MATHEMATICAL RELATIONS BETWEEN IMPACT FACTORS AND AVERAGE NUMBER OF CITATIONS

L. EGGHE
LUC, Universitaire Campus, B-3610, Diepenbeek, Belgium* and UIA, Universiteitsplein 1,
B-2610, Wilrijk, Belgium

Abstract—Instead of the two-year impact factor as used in the *Journal Citation Reports*, there is much in favor of using x-year impact factors ($x > 0$). These impact factors are studied as a function of x and compared with the average number of citations per paper to papers that appeared in the journal x years ago. It is shown that both are equal if and only if the derivative of the impact-factor function is zero. Based on this, a simple classification of impact-factor curves versus mean citation curves is established and examples are given. These results are also applied to recent practical data that were obtained by Rousseau.

1. INTRODUCTION

Let $c(k)$ be the number of citations that a journal receives in a fixed year to articles published k years before. Let $p(k)$ be the number of articles in that same journal k years before. Let

$$\alpha(k) = \frac{c(k)}{p(k)}. \tag{1}$$

The impact factor of a journal (in a given fixed year), as defined by Garfield [2] is given by

$$IF = \frac{c(1) + c(2)}{p(1) + p(2)}. \tag{2}$$

For example, in year 1987 (year 0)

$$IF = \frac{\text{number of citations in 1987 to articles in the journal of years 1985 and 1986}}{\text{number of articles in the journal of years 1985 and 1986}}.$$

Here 1986 is year 1 and 1985 is year 2.

However, as pointed out in [4] (see also [3]), in some research areas this value of IF is not ideal in the sense that, instead of considering two years, another number of years might produce a higher impact factor. If the two-year impact factor of Garfield is good for measuring the importance of life science journals, the same cannot be said for mathematics journals. As pointed out in [4] (see also [3]), one obtains the highest impact factor when taking a four-year period, that is, when calculating:

$$IF(4) = \frac{c(1) + c(2) + c(3) + c(4)}{p(1) + p(2) + p(3) + p(4)}. \tag{3}$$

One generally can define (cf. [3], [4])

$$\text{IF}(n) = \frac{\sum_{k=1}^{n} c(k)}{\sum_{k=1}^{n} p(k)}. \tag{4}$$

Of course, IF = IF(2).

The practical examples in [4] or [3] of the functions $\alpha(n)$ and IF(n) for some mathematics journals show that when IF is maximal in n, then $\alpha(n-1) >$ IF($n-1$) and $\alpha(n+1) <$ IF($n+1$), in all cases (see further on in this article for a reproduction of these examples). Also, in [4], it is shown that IF obtains its maximum later than α does. This second result has been shown mathematically in [4], but the first one is not explained.

All these regularities gave us the idea of making a mathematical theory on IF and α, as functions of time. In this article, time will be a continuous variable. This is not an unuseful abstraction; on the contrary, once a classification of the form and interrelation of the continuous curves of IF and α has been established, we also have a model for the values of $\alpha(n)$ and IF(n) for $n = 1,2,3,4,\ldots$.

In Section 3, we will develop the theory on the interrelationship between the functions α and IF. As a consequence, we will again derive Rousseau's result, which states that IF obtains its maximum later than α does, but we will also be able to explain the evolution of the α and IF values as obtained experimentally in [4] for some mathematics journals. We finally end up with a classification of α and IF curves: we will show that we have only two different interrelationships between α and IF, one of them even being very unlikely to be encountered in practice.

2. MATHEMATICAL NOTATIONS

In this article we use notations from mathematical analysis that might not be so well recognized by the nonmathematical reader. To overcome this problem, we will now indicate and explain most of the notations (see also [1], if necessary).

$\mathbb{R}^+$ = the set of the positive real numbers

$[a,b]$ = the closed interval with beginning a and ending b
= $\{x$ such that $a \le x \le b\}$

$]a,b]$ = $\{x$ such that $a < x \le b\}$

The definitions of $[a,b[$ and $]a,b[$ are similar. $]a,b[$ is called an open interval.

$\alpha(0+)$ = $\lim_{x \to 0} \alpha(x)$ is the limit of the function $\alpha(x)$ for x going to zero from the right side.

