ISSN-0971-6696

JISSI The International Journal of Scientometrics and Informetrics

Volume 1 Number 2 June 1995

Guest Editors :

Hildrun Kretschmer
Wolfgang Glänzel

Special Issue

In This Issue

Selected Papers from the Proceedings of the

Fourth International Conference on Bibliometrics, Informetrics and Scientometrics, Berlin 11-15 September 1993

bis BRZARK INFORMATION SYSTEMS

JISSI The International Journal of Scientometrics and Informetrics

International Society for Scientometrics and Informetrics (ISSI) has been officially founded on 5 October 1994 in Utrecht, The Netherlands, with Dr. Hildrun Kretschmer (Berlin) as the President and Dr. C. Le Pair (Utrecht) as the Secretary-Treasurer.

Membership information of ISSI is included in this issue of JISSI.

JISSI The International Journal of Scientometrics and Informetrics

Volume 1 Number 2 **Berlin Conference (1993) Special Issue** **June 1995**

Editorial Correspondence to :
Subir K Sen
Honorary Executive Editor
85 (old) Devi Nivas Road
P.O. - Motijhil
Calcutta 700 074, India
Fax : + 91-33-551 2180

Business Correspondence to :
Brzark Information Systems (P) Ltd.
112, Humayun Pur
(Near NCC Office)
Safdarjung Enclave,
New Delhi 110 029, India

Subscription Information :
Subscription rates include
air mail surcharge
For detailed subscription
information see inside
Reduced price
(two-third of the net price)
for advanced subscription
for first three years together
Subscriptions may start from
any time of the year

Payments may be made through :
Demand Draft/ Cheque/Money Order
payable to
Brzark Information Systems (P) Ltd.
For further enquiry and
for proforma invoice
write to the Publisher

Advertisements :
JISSI offers publication of
selected advertisements from
interested parties.
For rates and other conditions
please contact the publisher

Publishers :
Probhat Pandey
Bina Pandey

Cover design :
Bibhas Dutta

Price : Single Copy :
US $ 55.00 / Rs. 700.00

Our brave new field of sciento - infor-metrics is fast growing. The field may very well represent the quantitative (and qualitative) assassments of present cultural progress and advent into the information age. In this, JISSI has an important role to play.

The first issue of JISSI was released in March 1995 and was sent to more than 200 persons. The first issue contained the first Newsletter of ISSI and has been well received.

The present (second) issue is a special issue and contains ten selected papers (one in abstract) from the 4th international conference of 1993 held in Berlin.

Incidentally,publication of this issue happens to be timed with the 5th inter-national conference of our field being held at the Rosary College, Illinois, USA under the chairmanship of Professor Dean Michael Koenig.

The next issue of JISSI will be published as another special(joint) issue and will contain the rest of the selected papers from the Berlin conference. It is expected to be published in October 1995.

The delay in publishing these papers could not be avoided for several reasons. Also it took us some time to prepare for launching the journal.

Our regular issues will be forthcoming from March 1996. We solicit comments, criticisms, suggestions about JISSI and submission of papers and articles.

Selected Papers Presented At the Fourth International Conference On Bibliometrics, Informetrics And Scientometrics

In Memory of Derek John de Solla Price (1922-1983)

September 11- 15, 1993, Berlin (Germany)

Organized under the auspices of the Senator for Science and Research of the Land Berlin, Prof. Dr. Manfred Erhardt, by Association for the Promotion of the Fourth Internationl Conference of Science Measurement e.v., Association for Science Studies e.v., KAI e. V. and Deutsche Gesellschaft für Dokumentation e.V.*

Editors :

Wolfgang Glänzel (Hungary) and Hildrun Kretschmer (Germany)

*Sponsored by the Federal Ministry for Science and Technology (BMFT), the German Research Society (DFG), the German Foundation for International Development, the Robert Bosch Foundation, the Ministry for Science and Research of the Land Berlin, the Technnology Foundation (STW, Utrecht), the Donor's Association for Promoting Arts and Sciences in Germany (SV) and the OECD.

Preface

The International Conference on Bibliometrics, Informetrics and Scientometrics held in Berlin, September 11-15, 1993, and organized under the auspices of the Minister of Science and Research of the Land Berlin, Senator Manfred Erhardt, was the fourth in a series of increasingly prominent biennial conferences. Previous meetings were held in Belgium, 1987, in Canada, 1989, and in India, 1991. The fifth conference will take place in Chicago, 1995, and it is already planned to organize the sixth conference in Israel in 1997.

The Berlin Conference was the largest Bibliometrics meeting so far. About 180 scientists from 32 countries attended the conference. In all, 145 papers were presented as lectures or posters. The conference, however, proved to be a notable event in more respects. For the first time bibliometricians from East and West, North and South took the opportunity for personal contacts and for sharing their scientific views with colleagues from almost all regions of the world.

