article influence. These indicators have been recently added by Thomson Reuters to the Journal Citation Reports, in both science and social science editions, and are thus available for study and comparison over a sample of significative size. We find that the distribution associated with the eigenfactor largely differs from the distribution of the other surveyed measures in terms of deviation from the mean, concentration, entropy, and skewness. Moreover, it is the one that best fits to the lognormal theoretical model. Surprisingly, the eigenfactor is also the most variable indicator when computed across different fields of science and social science, while article influence is the most stable in this respect, and hence the most suitable metric to be used interdisciplinarily. Finally, the journal rankings provided by impact factors and article influence are relatively similar and diverge from the one produced by eigenfactor, which is closer to that given by the total number of received citations.

1. Introduction

The impact factor is, undoubtedly, the most popular and controversial bibliometric indicator available at the moment. It is defined, for a given journal, as the mean number of citations in a given census year to papers published in the journal during a target window consisting of the two previous years. It has been proposed by Eugene Garfield, working together with Irv Sher, to identify influential journals using the recent citations received from other journals (Garfield & Sher, 1963), and has been used at least since 1976, when the first Journal Citation Reports appeared as part of the Science Citation Index.

The use of the impact factor as a metric of journal status has been widely discussed (Bar-Ilan, 2008; Moed, 2002; Wilson, 1999). The strengths of the impact factor are comprehensibility, simplicity, robustness, wide and fast availability (Glänzel & Moed, 2002; Pendlebury, 2009). On the other hand, frequently mentioned flaws of the impact factor are that the target window of two years is too short, it does not represent a typical value since it is a mean of a highly skewed distribution, it does not consider the status of the citing journals, and it widely differs from one discipline to another (Amin & Mabe, 2000; Campbell, 2008; Seglen, 1997).

Thomson Reuters is not deaf to these criticisms and honestly admits the major drawbacks of the impact factor, warning, at the same time, against possible misuse of the metric, like assessing individual scholars or papers, that was never intended by the creators (Garfield, 2006; Pendlebury, 2009). In the 2007 editions of the Journal Citation Reports for science and social science, Thomson Reuters added three alternative journal performance measures, which try to avoid the mentioned flaws of the original 2-year impact factor. These are:

* Corresponding author. Tel.: +39 0432 558754; fax: +39 0432 558499.
E-mail address: massimo.franceschet@dimi.uniud.it.

1751-1577/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2009.12.002

- 5-*Year impact factor*. This is computed as the original impact factor but with a longer target window of 5 years.
- *Eigenfactor*. This is the sum of normalized citations received from other journals weighted by the status of the citing journals. Citations are normalized with respect to the total amount of cited references of the citing journal. The citation target period is 5 years.
- *Article influence*. This is the eigenfactor score divided by the number of articles published by the journal over the 5-year target period.

Given the availability of these four metrics over a large-scale journal set (6598 science journals and 1980 social science journals in the 2007 Journal Citation Reports), it is interesting to analyse and compare them. In particular, in this paper we address the following issues:

1. What is the form of the empirical distribution associated with the journal indicators in terms of deviation from the mean, concentration, entropy, and skewness? In particular, we investigate how well the empirical distributions follow the theoretical lognormal model.
2. Which is the indicator that is most stable across different disciplines? Do the new indicators reduce the well-known cross-field variability of the 2-year impact factor? This issue is related to the interdisciplinary usage of the metrics.
3. How similar are the journal rankings provided by the performance indicators?

Section 2 explores carefully the mentioned issues. Section 3 reports the related work, and Section 4 concludes the paper.

2. Factors of journal influence

Journal impact factors are defined as the mean number of citations in a given census year to papers published in the journal during an immediately previous target period. Typical target windows are 2 years long and 5 years long, giving rise to the 2-year impact factor (IF2) and the 5-year impact factor (IF5).

The computation method of the eigenfactor (EF) is more involved and exploits the entire citation network (Bergstrom, West, & Wiseman, 2008; Bollen, Rodriguez, & de Sompel, 2006; Pinski & Narin, 1976). Unlike the impact factor, the eigenfactor method weights journal citations by the status of the citing journals. As a result, the status of a journal is determined not only by the number of received citations, but also by the status of the citing journals. More precisely, let us fix a census year and let $C = (c_{i,j})$ be a journal–journal citation matrix such that $c_{i,j}$ is the number of citations from articles published in journal i in the census year to articles published in journal j during a target window consisting of the five previous years. Journal self-citations are ignored, hence $c_{i,i} = 0$ for all i. Moreover, let a be an article vector such that a_i is the number of articles published by journal i over the 5-year target window divided by the total number of articles published by all journals over the same period. A dangling node is a journal i that does not cite any other journals. The citation matrix C is transformed into a normalized matrix $H = (h_{i,j})$ such that all rows (journals) that are not dangling nodes are normalized by the row sum (the total number of citations made by the journal). Furthermore, H is mapped to a matrix $\hat{H}$ in which all rows corresponding to dangling nodes are replaced with the article vector a. A new matrix P is defined as follows:

$$P = \alpha \hat{H} + (1 - \alpha) A$$

where A is the matrix with identical rows each equal to the article vector a, and α is a free parameter of the algorithm, usually set to 0.85. Let π be the left eigenvector of P associated with the unity eigenvalue, that is, the vector π such that $\pi = \pi P$. The vector π, called the *influence vector*, contains the scores used to weight citations allocated in matrix H. Finally, the eigenfactor vector r is computed as

$$r = 100 \cdot \frac{\pi H}{\sum_i [\pi H]_i}$$

That is, the eigenfactor score of a journal is the sum of normalized citations received from other journals weighted by the eigenfactor scores of the citing journals. The eigenfactor scores are normalized such that they sum to 100 (West, Althouse, Bergstrom, Rosvall, & Bergstrom, 2009).