$\alpha(0)$ =: $\alpha(0+)$: we define $\alpha(0)$ to be $\alpha(0+)$ (i.e., =: means "definition")

α' = the derivative of α

α'' = the second derivative of α

$x \in A$ = x is an element of the set A

$V \subset]0,x[$ = the set V is a subset of $]0,x[$

$\square$ = the end of a proof of a statement

$\max_{x \in [0,x_0]} \alpha(x)$ = the maximum of the function $\alpha(x)$ for x ranging from 0 to x_0 (inclusive)

$\int_a^b c(x)dx$ = the integral of the function $c(x)$ over the interval $[a,b]$.

For discrete functions on the positive numbers, this is nothing but the sum $\sum_{k=a}^{b} c(k)$.

3. THEORY OF THE FUNCTIONS α AND IF AND EXPLANATIONS OF EXPERIMENTAL RESULTS

3.1 *Definitions*

Let us fix a year (to be considered as the present year from which we will study the past) and a journal. Let $x > 0$. Denote by $c(x)$ the distribution function of the citations in this year to articles in this journal that were published x years ago. Analogously, let $p(x)$ denote the distribution function of the articles in this journal, x years ago.

Remark 1. Of course, only $\int_0^1 c(x)dx$, $\int_1^2 c(x)dx$ and so on are known in practice (the same for the function p). Nevertheless, since time is continuous, we can interpret $c(x)$ and $p(x)$ for continuous time $x > 0$. We will see that our theory will yield new results for continuous $x > 0$. Then taking $x = 1,2,3,\ldots$, we will see that our theory is able to predict regularities that are encountered in practical examples.

Let $\alpha(x) = \dfrac{c(x)}{p(x)}$, for $x > 0$.

We assume the following, very weak and natural, properties of the functions α and p.

$$\alpha = [0,M] \to \mathbb{R}^+$$

(with M very high; in practice, this is the number of years ago that the journal was founded).

$$\alpha(0) = \alpha(0+) = \lim_{x \gtrsim 0} \alpha(x).$$

α is continuously differentiable and $\alpha'(x) > 0$ for every $x \in [0,\delta]$, for a certain $\delta > 0$. We also suppose that there exists an $x \in \,]0,M[$, such that $\alpha'(x) = 0$ and $\alpha''(x) < 0$.

Remark 2. Note that the second-to-last property of the function α does not follow from the other ones; to see this, just consider the function $\alpha(x) = x + x \sin 1/x$.

Remark 3. The fact that we require α' to be strictly positive, at least in a small interval starting in zero, is very natural and is always encountered in practice. Indeed, requiring $\alpha' > 0$ in a small interval starting in zero is equivalent with saying that α is strictly increasing in the beginning. This is natural, except for those articles that are never cited (α is constantly zero then), every article starts influencing others for a small or longer period in time. Gradually, the article is used more and more in this time period. In the life sciences, this time period is approximately two years; for mathematics this period is approximately four years. How long this period is is of no importance here; the existence of it suffices in our arguments. That is why we suppose the existence of a $\delta > 0$ (possibly small) such that $\alpha' > 0$ on $[0,\delta]$.

Remark 4. Requiring that $\alpha'(x) = 0$ and $\alpha''(x) < 0$ for a certain $x \in \,]0,M[$ is also very logical: the older the volumes of a journal are, the less they are used, at least after a certain time; it is never true that the older a journal volume is, the more it is used; this can be so in the beginning years but certainly not forever; $\alpha'(x) = 0$ and $\alpha''(x) < 0$ for a certain $x > 0$ expresses this fact.

About the function p we just assume

$$p = [0,M] \to \mathbb{R}^+$$

$p(x) > 0$ for $x > 0$ and p continuous.

Finally, define

$$\text{IF}(x) = \frac{\displaystyle\int_0^x \alpha(x')p(x')dx'}{\displaystyle\int_0^x p(x')dx'} \tag{5}$$

for every $x > 0$.

3.2 *Basic theory of* α *and* IF

All results taken from infinitesimal analysis can be found, for instance, in [1].

PROPOSITION 1

IF$(0+) = \alpha(0+)$.

Proof. Using de l'Hôspital's rule, we see that

$$\text{IF}(0+) = \lim_{x \gtrsim 0} \text{IF}(x) = \lim_{x \gtrsim 0} \frac{\alpha(x)p(x)}{p(x)} = \alpha(0+). \qquad \square$$

THEOREM 1

IF$'(x)$ *has the same sign as* $\alpha(x) - \text{IF}(x)$.