The program covered many topics of our field. Of course, research evaluation was one of the focal points of the meeting. The great number of "evaluative" studies also reflects the increasing demand for such research results in science policy. A part from applied studies (evaluation of research in selected fields and/or special regions, countries or institutions; analysis of scientific collaboration), a series of methodological papers were presented marking the advances in recent "basic" research in scientometrics. Several theoretical papers addresed topics like information measures, library circulation models, and generalizations of classical bibliometric laws (Bradford's law, regularities of growth and obsolescence of scientific literature, etc.) and proved the quality and importance of informetric research in our field.

The Berlin Conference was held in memory of Derek de Solla Price, one of the pioneers and founders of scientometrics, who died in 1983. In the inaugural address by the honorary chairman of the conference, Eugene Garfield, and in the plenary session commemorating Derek de Solla Price, his personality and the impact of his work on present bibliometric research were stressed. A report on the history of the Price Medal, which is awarded by the journal *Scientometrics* since 1984, and a review of the scientific work of the Price awardees were given.

The plenary session "Bridging the gaps between Bibliometrics, Scientometrics and Informetrics" was a further high-light of the conference. The plenary lectures showed the great concern bibliometricians have for the development and future of their field. In this context the foundation of the International Society for Scientometrics and Informetrics (ISSI) during the conference can be seen, too. The society aims at encouraging communication in our field, coordinating organizational tasks and helping to establish scientometrics and informetrics as a scientific discipline.

We would like to thank all the referees and other persons who have helped in reviewing the some 110 papers submitted for publication in the conference proceedings. The selected papers are published in four separate proceedings volumes. Three of them are special issues of the international journals *Scientometrics, Research Evaluation*, and *International Journal of Science and Science of Science*. The fourth volume, focussed on informetric topics, is now being published in parts by the present journal, *JISSI : The International Journal of Scientometrics and Informetrics.*

Wolfgang Glänzel
Hildrun Kretschmer

Guest Editors
(Editors of the Conference Proceedings)

Message of Greeting
from
the Senator of Science and Research,
Prof. Dr. Manfred Erhardt,
to the 4th World Congress on Bibliometrics, Informetrics and Scientometrics
from 11 to 15 September 1993

Ladies and gentlemen,

I welcome most cordially all participants of your 4th World Congress to Berlin, the old and the new German capital.

I was happy to assume the patronage of your meeting and I pay tribute to the excellence of the founder of your young scientific discipline, Derek de Solla Price, the tenth anniversary of whose death you are commemorating.

I am pleased that you have chosen our city as the venue of your congress for it enables Berlin once again to fulfil its new geopolitical function as a central European metropolis : we are the host and meeting place of scientists from the North and the South, from the East and the West.

My special greetings go to all who are able, after the fall of the Wall, to pursue their international contacts free from external and internal constraints at long last.

About 150,000 students are enrolled at 17 tertiary education institutions in the city of science, Berlin, among them three universities with three teaching hospitals, four colleges of art and nine special technical colleges. There exist, in addition, more than 70 extra-university research institutions.

Berlin is also a city with a rich cultural life and a surrounding countryside worth seeing. I hope you will have time to enjoy both outside your work at the meeting.

I wish your congress every success and I wish all of you a pleasant stay in our city.

Yours sincerely,

Signature
Professor Dr. Manfred Erhardt

Berlin, 30 August 1993

Concentration Measures in Random Hierarchical Distributions

Aparna Basu
National Institute of Science Technology & Development Studies
NISTADS, K.S.Krishnan Road, New Delhi 110012, India.*

The Lorenz Curve, used to depict inequality in the attributes of a class of entities, is unable to give an unambiguous comparison of inequality in situations where the curves cross. Commonly used concentration measures, the Gini Coefficient or Pratt index among others, involve loss of information, since they no longer define a unique Lorenz curve. We suggest that a single concentration index may be insufficient to capture the variation in concentration along a Lorenz curve, and propose the use of two indices related to the average concentration and its skewness. These are additional parameters introduced in our earlier mathematical expression for the Lorenz curve, obtained in the context of Bradford's Law using an information theoretic model, which we have called the Random Hierachical model. The indices retain total information from which the Lorenz curve may be reconstructed.

1. Introduction

Concentration or uneven dispersion is a phenomenon encountered frequently in widely varying circumstances, some examples being the inequality of incomes, differential shares of firms in an industry, population dispersion in geographic regions, and so on. In bibliometrics, this is referred to as the *80-20 Rule (Trueswell 1969)* which states that 80% of the needs of the library users are satisfied by 20% of the acquisitions. This rule, though by no means exact, has since found application in various functions related to library science such as journal acquisition, weeding out of stock etc. Concentration phenomena encountered in other areas such as the numbers of articles recovered from a ranked set of journals, the productivity of authors of journal papers, or word use patterns in linguistics etc., are known under the names of Bradford, Lotka, and Zipf Laws *(Bradford 1934; Lotka 1926; Zipf 1949).* All of these phenomena fall into a class of stable inequality distributions. It has been observed that the *80-20 Rule* corresponds to a much higher degree of concentration than is typically observed in situations other than library circulation, and as much as 45% of *'sources'* may be needed to yield 80% of the *'items'*. (The terms *source* and *item* will be used here in the usual generic sense, for example in the case of journal productivity, the journal is the *source* and the article the *item*).