Finally, the article influence (AI) for a journal is simply its eigenfactor divided by the number of articles published by the journal over the 5-year target period; hence, it corresponds to the journal eigenfactor score per published article (Bergstrom et al., 2008).

All these measures are available at Thomson Reuters Journal Citation Reports (JCR) for the indexed science and social science journals. Furthermore, eigenfactor and article influence scores are published on the eigenfactor web site for journals listed in JCR and also for those journals that do not belong to JCR but are cited by other JCR journals[1] (West et al., 2009). In the present study, we use JCR 2007 science and social science editions.

[1] Eigenfactor scores are added to the eigenfactor web site 6 months after they are published in JCR.

Table 1
Descriptive statistics for factors of journal influence: 1st quartile, median (2nd quartile), mean, 3rd quartile, maximum, deviation (dev), concentration (conc), entropy (ent), and skewness (skew).

Index	1st Qu	Median	Mean	3rd Qu	max	dev	conc	ent	skew
IF2	0.625	1.210	1.850	2.200	74.580	1.43	0.50	0.40	8.10
IF5	0.772	1.478	2.138	2.543	50.770	1.31	0.48	0.51	6.46
AI	0.257	0.503	0.809	0.8815	24.860	1.67	0.54	0.45	7.53
EF	0.00106	0.00288	0.01195	0.00812	1.76400	4.27	0.77	0.13	19.80

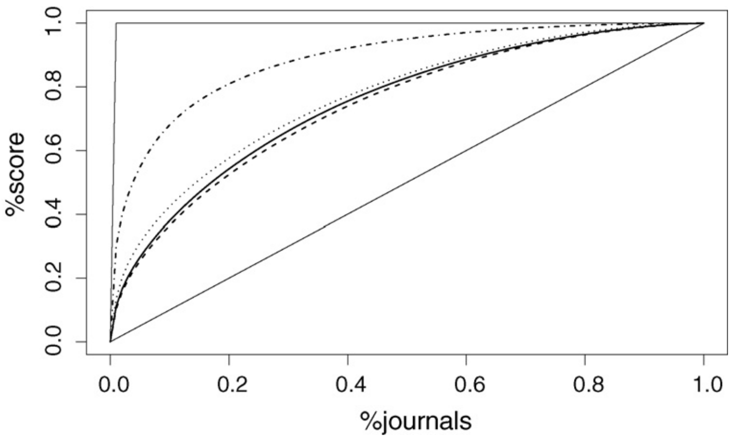

Fig. 1. Inverted Lorenz diagram with curves for IF2 (solid), IF5 (dashed), AI (dotted), and EF (dot-dashed). The curves are contained is a triangle representing the extreme cases of equidistribution (the side of the triangle corresponding to the segment with slope 1) and maximum concentration (the other two sides of the triangle).

2.1. Variability, skewness, and model fitting

This section describes the form of the distribution of the different measures of journal influence in terms of deviation, concentration, entropy, and skewness. Moreover, we investigate the fitting of the indicator distributions with respect to the lognormal model. Finally, we study the cross-field variability of the indicator scores, and discuss the reasons for such a variability.

Table 1 summarizes centrality, variability, and skewness descriptive statistics we have computed for the indicators at hand. The *deviation* of a distribution is the attitude of the data to deviate from the typical value, the mean, of the distribution. A relative measure of deviation is the coefficient of variation CV, defined as the ratio between the standard deviation σ and the mean μ of the distribution:

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(x_i - \mu)^2}}{\mu}$$

The eigenfactor largely dominates the other measures in terms of deviation: its standard deviation if more than four times its mean (Table 1, column *dev*). The coefficients of variation of the other three measures are close, with 5-year impact factor having the lowest deviation from the mean.

Concentration measures how the character is equally distributed among the statistical units. The two extreme situations are equidistribution, in which each statistical unit receives the same amount of character, and maximum concentration, in which the total amount of the character is attributed to a single statistical unit. A typical application of concentration is the analysis of the allocation of wealth among individuals; Pareto (1897) originally observed that a larger share of wealth of any society (approximately 80%) is owned by a smaller fraction (about 20%) of the people in the society (Pareto principle or 80-20 rule). Concentration is a particular measure of variability that is conceptually different from deviation, although the extreme cases of maximality and minimality are the same for both measures.

Fig. 1 depicts an (inverted) Lorenz diagram with curves representing the concentration of the different indicators.[2] The curve for a given indicator is obtained by sorting journals in decreasing order with respect to the indicator scores. Then, the share of top journals collecting a given percentage of indicator score is plotted. Notice that all four curves are contained

[2] With respect to the traditional Lorenz curve, the statistical units are sorted in *decreasing* order.

inside a solid bounding triangle: the side of the triangle corresponding to the segment with slope 1 leading from point (0, 0) to point (1, 1) represents the extreme situation of equidistribution: each journal receives the same amount of total score. The other two sides of the triangle represent the alternative extreme situation of maximum concentration: the total amount of score is assigned to a single journal (the first is the sorted sequence). A relative measure of concentration is the concentration ratio *CR*, which is the ratio between the area R_C contained between the curve and the equidistribution line and the area R_T of the bounding triangle.[3] It turns out that

$$R_C = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}(q_{i-1}+q_i) - 1\right)$$

$$R_T = \frac{1}{2}\frac{n-1}{n}$$

$$CR = \frac{R_C}{R_T} = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(q_{i-1}+q_i) - 1\right)$$

where $q_0 = 0$ and, for $i > 0$, q_i is the percentage of total score collected by the top i journals. Notice that the concentration ratio ranges between 0 and 1 with 0 representing equidistribution and 1 representing maximum concentration.