Proof.

$$\text{IF}'(x) = \frac{\int_0^x p(x')dx' \cdot \alpha(x)p(x) - \int_0^x \alpha(x')p(x')dx' \cdot p(x)}{\left(\int_0^x p(x')dx'\right)^2}$$

$$= \frac{\alpha(x)p(x) - \text{IF}(x)p(x)}{\int_0^x p(x')dx'}$$

$$= \frac{p(x)}{\int_0^x p(x')dx'} [\alpha(x) - \text{IF}(x)].$$

Since $\dfrac{p(x)}{\int_0^x p(x')dx'}$ is strictly positive, the result follows. $\qquad \square$

COROLLARY 1

IF$'(x) = 0$ *if and only if* $\alpha(x) = \text{IF}(x)$.

Proof. This follows immediately from the proof of Theorem 1. $\qquad \square$

Definition 1. Let $x_0 > 0$ be the first for which $\alpha'(x_0) = 0$ and $\alpha''(x_0) < 0$. Let $x_1 > 0$ be the first for which IF$'(x_1) = 0$. We suppose in this section that x_1 exists.

That x_0 exists was supposed in Section 3.1. x_1, however, does not have to exist. In Section 3.3 we will see when x_1 exists or not.

THEOREM 2

For every $x \in]0, x_0]$, we have that IF$(x) < \alpha(x)$.

Proof. Since there exists a $\delta > 0$ such that, for every $x \in [0, \delta]$ we have $\alpha'(x) > 0$ and since x_0 is the first for which $\alpha'(x_0) = 0$, we have, since α' is continuous that $\alpha'(x) > 0$ for every $x \in [0, x_0[$. Hence, the function α is strictly increasing on the interval $[0, x_0[$, hence also on the interval $[0, x_0]$ (since α is continuous). Let $x \in]0, x_0]$ arbitrary. So $\alpha(x) > \alpha(y)$ for every $y \in [0, x[$. Since $p(y) > 0$ for every $y > 0$, we hence have

$$(\alpha(x) - \alpha(y))p(y) > 0, \qquad \text{for every } y \in]0, x[. \tag{6}$$

Since the function

$$y \to (\alpha(x) - \alpha(y))p(y)$$

is continuous, we see that there exists an $\epsilon > 0$ and an open interval $V \subset]0,x[$ such that

$$(\alpha(x) - \alpha(y))p(y) > \epsilon, \tag{7}$$

for every $y \in V$.

Properties (6) and (7) together now yield

$$\int_0^x (\alpha(x) - \alpha(y))p(y)dy > \int_V (\alpha(x) - \alpha(y))p(y)dy > 0.$$

Since $p(y) > 0$ for every $y > 0$ and since p is continuous, we have (by the same argument as above) that

$$\int_0^x p(y)dy > 0.$$

The last two results together yield

$$\frac{\int_0^x (\alpha(x) - \alpha(y))p(y)dy}{\int_0^x p(y)dy} > 0.$$

Hence

$$\alpha(x) - \frac{\int_0^x \alpha(y)p(y)dy}{\int_0^x p(y)dy} > 0$$

or $\alpha(x) > \mathrm{IF}(x)$, for every $x \in]0,x_0]$.   □

COROLLARY 2 (ROUSSEAU [4])

$$x_1 > x_0.$$

Proof. For every $x \in]0,x_0]$, Theorem 2 yields that $\mathrm{IF}(x) < \alpha(x)$. By definition of x_1 and by Corollary 1 it follows that $\alpha(x_1) = \mathrm{IF}(x_1)$. Hence, $x_1 > x_0$, necessarily.   □

COROLLARY 3

Suppose that the function α has only one local maximum. Then there exists an $\epsilon > 0$ such that

For every $x \in]x_1 - \epsilon, x_1[: \alpha(x) > \mathrm{IF}(x)$ and
for every $x \in]x_1, x_1 + \epsilon[: \alpha(x) < \mathrm{IF}(x)$.

Proof. We have that

$$\alpha(x_0) = \max_{x \in [0,x_0]} \alpha(x),$$

as is shown in the first part of the proof of Theorem 2 and $\alpha'(x_0) = 0$ and α has only one local maximum. Hence, the function α is decreasing in $[x_0, M]$. But $\alpha(x_1) = \mathrm{IF}(x_1)$ by definition of x_1 and by Corollary 1 and $x_0 < x_1$ by Corollary 2. This completes the proof.
                                                                                            □

COROLLARY 4

For every $x \in]0,x_1]$, $IF'(x) > 0$ and hence IF is strictly increasing on $[0,x_1]$.