The concept of concentration is intuitively clear, and heuristic measures of concentration are routinely used. The use of some of these measures may lead to ambiguities. This is discussed later in relation to a few commonly used measures of concentration (Section 2). In this article, we hope to show that some of these ambiguities may be avoided by using two measures, simultaneously, to characterize a single set of data. These measures are parameters that have been introduced to generalize the Random Hierarchical (RH) model, proposed by us earlier in the context of Bradford's law (Basu, 1992). In Section 3, the Random Hierarchical model is briefly described, and generalized to include two free parameters α and β, to obtain the Generalized Random Hierarchical (GRH) model. This model is tested for goodness of fit to 10 diverse sets of data

* E - mail : nistads @ sirnetd. ernet. in ; FAX: +91-11-575-4640; Tel : +91-11-572-6406 / 9151

in the area of bibliometrics (Section 4). The two parameters, together with their interpretations, are discussed in the concluding section.

2. Some Common Inequality Measures

One of the most simple measures of concentration is the Concentration Ratio (C_n) often used to depict the level of concentration in an industry. It is the ratio of the sales of the top n firms to the total industry sales. The choice of n is arbitrary, the more comm- only used ratios being C_4, C_5; C_8 and C_{10}. These non-rigorous measures are easy to compute and quite adequate under certain circumstances were it not for cases where not all the ratios C_n in one set of data are either higher or lower than the corresponding ratios in another set. This situation leads to a conflict in comparing the degree of concentration due to which the ratios fail to provide a satisfactory and unambiguous measure.

2.1. Lorenz Curve

A more complete depiction of inequality is offered by Lorenz curve. Proposed initially as a graphical description of income inequality, it showed the total income earned by all persons below a certain income level, as a function of number of persons in that class. Although used initially for income distribution, Lorenz curve may be used in other situations reflecting inequality.

The Lorenz Curve may be formally defined in the following fashion. If x is a random variable taking integer values x_1, x_2, x_3, ... such that $x_1 > x_2 > x_3$ ... etc., and $f(x)$ is the discrete frequency distribution of x such that $f(x_i)$'s are integers, the Lorenz curve may be written in parametric form as $L(\phi_k)$, where

$$\Phi_k = \sum_{i=1}^{k} f(x_1) \qquad \text{.......(1) and}$$

$$L(\Phi_k) = \sum_{i=1}^{k} x_i f(x_i) \qquad \text{.......(2)}$$

(Stuart & Kendall, 1977)

When defined in this way, $L(\Phi_k) > \Phi_k$ everywhere along the Lorenz curve, unlike the original definition of the Lorenz curve where $x1 < x2 < x3$, and $L(\Phi_k) < \Phi_k$ at all points. There is an advantage in defining the curve in the manner described above, since the data are often 'clouded' or less precisely known at the small values of x, and the process of cumulation implicit in the definition of L (Eqn. 2), renders the entire Lorenz curve imprecise when defined in the older way.

The continuous form of equations (1) and (2) may be written as

$$q_k = \frac{\sum_{i=1}^{k} f(x_i)}{\sum_{all\ i} f(x_i)} \qquad \text{.......(1a) and}$$

$$p_k = \frac{\sum_{i=1}^{k} x_i f(x_i)}{\sum_{all\ i} x_i f(x_i)} \qquad \text{.......(2a)}$$

Where $0 < q < 1$, $0 < p < 1$, and the subscript k may now be dropped since p and q are continuous. The graph of p vs. q defines a Lorenz curve. If the Lorenz curve of one data set lies entirely below the Lorenz curve of another, i.e. closer to the diagonal $p = q$, then the concentration of the former is strictly less than that of the latter. The situation is ambiguous when two Lorenz curves cross.

Other measures of concentration like the Gini index, which is related to the area between the Lorenz curve and the diagonal $p = q$, and indices such as the Pratt measure in bibliometrics *(Pratt, 1977)*, are useful compact measures as compared to the Lorenz curve, which involves the same amount of information handling as the original data. However their use involves loss of information, since the Lorenz curve cannot be reconstructed from a knowledge of these indices. Neither is it possible to predict if the Lorenz curves of two sets of data will cross from the values of the corresponding Gini coefficient. It is apparent that single measures of concentration are unable to capture the variation of concentration in the data depicted by the Lorenz curve, or even identify situations where ambiguity is likely to exist.