By inspecting the different curves in Fig. 1 and the corresponding concentration ratios contained in column *conc* of Table 1, we observe that the concentration of impact factors and that of article influence are quite close, with 5-year impact factor being the most equidistributed sample. On the other hand, the eigenfactor is far more concentrated than the other journal metrics. In particular, the top 5% of the journals collect more than half of the eigenfactor score, and the top half of the journals harvest 95% of the eigenfactor score. Interestingly, exactly 80% of the eigenfactor score comes from 20% of the top journals, perfectly matching the above-mentioned Pareto principle. As for the less concentrated metric, that is article influence, we have that the top 16% of the journals collect more than half of the article influence score, and the top half of the journals represent the 83% of the article influence score.

Entropy is a measure of heterogeneity that captures the amount of uncertainty or randomness in a distribution. It is also related to the theoretical compression rate that is possible to achieve when the distribution sample is viewed as a message to be transmitted across a channel (Shannon, 1948). For each indicator, let us divide the domain interval $[m, M]$, where m is the minimum indicator score and M is the maximum indicator score in the sample, into k intervals of equal size. Let f_i be the relative frequency of elements contained in the ith segment, that is, f_i is the ratio between the number of sample elements belonging to the ith interval and the total number of sample elements. The entropy H of the distribution sample is as follows:

$$H = -\sum_{i=0}^{k} f_i \log_2 f_i$$

where $f_i \log_2 f_i = 0$ if $f_i = 0$. The entropy is minimum (equal to 0) when there is some j such that $f_j = 1$ and $f_i = 0$ for all $i \neq j$. That is, all sample elements are contained in the jth interval and, therefore, there is no uncertainty about which interval contains an element randomly chosen from the sample. The entropy is maximum (equal to $\log_2 k$) when $f_i = 1/k$ for all i. That is, each interval contains the same number of sample elements and, therefore, there is maximum uncertainty about which interval a random sample element belongs to. A relative measure of uncertainty is the ratio $H/\log_2 k$ which ranges between 0 (minimum uncertainty) and 1 (maximum uncertainty). Entropy is a measure of heterogeneity different from deviation and concentration. In particular, the situation of minimum concentration (or deviation) corresponds to the situation of minimum entropy, but the situation of maximum concentration (or deviation) and maximum entropy are different.

We computed the uncertainty ratio for the journal influence factor distributions setting the number of intervals $k = 100$ (Table 1, column *ent*). The eigenfactor conveys the lowest uncertainty: the absolute entropy is 0.88 bits, which amounts to 13% of the maximum entropy for the fixed valued of k, thus the theoretical compression rate is 87%. On the other hand, 5-year impact factor is the most entropic measure: the absolute entropy is 2.99 bits, which corresponds to 45% of the maximum randomness, achieving a lower theoretical compression rate of 55%. The values of entropy for the other two indicators (IF2 and AI) are closer to that of IF5 than to the one for EF.

Finally, *skewness* measures the symmetry of a distribution. A distribution is symmetric if the values are equally distributed around its mean; a well-known example is the normal distribution. A distribution is right-skewed if it contains many low values and a relatively few high values. It is left-skewed if it comprises many high values and a relatively few low values. As a rule of thumb, when the mean is larger than the median the distribution is right-skewed and when the median dominates the mean the distribution is left-skewed. A numerical indicator of skewness is the third standardized central moment *SK*,

[3] This corresponds to the Gini coefficient for the traditional Lorenz diagram.