Proof. From Theorems 1 and 2 we have already that IF is strictly increasing on the interval $]0,x_0]$. Suppose now that IF does not increase strictly on $]x_0,x_1]$. So, there is a $y \in]x_0,x_1[$ and an $\epsilon > 0$, such that $IF'(x) \leq 0$ for every $x \in]y - \epsilon, y + \epsilon[$ (since IF is continuous). Since $IF'(x) > 0$ for every $x \in]0,x_0]$, we must have that $IF'(y) = 0$ for a certain $y_0 \in]x_0,y] \subset]0,x_1[$ (hence, $y < x_1$).

But x_1 was supposed to be first for which $IF'(x_1) = 0$, a contradiction. Hence, IF is strictly increasing on the interval $]0,x_1]$. From this, of course, it follows that $IF'(x) > 0$, for every $x \in]0,x_1]$.   □

COROLLARY 5

For every $x \in]0,x_1[$

$$IF(x) < \alpha(x).$$

Proof. This follows immediately from Theorem 1 and Corollary 4.   □

Remark 5. By the special form of eqn (5), we can also apply the weighted mean value theorem for integrals (since $\alpha \geq 0$ and continuous in $[0,M]$ and $p > 0$ and continuous in $]0,M]$) (cf. [1], p. 325). Hence we find for every $x \in]0,M]$, there is a point $c_x \in [0,x]$ such that

$$\alpha(c_x) = IF(x).$$

From this, we can conclude that $IF(x)$ takes the α values in a "retarded" way (since $c_x \leq x$).

3.3 *Classification of the possible relations between α and IF*

In all cases, both α and IF start (in 0) at the same point, by Proposition 1. If the value in 0 is not 0, this shows the existence of invisible colleges, which can indeed be the case. Only invisible colleges are responsible for $\alpha(0) > 0$: fast, direct communication of research results through preprints or oral communication; journals are not likely to react upon each other with this speed (except in special circumstances where the publication date of a journal differs substantially from the date that is printed on it).

In all cases, α starts increasing (by assumption), and hence, since IF is a weighted average of α values, the same is true with IF. Because it is an average, $IF(x) < \alpha(x)$ in a certain interval $[0,y]$, $y > 0$.

In Section 3.1 we showed that it was natural to suppose that there is an $x > 0$ such that $\alpha'(x) = 0$ and $\alpha''(x) < 0$. So, by Definition 1, x_0 exists. Thus, in any case, the function α has its first local maximum in x_0.

Now, after x_0, we can have that the function α keeps on decreasing, or it starts increasing in a certain point $y > x_0$. Suppose first that α keeps on decreasing. In this case x_1, being the intersection of α and IF (by Corollary 1 and Definition 1) always exists. Indeed, so far IF was increasing; it can only start decreasing after a point where IF' is zero. But then, by Corollary 1, the curves α and IF intersect necessarily. In Definition 1, x_1 was called the first such point.

After x_1 we have (since $IF' < 0$): $\alpha(x) < IF(x)$: see Fig. 1. In fact, this was the case described by Corollary 3.

Suppose now that the function α starts increasing in a certain point. Then we have two cases: either α decreases so much that it intersects IF or α decreases not enough to intersect IF. In the latter case, x_1 does not exist and we have a situation as in Fig. 2.