3. Generalized Random Hierarchical Model.

The Random Hierarchical (RH) model is an information theoretic model, from which we had earlier obtained the most probable distribution that arises from the random but unequal partitioning of a number of entities into a set of classes (Basu, 1992). The condition of inequality in the partitioning is both a necessary and a sufficient condition for the generation of a ranked hierarchy. It is given by the relation

$$p = q - q \ ln \ q \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

In the context of Bradford's law, p is the proportion of articles one may hope to retrieve from a proportion q of journals, selected from the top of a set ranked in order of decreasing *productivity*. The relation essentially defines a Lorenz curve.

As it stands, the *RH* model is not able to take into account the varying levels of concentration that one may expect from the diverse nature of the underlying inequality generation processes that give rise to Lorenz curves in different situations. We had attempted to rectify this problem by introducing a parameter to reflect concentration. *(Basu, 1992)*. However, as seen in section 2, single measures of concentration are not free from ambiguity. We therefore generalize the *RH* model by introducing two parameters α and β, to obtain the Generalized Random Hierarchical *(GRH)* model, defined by the relation.

$$p = [\ q - \beta\ q\ ln\ q\ ]^{\alpha} \quad \ldots\ldots (4)$$

It states that if a fraction q is selected from the top of a set of sources, ranked in decreasing order of productivity, one may expect to recover a fraction p of the total items, where p is determined by the values of the parameters α and β, characteristic to the data. The parameters have been introduced so as to leave the boundary conditions, $q = 0$, $p = 0$, and $q = 1$, $p = 1$, unchanged. This also decides the range of acceptable values of α and β, together with the additional constraint $0 \leq p, q \leq 1$. Theoretical curves plotted for different values of α and β show that Eqn. (4) is able to generate Lorenz curves that intersect each other. Moreover, it is seen that inequality increases in the direction of increasing β and decreasing α (Fig la-b).

3.1. Interpretation of α and β as Measures of Concentration

It is readily seen from Eqn.(4) that the parameters α and β bear some relation to concentration. If β is set to zero, and α to 1, then a situation of perfect equality or zero concentration is obtained with p equal to q everywhere. On the other hand, if α is set equal to zero, then we obtain a situation of total concentration with $p = 1$ even at infinitesimally small values of q.

A possible interpretation of the parameter β follows from a relation similar to Eqn. (4), that was obtained by Burrell from a stochastic model of library circulation for the special case of a Geometric distribution (*Burrell, 1985*). Burrell's model reduces to the *GRH* for the special case $\alpha = 1$, and $\theta = -\beta$ is given by,

$$\beta = -[\mu\ ln(1-1/\mu)]^{-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where the *average productivity* μ is the ratio of the total number of *items* to the total number of *sources*. As observed by Burrell, his relation generates a series of non intersecting Lorenz curves, with zero concentration as μ approaches zero. It reduces to the *RH* model (Eqn. 3) in the limiting case of μ tending to infinity. If we use the expression for β given in Eqn. (5) in the *GRH* model, we conclude that β depends only on the average or gross concentration as represented by μ.

Since the parameter α occurs as a power, its effect is felt more at small values of q. At values of q near unity, it leaves the Lorenz curve virtually unchanged. The smaller the value of alpha, the higher is the concentration in the more productive sources. α is thus value of alpha, the higher is the concentration in the more productive sources. α is thus interpreted as the degree of skewness in the Lorenz curve. The effect of varying the parameters may be observed in the theoretical curves plotted for different values of α and β (Fig. 1-2).

4. Fitting the GRH Model to Bibliometric Data

The *RH* model has not been tested for its ability to reproduce data, except in a single case for the bibliography of Agriculture *(Lawani 1973)* where a parameter α had been introduced to provide best fit to the data *(Basu, 1992)*. Even in other fairly extensive studies of different models of Bradford's Law and their fit to data, no model of this form seems to have been represented. *(Asai, 1981, Qiu, 1990)* even though a similar model had been discussed in the slightly removed context of library circulation *(Burrell, 1981, Burrell 1985)*.

As a test of the *GRH* model the data selected are those used recently by Burrell in testing Sichel's Generalized Inverse Gaussian Process (*GIGP*) model (*Burrell, 1993, Sichel 1985*). These representative data sets have marked the inception and development of the field of bibliometric studies, and relate to a wide variety of bibliometric phenomena (Physics Abstracts, Keenan and Atherton 1964; Schistosomiasis and Mast Cells, Goffman and Warren 1969; Operations Research, Kendall 1960; MEDLARS, Nelson andTague 1985; Chemical Abstracts, Lotka 1926; Economics, Ravichandra Rao 1990; Geophysics and Lubrication, Bradford 1934; Library of Saskatchewan Circulation data - Tague and Ajiferuke, 1987) The data characteristics are suffciently varied, from short to long data sets (7 to 42 data points), average 'productivity' μ ranging from 46 to about 2, and data with a pronounced Groosdroop *(Keenan & Atherton, 1964)*. The basic characteristics of the data are given in Table I, where the data sets have been ordered in terms of decreasing productivity μ and therefore decreasing *Concentration* according to the definition of Burrell.