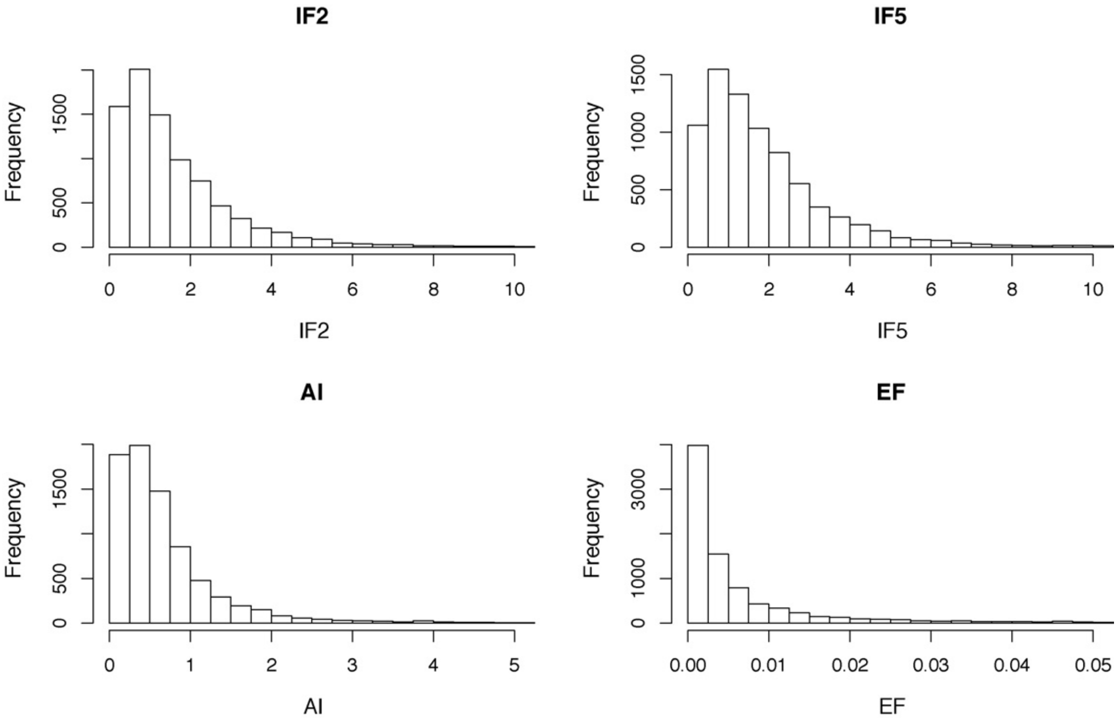

Fig. 2. Histograms for factors of journal influence.

that is:

$$SK = \frac{(1/n)\sum_{i=1}^{n}(x_i - \mu)^3}{\sigma^3}$$

Positive values for the skewness indicator correspond to right skewness, negative values correspond to left skewness, and values close to 0 mean symmetry.

Fig. 2 illustrates the histograms for the four journal influence factors under investigation. From a visual inspection of the histograms it is clear that all factor distributions are right-skewed and that the skewness of the eigenfactor distribution is more pronounced that the asymmetry of the other three distributions. The outcomes of the visual analysis are confirmed by the computation of the skewness indicator (Table 1, column *skew*): the skewness of EF is more than three times higher than the skewness of IF5, whose distribution is the most symmetric one. Furthermore, notice that, for each indicator, the mean is greater than the median and the gap between the median (2nd quartile) and the 1st quartile is less than the gap between the 3rd quartile and the median. In particular, for the eigenfactor, the mean is even greater than the 3rd quartile.

Next, we compare the empirical distributions of the journal influence indicators with the theoretical lognormal model. The lognormal probability density function is defined in terms of parameters μ and $\sigma > 0$ as follows:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\log(x)-\mu)^2/2\sigma^2}$$

for $x > 0$. A variable is lognormally distributed if the logarithm of the variable is normally distributed. The lognormal distribution is commonly used to model bibliometric phenomena, including scientific productivity of scholars (Shockley, 1957) and citations to journal articles (Radicchi, Fortunato, & Castellano, 2008; Stringer, Sales-Pardo, & Amaral, 2008). We computed the logarithm of the indicator values and checked the fitting of the resulting distributions with the normal distribution using the Kolmogorov–Smirnov test.[4] We estimated the distribution parameters using the maximum likelihood method. Table 2 contains the results we have found for the logarithm of the original distributions. The best fitting with respect to the lognormal model is observed for the eigenfactor distribution: the mean is very close to the median and the skewness indicator is close to 0. Furthermore, the eigenfactor has the lowest Kolmogorov–Smirnov statistic and the highest *p*-value

[4] The test compares an empirical sample and a theoretical model by computing the maximum absolute difference between the empirical and theoretical cumulative frequencies.

Table 2
Statistics for the logarithm of indicator scores: median, mean, standard deviation (sd), skewness, Kolmogorov–Smirnov (KS) statistic and corresponding *p*-value for the lognormal fitting.

Index	Median	Mean	sd	skew	KS	*p*-Value
IF2	0.1906	0.1253	1.0383	−0.82	0.049	0.062
IF5	0.3907	0.3123	0.9792	−0.57	0.054	0.041
AI	−0.6872	−0.7887	1.1176	−0.71	0.070	0.003
EF	−5.8500	−5.8520	1.6629	−0.18	0.040	0.178

(greater than the usual significance level of 5%). The *p*-values for the impact factors are also significative (IF2 passed the test at the 5% significance level and IF5 at the 1% level), while the article influence sample does not seem to fit well the lognormal model.

We conclude this section by studying the variability of the scores for the journal influence factors across different disciplines within science and social science. The mapping of science journals into disciplines is taken from the map of science based on Thomson Reuters subject categories that was recently computed by Leydesdorff and Rafols (2009). The authors used exploratory factor analysis to cluster the 175 subject categories into 14 factors corresponding to disciplines in science, e.g., biomedical sciences, engineering, geosciences. For each discipline, we selected the five subject categories with the highest factor loadings on the cluster identified by the discipline; they correspond to the most representative categories of the discipline. For instance, for the discipline computer sciences, we selected categories hardware and architecture, information systems, artificial intelligence, electrical and electronic engineering, and theory and methods. Finally, we included in each discipline sample all journals belonging to the selected subject categories. The resulting science sample comprises 4318 journals. Leydesdorff and Rafols chose to exclude the social sciences from their map. In order to populate the social science samples, we selected the JCR subject categories in the social sciences with the highest number of journals and aggregated them into six disciplines using domain knowledge. The resulting social science sample contains 1018 journals.

Table 3 contains the disciplines, sorted in increasing order with respect to the median value of each of the four influence indicator that we consider. Disciplines neurosciences, biomedical sciences, and infectious diseases are those with the highest scores irrespective of the indicator. On the other hand, education and educational research and sociology are the disciplines with the lowest scores. In general, the scores for science disciplines are higher than the scores for social science disciplines, and the scores for biology-medical disciplines dominate the scores for the other science disciplines. We measured the variability across disciplines of the different influence measures by computing the coefficient of variation of the median discipline scores. The most stable indicator is AI (0.31), followed by IF5 (0.45), IF2 (0.48) and EF (0.59).