In the former case, α and IF intersect at a certain point $x_1 > x_0$. However, by Corollary 1, $IF'(x_1)$ must necessarily be zero. So we have again a picture as in Fig. 1 (α starts only increasing after x_1 here), but it might also be that α and IF are tangent in x_1. Since $IF'(x_1) = 0$, we hence have also $\alpha'(x_1) = 0$ and hence the intersection of α and IF is in
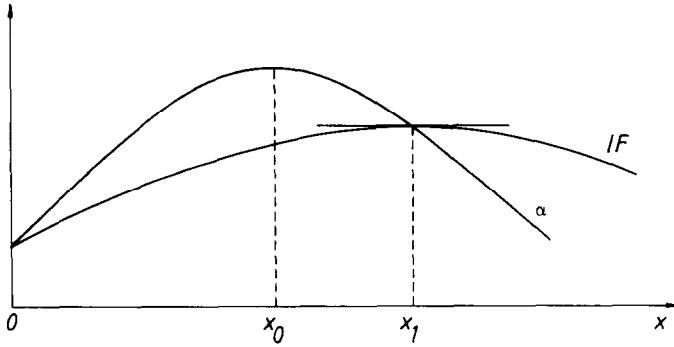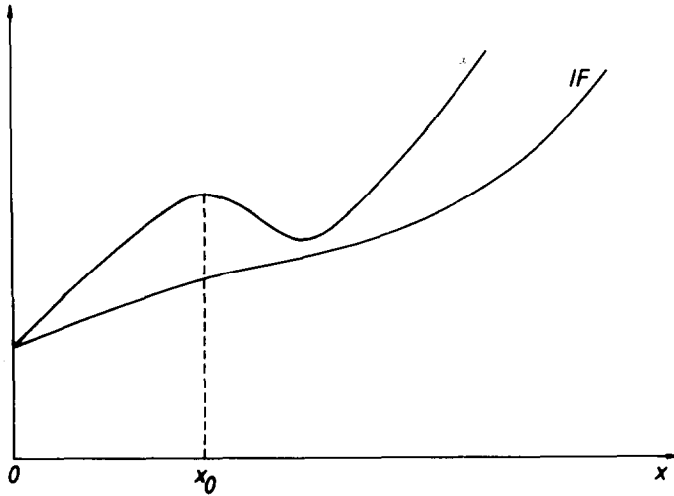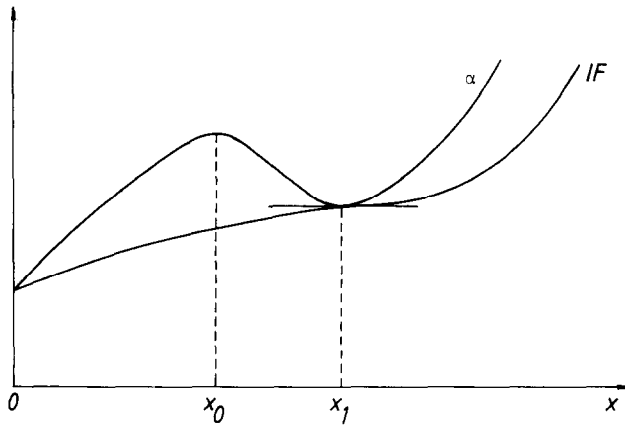
Fig. 1.

Fig. 2.

Fig. 3.

the point where α attains a local minimum. Since α increases then after x_1, IF must do the same since it is a weighted average of α values and since $\alpha(x_1) = \text{IF}(x_1)$. Hence, $\text{IF}' > 0$ after x_1. So by Theorem 1, $\alpha(x) > \text{IF}(x)$ now, at least for an interval $[x_1, y[$ after x_1. So, only in this case (see Fig. 3) IF remains under the curve of α in the whole interval $[0, y[$.

Remark that in the case of Fig. 3, we have also that $IF''(x_1) = 0$. This follows from the fact that

$$\alpha'(x) - IF'(x) = IF''(x) \frac{\displaystyle\int_0^x p(x')dx'}{p(x)} + IF'(x)\left[1 - \frac{\displaystyle p'(x)\int_0^x p(x')dx'}{p(x)^2}\right]$$

and from the fact that $\alpha'(x_1) = 0$ and $IF'(x_1) = 0$ (since $\alpha(x_1) = IF(x_1)$ and by Corollary 1).

Remark 6. As pointed out in Remark 4, the situations in Figs. 2 and 3 are not very likely to happen, but they must be included, since they are mathematically possible.