From the data, a set of *(p,q)* values are generated for each set, by calculating the cumulative number of *items* that may be obtained from a collection of *sources* selected from the top of a set ranked in order of decreasing productivity. The parameter β obtained using Eqn. (5), depends only on the average concentration.

Table I : Characterstics of Some Bibliometric Data Sets

Data	Symbol	Classes	m
Physics Abstracts	P	39	46.64
Economics	E	24	5.41
Schistosomiasis	S	34	5.26
Mast Cells	A	18	4.01
Operations Research	O	14	3.84
Medlars	M	42	3.83
Geophysics	G	13	3.49
Chemical Abstracts	C	44	3.28
Library : Saskatchewan	L	17	2.37
Lubrication	U	7	2.29

To obtain α, the *GRH* model is fit to each data set using a nonlinear regression procedure in the statistical package **SYSTAT** and the parameter α that gives the best fit to data is determined for each set. The *F-test* and *t-test* indicate a good degree of fit, with α significant at the 99.5% level for all data sets except one, the Keenan Atherton data on Physics abstracts (Table II). The data sets are plotted in Figs. (3a-c) along with model estimates, and a plot of the position of the data in α -β space (Fig 4). It may be seen that the more concentrated data sets have higher values of β and lower values of α as compared to the less concentrated data, as also observed from the Lorenz curves in Fig 1-2.

Another point that needs to be emphasised is that all the information contained in the original Lorenz curve can be recovered from the Generalized Hierarchical model by substituting for the values of α and β in Eqn. 4. This is a complete departure from the usual extraction of concentration indices, which involve real loss of information. The parameters α and β may thus be regarded roughly as the counterparts of the mean and standard deviation of a normal distribution, in the case of skew distributions, which provide complete information regarding the distribution.

Table II : Goodness of Fit : F-values and t_α

Data	F-value	t_α
Schistosomiasis	11061	60.29
Physics Abstracts	7011	18.47
Lubrication	2503	35.38
Lib Saskatechewan	246800	239.50
Economics	7763	62.82
Operations Research	5119	49.18
Geophysics	4504	44.47
Chemical Abstracts	15170	97.17
MEDLARS	7293	83.29
Mast Cells	6349	37.86

5. Indicators of Concentration

In this section we shall consider how to obtain a concentration index from the parameters α and β . First we shall visually order the data in decreasing order of concentration by examining their Lorenz curves. We then compute an index K,

$$K = 1/2 + (\beta - \alpha) / 2 \quad \ldots\ldots\ldots\ldots\ldots (6)$$

For $0 < \alpha, \beta < 1$, K appears to be a good measure of concentration. If $\alpha = 1$ and $\beta = 0$, it follows that $K = 0$, implying that there is no concentration. If $\alpha = 0$ and $\beta = 1$, the concentration is maximum and K is equal to 1. We also find that the values of K reflect the intuitive visual ordering of the data sets with respect to concentration. (Table III) On the other hand, the proportion A of *sources* that yields 80% of the *items* does not follow this ordering. This is also found to be the case for the intitial ordering that was done on the basis of the average concentration μ, which is effectively equivalent to ordering the data in terms of decreasing β. From Table III, it may be concluded that there is no ambiguity in situations where $\alpha_1 < \alpha_2$ and $\beta_1 > \beta_2$, in which case the concentration of data set 1 is strictly greater than that of data set 2. Ambiguity arises only in situations where this ordering is not maintained, as in the case of the data on Operations Research as compared to Economics, or the data of Lubrication as compared to that of the Library of Saskatchewan.

We end with a short prescription for obtaining an approximate estimate of α, without recourse to elaborate data fitting procedures. if we find the point (p,q) on the Lorenz curve at which the tangent is parallel to the diagonal, then α may be obtained form the relation

$$\alpha = (q /p) /(1-\beta /(1 - \beta \ln q)) \quad \ldots\ldots (7)$$

evaluated at the point where the tangent to the Lorenz curve is parallel to the diagonal. β may be obtained from Eqn (5). The parameter that corresponds with our intuitive understanding of concentration is K, which may then be obtained from Eqn(6).

Summary

In this paper we have tried to point out

Table III : Measures of Concentration

Data	α	β	K	A
Physics Abstracts	.35	.99	.82	20
Schistosomiasis	.42	.90	.74	28
Mast Cells	.53	.87	.67	30
Operations Research	.54	.86	.66	40
MEDLARS	.58	.86	.64	40
Chemical Abstracts	.58	.84	.63	40
Economics	.69	.90	.61	38
Geophysics	.67	.85	.59	40
Lubrication	.74	.76	.51	50
Library : Saskatchewan	.97	.77	.40	52

Note : Bibliometric data arranged in decreasing order of "concentration" by visual inspection of their Lorenz Curves. The concentration index K (Eqn. 6) correctly reproduces this order, but A, derived from an *80/A Rule*, using information essentially from a single point of the Lorenz Curve, does not.

that the usual concentration measures or the Lorenz curve may not give a conclusive comparison of the degree of 'inequality' that may exist in two comparable situations. We therefore suggest the use of two measures b and a which correspond to the average concentration and its skewness. Together, these indices, coupled with the Generalized Random Hierarchical model, contain the full information of the Lorenz curve. There is thus no loss of information, as in computing of other indices of cocentration such as the Gini index. A simple linear combination of the indices a and b gives a concentration index K which satisfies the usual properties of concentration indices, and also agrees with our intuitive notion of the relative degree of concentration as obtained by a visual inspection of the Lorenz curves. A prescription for obtaining the index a without recourse to elaborate data fitting procedures is also provided.