These outcomes deserve a broader discussion. In principle, one might expect that the average influence of journals in each discipline, and hence the scores of the corresponding indicators, is comparable. In practice, however, this is not the case, and this prevents the fair comparison of bibliometric indicators across disciplines. Impact factors are well known to vary widely across disciplines (Moed, Burger, Frankfort, & Raan, 1985; Seglen, 1997). Recently, Althouse, West, Bergstrom, and Bergstrom (2008) analyse the sources of these variations and conclude, surprisingly, that the greatest contributor to

Table 3
Disciplines sorted in increasing order with respect to the discipline median indicator score. Discipline names are abbreviated as follows: ARG (agriculture), BIOM (biomedical sciences), CHE (chemistry), CLM (clinical medicine), CS (computer sciences), ECO (ecology), ENG (engineering), ENV (environmental sciences), GMH (general medicine and health), GEO (geosciences), IND (infectious diseases), MAT (material sciences), NEU (neurosciences), PHY (physics), EMB (economics, management, and business), EDU (education and educational research), LAW (law), POL (political science), PSY (psychology) and SOC (sociology).

IF2	Score	IF5	Score	AI	Score	EF	Score
EDU	0.55	POL	0.63	AGR	0.282	SOC	0.0011
POL	0.56	EDU	0.71	EDU	0.330	EDU	0.0012
SOC	0.63	SOC	0.75	SOC	0.364	LAW	0.0013
ENG	0.71	LAW	0.81	CS	0.374	POL	0.0015
EMB	0.75	ENG	0.82	MAT	0.386	AGR	0.0020
CS	0.80	CS	0.91	LAW	0.388	PSY	0.0022
LAW	0.87	EMB	0.98	CHE	0.391	CS	0.0023
AGR	0.91	AGR	0.99	ENG	0.400	EMB	0.0024
MAT	0.95	MAT	0.99	ENV	0.409	ENG	0.0025
ENV	0.97	PHY	1.08	POL	0.416	ECO	0.0027
GEO	1.02	ENV	1.10	CLM	0.446	ENV	0.0028
ECO	1.13	GEO	1.22	ECO	0.450	GEO	0.0031
PSY	1.14	CHE	1.28	GMH	0.466	MAT	0.0040
PHY	1.18	ECO	1.33	GEO	0.540	GMH	0.0043
CHE	1.22	GMH	1.42	EMB	0.542	CHE	0.0046
GMH	1.40	PSY	1.44	PHY	0.577	CLM	0.0053
CLM	1.43	CLM	1.48	PSY	0.586	PHY	0.0058
NEU	2.22	IND	2.35	IND	0.728	NEU	0.0065
BIOM	2.37	BIOM	2.42	BIOM	0.799	BIOM	0.0073
IND	2.39	NEU	2.53	NEU	0.815	IND	0.0078

Table 4
Disciplines sorted in increasing order with respect to the mean journal size (Size) and mean cited half-life (CHL). Discipline names are abbreviated as in Table 3.

Disc	LAW	SOC	EDU	POL	EMB	PSY	ECO	GEO	ENG	CS
Size	31	33	36	41	49	51	91	93	95	106
Disc	AGR	GMH	NEU	ENV	IND	BIOM	CLM	MAT	PHY	CHE
Size	110	110	118	119	163	166	170	261	276	284
Disc	IND	MAT	BIOM	GMH	CS	PHY	CHE	CLM	ENV	NEU
CHL	5.8	6.0	6.2	6.5	6.5	6.5	6.6	6.6	6.6	6.8
Disc	ENG	POL	GEO	AGR	ECO	LAW	EDU	EMB	PSY	SOC
CHL	7.1	7.2	7.5	7.5	7.6	7.9	7.9	8.0	8.0	8.6

differences across fields is discipline internal coverage of Thomson Reuters Web of Science data source, that is, the fraction of references cited in articles of the discipline that match papers contained in Web of Science. Internal coverage of Web of Science largely varies across disciplines; for instance, the internal coverage of molecular and cell biology is 0.802, that of psychology is 0.538, and that of computer science is 0.266. Additional factors of cross-field variability are average length of article bibliographies and the fraction of references that were published in the 2-year target window used by the traditional impact factor. Notice that discipline internal coverage has nothing to do with citation practices of the discipline, but relates to the content of the particular database only.

The creators of the eigenfactor claim that this method mitigates the impact of different discipline citation practices, because it normalizes in its formula citations from a source item to a target item by the total number of citations given by the source item (West et al., 2009). Furthermore, it uses a 5-year target window, larger than the 2-year period commonly exploited by the impact factor. This allows a more suitable evaluation for journals with longer cited lives. Nevertheless, the eigenfactor cross-discipline variability is the highest among the analysed factors. The main reason for this is that eigenfactor is size-dependent: with all else equal, bigger journals will have larger eigenfactor scores, since they have more articles and hence we expect them to be cited more often. But journal size strongly differs across disciplines: the average size (number of articles published in one year) of a journal in chemistry is 284, that of a journal in computer science is 106, that of a journal in law is 31 (see Table 4). Indeed, the cross-field variability of article influence, which corresponds to eigenfactor per article and thus normalizes with respect to the size factor, is much lower than that of eigenfactor, and, notably, is the lowest among the surveyed indicators. Hence, the original intuition of the eigenfactor inventors is correct, but it is capitalized only by the article influence version of the measure. Causes of the remaining cross-field variability in the article influence scores are the different discipline internal coverage of the data source Web of Science and the fraction of references that were published in the 5-year target period used by the article influence. As for the latter variability factor, we noticed that the mean journal cited half-life – the median age of the journal articles cited in a given year, an indicator of the velocity of accumulation of citations for the journal – varies across fields, from a maximum of 8.6 years for sociology to minimum of 5.8 years for infectious diseases (see Table 4).