Examples of all three types of figures can be given. An example of Fig. 1 is

$$\alpha(x) = \sin x + 2, \, x \in [0, 3\pi]$$

$$p(x) = p = \text{a constant}$$

$IF(x)$ can then be calculated, yielding,

$$IF(x) = 2 + \frac{1 - \cos x}{x}.$$

Figure 2 happens in a lot of cases:

$$\alpha(x) = ax^3 + by^2 + cx + d$$

$$p(x) = p = \text{a constant}$$

where $a > 0$, gives

$$IF(x) = \frac{ax^3}{4} + \frac{bx^2}{3} + \frac{cx}{2} + dx.$$

Figure 2 applies in most choices of a, b, c, or d. In fact, to have Fig. 3, conditions on a, b, c, and d apply. These are found by expressing that there must exist an $x > 0$ such that

$$\alpha(x) = IF(x)$$

$$\alpha'(x) = IF'(x).$$

This yields the following conditions

$$a > 0$$

$$c = \frac{8}{27}\frac{b^2}{a}$$

$$x = -\frac{4}{9}\frac{b}{a} > 0.$$

So, if we take f.i. $a = 1$, $b = -6$, $c = 32/3$ and $d = 0$, then we find that α's osculation point is in $x = 2$ and that the tangent point $x_1 = 8/3$.

So all figures are possible.

Of course, as said in Remark 5, only Fig. 1 is likely to happen in practical examples. That this is so will be seen in the next section.

3.4 *Examples*

In [4] and [5], Rousseau shows a few tables of $\alpha(n)$ and corresponding IF(n) values, for discrete n, for several mathematics journals in 1985.

His findings are shown in Tables 1–3.

Furthermore, all 43 mathematics journals that Rousseau studied were joined into one "macromathematics" journal. The data for this (see [4]) are shown in Table 4.

In each of these tables we underlined the local maxima of α and of IF. Of course, working discretely, the real local maxima might occur on noninteger x values, but this is not important here. The point is that all tables above are explained through Fig. 1, which was derived from our continuous theory. We see indeed always that $x_0 < x_1$ and furthermore, if IF(n) is a local discrete maximum one has IF$(n - 1) < \alpha(n - 1)$ and IF$(n + 1) > \alpha(n + 1)$ in all four tables ($n = 3$ in Table 1, $n = 8$ in Table 2, $n = 5$ in Table 3, and $n = 4$ in Table 4). That IF is not really maximal in those values of n can also be deduced by the fact that in these n, IF$(n) \neq \alpha(n)$, which is a requirement (by Corollary 1): this can never be deduced through a discrete theory, but, conversely, our continuous theory was able to explain the above examples and furthermore gives a detailed insight in the relationship between α and IF.

Of course, other examples (agreeing more with Figs. 2 and 3) might exist and can be explained in this way. We point out however that the examples above are most likely to occur. Examples conforming with Figs. 2 or 3 can be expected in the case of "revived interest" (α increasing at least twice): an old method that can be reapplied in the solution of a new problem, a viewpoint that becomes active again. Also in these cases, Figs. 2 and 3 predict the interrelationship between IF and α.

Table 1. *Acta Math-Djursholm*

n	1	2	3	4	...
$\alpha(n)$	1.471	<u>2.286</u>	2.278	1.636	.....
IF(n)	1.471	1.839	<u>2.000</u>	1.887	.....

Table 2. *Adv Math*

n	1	2	...	7	8	9
$\alpha(n)$	0.600	1.105	.....	<u>2.041</u>	1.767	1.092
IF(n)	0.600	0.846	.....	1.256	<u>1.317</u>	1.283

Table 3. *Ann Sci Ecole Norm S*

n	1	2	3	4	5	6	...
$\alpha(n)$	0.350	0.083	0.095	<u>0.625</u>	0.421	0.063	.....
IF(n)	0.350	0.205	0.169	0.292	<u>0.315</u>	0.282	.....

Table 4. Macromathematics journal

n	1	2	3	4	5	...
$\alpha(n)$	0.371	<u>0.595</u>	0.580	0.572	0.515	.....
IF(n)	0.371	0.483	0.515	<u>0.529</u>	0.526	.....

REFERENCES

1. De Lillo, N.J. Advanced Calculus with Applications. London: Collier Macmillan Publ.; 1982.
2. Garfield; E. Citation analysis as a tool in journal evaluation. Science 178: 471–479; 1972.
3. Rousseau, R. Impact factors calculated over several periods with an application to mathematical journals. Proceedings 3rd National Conference with international participation on Scientometrics and Linguistics of the Scientific Text, Varna (Bulgaria), 14–16 May 1987. To appear.
4. Rousseau, R. Citation distribution of pure mathematics journals. Proceedings 1st International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval (L. Egghe and R. Rousseau, eds.) LUC, Diepenbeek, 24-28 August 1987, pp. 249–262, 1988.