It may be pertinent to enquire why there is a need to use two indices of concentration. We attempt an explanation which traces the ambiguity to the use of the word concentration in common parlance. For example, if two spoons of salt are stiirred into one glass of water, and one spoon into another, we would say that the former is the more *concentrated* solution. If, on the other hand, equal quantities of salt are added to the glasses, and one is stirred, while the other is not, the latter would be more *concentrated* in the technical sense, indicating that a gradient or differential exists in the degree of salt in each layer, it is apparent from this example, that we have, till now, been attempting to measure two different, though related, aspects using a single word, and, by extension, a single measure. In this paper we have tried to separate these aspects into two independent measures α and β to capture the two aspects of concentration.

Acknowledgements

I thank P.S. Nagpal, PVS Kumar, Lalita Sharma and Ravish for discussions and computational assistance.

References

1. Asai, I (1981) A general formulation of Bradford's distribution: The Graph oriented approach, *Journal of American Society for Information Science*, 32, 113-119.
2. Basu, A. (1992) Hierarchical Distibutions and Bradford's Law, *Journal of American Society for Information Science*, 43, 494-500
3. Bradford, S.C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85-86
4. Burrell Q.L., (1985) Library Lore or Statistical Law *Journal of Documentation*, Vol 41, No 1, pp. 24-39.
5. Burrell, Q.L. (1980) A simple stochastic model for library loans. *Journal of Documentation*, 36, 115-32.
6. Goffman, W., & Warren K.S. (1969) Dispersion of papers among journals based on a mathematical analysis of two diverse medical literature. *Nature*, 221, 1205-1207.
7. Keenan, S. & Atherton, P. (1964) *The Journal Literataure of physics*. New York: American Institute of Physics.
8. Kendall M.G. (1960) The bibliography of operational research *Operational Research Quarterly*, 11,31-36.
9. Lawani, S.M. (1973) Bradford's law and the literature of Agriculture. *International Library Review*, 5, 341-350.
10. Lotka, A.J. (1926) The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*. 16, 317-323.
11. Nelson, M.L. & Tague, J.M. (1985) Split size-rank models for distribution of index items, *Journal of American Society for Information Science*, 36, 282-296.
12. Pratt, A.D. (1979) Pratt's measure of class concentration and the Gini index, *Journal of American Society for Information Science*, 30, 317-318.
13. Qiu, L. (1990) An empirical examination of the existing models of Bradford's law, *Information Processing and Management*, 26, 650-672.
14. Ravichandra Rao, I.K. (1990) The Distribution of Scientific Productivity & Social Change. *Journal of American Society for Information Science*, 31, 111-122.
15. Stuart, A. and Kendall M. (1977) Advanced Theory of Statistics, Charles Griffin, London
16. Tague, J., and Ajiferuke, I. (1987). The Markov and the mixed - poisson models of library circulation compared. *Journal of Documentation*, 43, 212-231.

APPENDIX : FIGURES

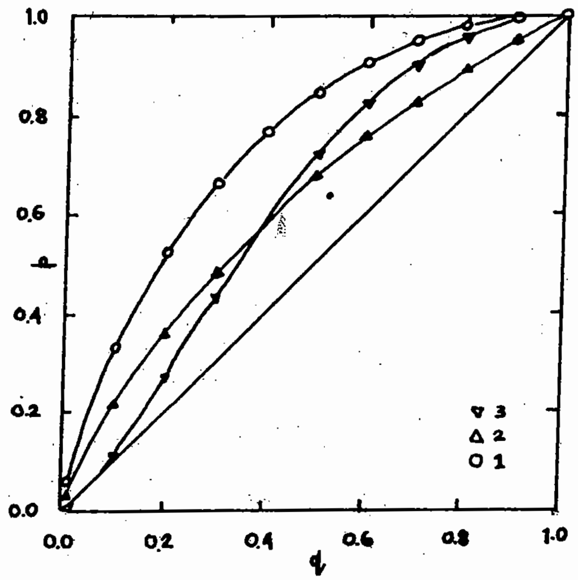


Fig. 1 : Lorenz curves with inequality less than the Random Hierarchical Distribution (RHD)
(1) α=1, β=1; (2) α=1, β=0.5; (3) α=2, β=1