2.2. Correlation analysis

In this section we study the correlation between the journal influence indicators covered in the present study.

We start by comparing the discipline ranking contained in Table 3. We look for disciplines having strongly diverging ranks with respect to the following three pairs of indicators: (IF2, IF5), (IF5, AI), and (AI, EF). The discipline rankings provided by IF2 and IF5 are strongly correlated: the Spearman correlation coefficient is 0.95. Hence, we noticed no strong divergence. The association between the discipline rankings given by IF5 and AI is less significative (Spearman 0.74). The main difference between IF5 and AI is that the latter takes into account the status of the citing journal. Disciplines that loose an important number of positions when moving from IF5 to AI rankings are agriculture (−7 positions), chemistry (−6 positions), and clinical medicine (−6 positions). On the other hand, disciplines that gain a significative number of positions when shifting from IF5 to AI rankings are political science (+9 positions) and economics, management, and business (+8 positions). The correlation between the discipline rankings according to AI and EF is 0.68, the lowest among the analysed pairs. Recall that EF measures the cumulative performance of a journal, while AI measures the performance per article. Hence, EF is a size-dependent indicator that, with all else equal, favours journals that publish more papers (see Table 4). Compared to EF, the AI ranking mostly favours disciplines psychology (+11 positions), economics, management, and business (+7 positions), and political science (+6 positions), and it mostly hampers material sciences (−8 positions) and chemistry (−8 positions).

In the rest of the section we study the correlation among the journal rankings (and not the discipline ones) provided by the surveyed indicators. Table 5 contains the Spearman rank-based correlation matrix for the four indicators we consider in this study:

- Indicators IF2, IF5, and AI form a correlation cluster (Franceschet, 2009), with an average mutual correlation of 0.90. These indicators are size-independent, that is, they do not depend on the size (number of published articles) of the journal.
- The size-dependent indicator EF does not join the main correlation cluster. On the other hand, we observed that it is strongly associated with the total number of citations received by the journal in the JCR year (correlation 0.93).

Table 5
Spearman correlation matrix for the surveyed indicators (correlations ≥0.90 are highlighted in bold).

	IF2	IF5	AI	EF
IF2	–	**0.96**	0.84	0.77
IF5	**0.96**	–	**0.90**	0.77
AI	0.84	**0.90**	–	0.76
EF	0.77	0.77	0.76	–

Despite the statistically significant correlation between the journal rankings provided by the impact factor and the eigenfactor metrics, a close analysis reveals that the journal compilations according to the three metrics contain more than a few marked discrepancies (Franceschet (2010); West et al., 2009).

Fig. 3 illustrates the linear regression models for the six pairs of indicators. We notice that:

- The linear model well captures the pairwise relationships among the size-independent indicators (IF2, IF5 and AI). The falling star-shaped scatterplots for the three pairs of metrics are quite similar to each other (the outlier in the first two plots corresponds to CA: A Cancer Journal for Clinicians). The regression equations along with the coefficients of determination