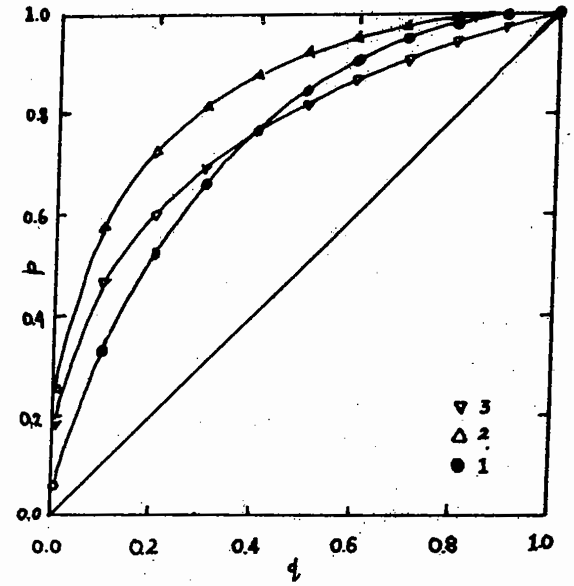


Fig. 2 : Lorenz curves with inequality higher than the Random Hierarchical Distribution (RHD)
(1) α=1, β=1; (2) α=0.5, β=1; (3) α=0.5, β=0.5