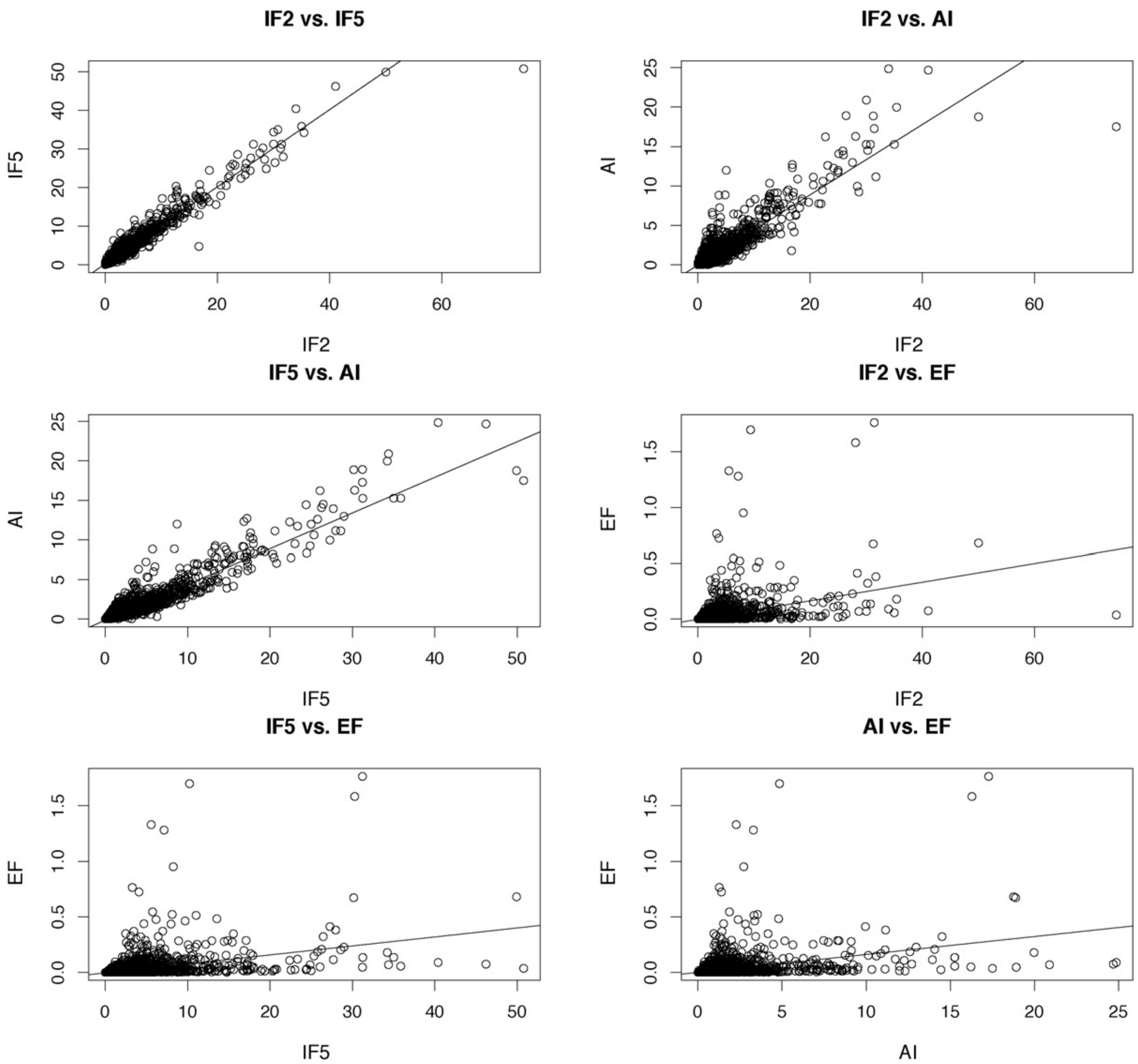

Fig. 3. Linear regression models for each pair of indicators.

R^2 are given below:

$$\mathrm{IF5} = 0.999 \cdot \mathrm{IF2} + 0.233, \quad R^2 = 0.95$$
$$\mathrm{AI} = 0.446 \cdot \mathrm{IF2} - 0.041, \quad R^2 = 0.81$$
$$\mathrm{AI} = 0.452 \cdot \mathrm{IF5} - 0.157, \quad R^2 = 0.88$$

Interestingly enough, the 5-year impact factor for a journal can be approximated by adding to the journal 2-year impact factor score a constant value of 0.233.

- On the other hand, the linear model does not explain well the mutual relationships between EF and the size-independent indicators. Indeed, the coefficients of determination are much lower: 0.17 (IF2), 0.18 (IF5), and 0.17 (AI). Many more outliers are present, in particular journals with high EF scores with respect to the values for the other metrics. Also in this case, the shapes of the corresponding three scatterplots resemble each other.

3. Related work

The rankings provided by impact factor and eigenfactor methods have been object of previous investigation. Bollen et al. (2006) compare journal PageRank with 2-year impact factor on 2003 science edition JCR dataset. The Spearman rank correlation between the whole rankings is 0.61, while that for physics, computer science, and medicine is 0.59, 0.63, and 0.77, respectively. The main difference between journal PageRank and eigenfactor is that the former includes journal self-citations in the computation. Moreover, to enforce irreducibility of the citation matrix, and hence convergence of the method, in both approaches the original citation matrix is perturbed by adding artificial transitions, with low probability, among journals. In the eigenfactor method the weight of each artificial transition is proportional to the number of article published by the target journal, whereas in the journal PageRank approach artificial transitions are uniformly distributed over all journals. Franceschet (2010) makes a thorough comparison of the rankings provided by eigenfactor and 5-year impact factor for science and social science journals included in the 2007 edition of JCR. The author finds that, although the two bibliometric measures are generally statistically correlated, they also significantly diverge in some cases. The two methods diverge more for the hard sciences, including physics, engineering, material sciences, and computer sciences, than they do for the geosciences, for biology-medical disciplines, and for the social sciences. Moreover, the author identifies the science and social science journals with the highest diverging ranks with respect to eigenfactor and 5-year impact factor as well as with respect to article influence and 5-year impact factor.

Furthermore, Davis (2008) compares the rankings according to eigenfactor and 2-year impact factor for 165 journals from the category medicine (general and internal). The author finds a significant correlation between the two measures (Spearman 0.84), and an even higher association (Spearman 0.95) between eigenfactor and the total number of citations. An impressive correlation between the PageRank method and the total number of citations is also noticed in (Chen, Xie, Maslov, & Redner, 2007; Ma, Guan, & Zhao, 2008), where the PageRank algorithm is used to find the influence of scientific papers instead of that of scientific journals. In particular, Chen et al. (2007) analyze all publications in the Physical Review family of journals from 1863 to 2003 and measure a Spearman correlation of 0.91 between the article rankings provided by PageRank and total number of citations. Ma et al. (2008) analyze papers published in period 2000–2005 in the field of molecular chemistry and molecular biology that are included in Web of Science and find a Spearman correlation of 0.98 between PageRank and total number of citations.

A couple of papers use factor analysis with the aim of clustering different scientific impact measures including impact factor and journal PageRank as well as social network centrality indexes (Bollen, de Sompel, Hagberg, & Chute, 2009; Leydesdorff, 2009). They assign impact factor and journal PageRank to different clusters; in particular the latter is aggregated with centrality measures.

Finally, Saad (in press) calculates pairwise correlations between 2-year impact factor, eigenfactor, article influence as well as, interestingly, journal h index for different (small) samples of journals. The author notices that the impact factor is better correlated to article influence than it is to eigenfactor and that there is an important correlation between eigenfactor and journal h index. Similar conclusions are obtained by Rousseau and STIMULATE 8 Group (2009) on a sample of 77 journals from different fields. Rousseau investigates the relationships between 2-year and 5-year impact factors. Results indicate that the two measures lead statistically to the same rankings per category and that, in the majority of cases, the medium term impact factor scores are larger than the short term ones (Rousseau, 2009). These conclusions are coherent with the outcomes of the present investigation.

4. Conclusion

We performed a thorough comparison of four main indicators of journal influence, namely 2-year and 5-year impact factors, eigenfactor and article influence. We retrieved the indicator scores for all science and social science journals indexed in Thomson Reuters JCR 2007 and studied the form of the retrieved samples in terms of deviation, concentration, entropy, and symmetry. We compared the empirical distributions of the journal influence measures with the theoretical lognormal model. Furthermore, we investigated the variability of the indicator scores across the different disciplines forming the sciences and the social sciences. Finally, we analysed the correlation between the journal rankings provided by the mentioned indicators.

The eigenfactor is the indicator with the highest deviation, concentration, and skewness and with the lowest entropy. On the other hand, 5-year impact factor is the measure with the lowest deviation, concentration, and skewness, and it is the most entropic (least foreseeable) one. Article influence and 2-year impact factor are close to 5-year impact factor with respect to the considered distribution variables. The empirical distribution of the eigenfactor fits very well the theoretical lognormal model, those of impact factors match moderately well the lognormal model, and that of article influence is not well explained by the lognormal distribution. The article influence measure is the most stable indicator across different disciplines, while the eigenfactor scores widely vary across fields. Finally, impact factors and article influence form a cluster of mutually correlated indicators, and eigenfactor and total number of citations produce similar journal rankings.

As observed by Rousseau and STIMULATE 8 Group (2009), the association between article influence and impact factor and the fact that the former measure is freely accessible at `eigenfactor.org` with a limited temporal delay with respect to the publication in JCR are good news for developing countries[6] who might not have the resources to procure access to commercial data sources.

References

Althouse, B. M., West, J. D., Bergstrom, C. T., & Bergstrom, T. (2008). Differences in impact factor across fields and over time. *Journal of the American Society for Information Science and Technology*, *60*(1), 27–34.

Amin, M., & Mabe, M. (2000). Impact factors: use and abuse. *Perspectives in Publishing*, *1*, 1–6.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, *2*(1), 1–52.

Bergstrom, C. T., West, J. D., & Wiseman, M. A. (2008). The eigenfactor metrics. *Journal of Neuroscience*, *28*(45), 11433–11434.

Bollen, J., de Sompel, H. V., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, *4e*, 6022.

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

Campbell, P. (2008). Escape from the impact factor. *Ethics in Science and Environmental Politics*, *8*, 5–7.

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, *1*(1), 8–15.

Davis, P. M. (2008). Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *Journal of the American Society for Information Science and Technology*, *59*(13), 2186–2188.

Franceschet, M. (2009). A cluster analysis of scholar and journal bibliometric indicators. *Journal of the American Society for Information Science and Technology*, *60*(10), 1950–1964.

Franceschet, M. (2010). The difference between popularity and prestige in the sciences and in the social sciences: a bibliometric analysis. *Journal of Informetrics*, *4*(1), 55–63.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*, *295*(1), 90–93.

Garfield, E., & Sher, H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, *14*, 195–201.

Glänzel, W., & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics*, *53*(20), 171–193.

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology*, *60*(7), 1327–1336.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, *60*(2), 348–362.

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing & Management*, *44*(2), 800–810.

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

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Raan, A. F. J. V. (1985). The application of bibliometric indicators: important field- and time-dependent factors to be considered. *Scientometrics*, *8*(3–4), 177–203.

Pareto, V. (1897). *Cours d'économie politique* (Vol. 2). Lausanne: Université de Lausanne.

Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, *57*(1), 1–11.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, *12*(5), 297–312.

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

Rousseau, R. (2009). What does the Web of Science five-year synchronous impact factor have to offer? *Chinese Journal of Library and Information Science*, *2*(3), 1–7.

Rousseau, R., & STIMULATE 8 Group (2009). *On the relation between the WoS impact factor, the eigenfactor, the SCImago journal rank, the article influence score and the journal h-index* (Tech. rep.). Retrieved January 1, 2010 from http://eprints.rclis.org/16448/.

Saad, G. (in press). Convergent validity between metrics of journal prestige: The eigenfactor, article influence, h-index scores, and impact factors. *Journal of the American Society for Information Science and Technology*.

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

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

Shockley, W. (1957). On the statistics of individual variations of productivity in research laboratories. In *Proceedings of the IRE 45* (pp. 279–290).

Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2008). Effectiveness of journal ranking schemes as a tool for locating information. *PLoS ONE*, *3e*(2), 1683.

West, J., Althouse, B., Bergstrom, C., Rosvall, M., & Bergstrom, T. (2009). *Eigenfactor. org—Ranking and mapping scientific knowledge*. Accessed January 1, 2010 at http://www.eigenfactor.org.

Wilson, C. S. (1999). Informetrics. *Annual Review of Information Science and Technology*, *34*, 107–247.

[6] STIMULATE is an international training program in information management aimed at young scientists and professionals from developing countries with the purpose of developing the professional skills of all participants and transferring the acquired knowledge and skills to colleagues and stakeholders in their home countries.