PHYSICS ABSTRACTS

ECONOMICS

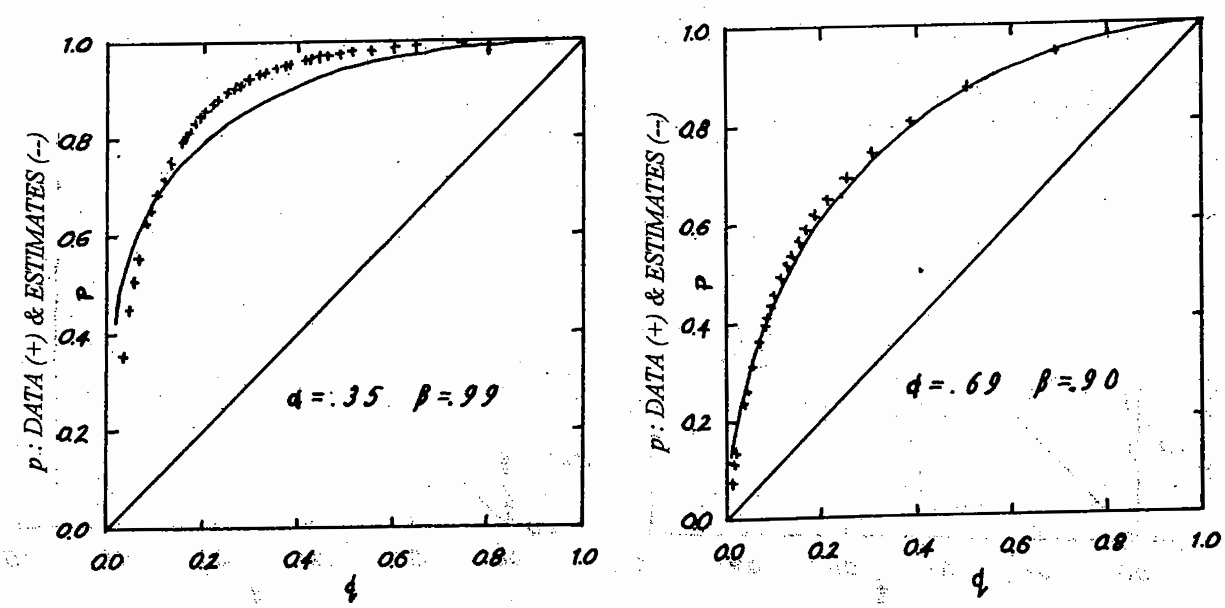


Fig. 3a

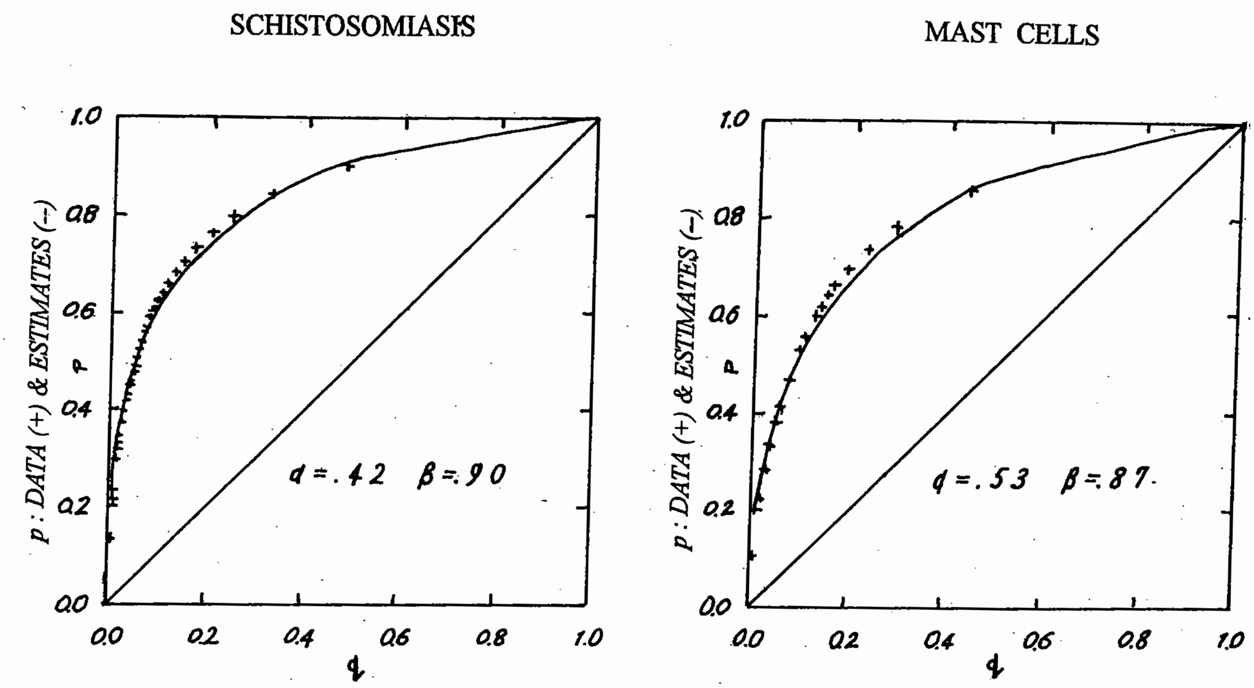

SCHISTOSOMIASIS
MAST CELLS
p : DATA (+) & ESTIMATES (–)
p
q
α = .42 β = .90
α = .53 β = .87
0.0
0.2
0.4
0.6
0.8
1.0

Fig. 3a

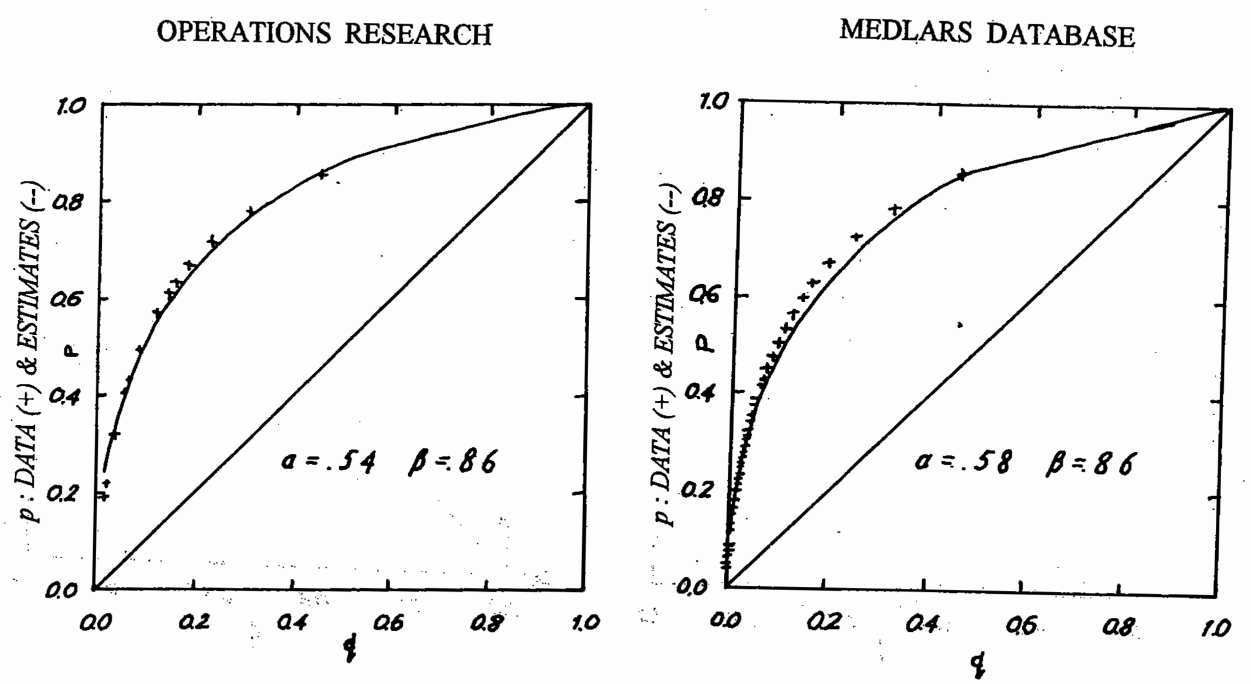

OPERATIONS RESEARCH
MEDLARS DATABASE
p : DATA (+) & ESTIMATES (–)
p
q
a = .54 β = .86
a = .58 β = .86
0.0
0.2
0.4
0.6
0.8
1.0

Fig. 3b

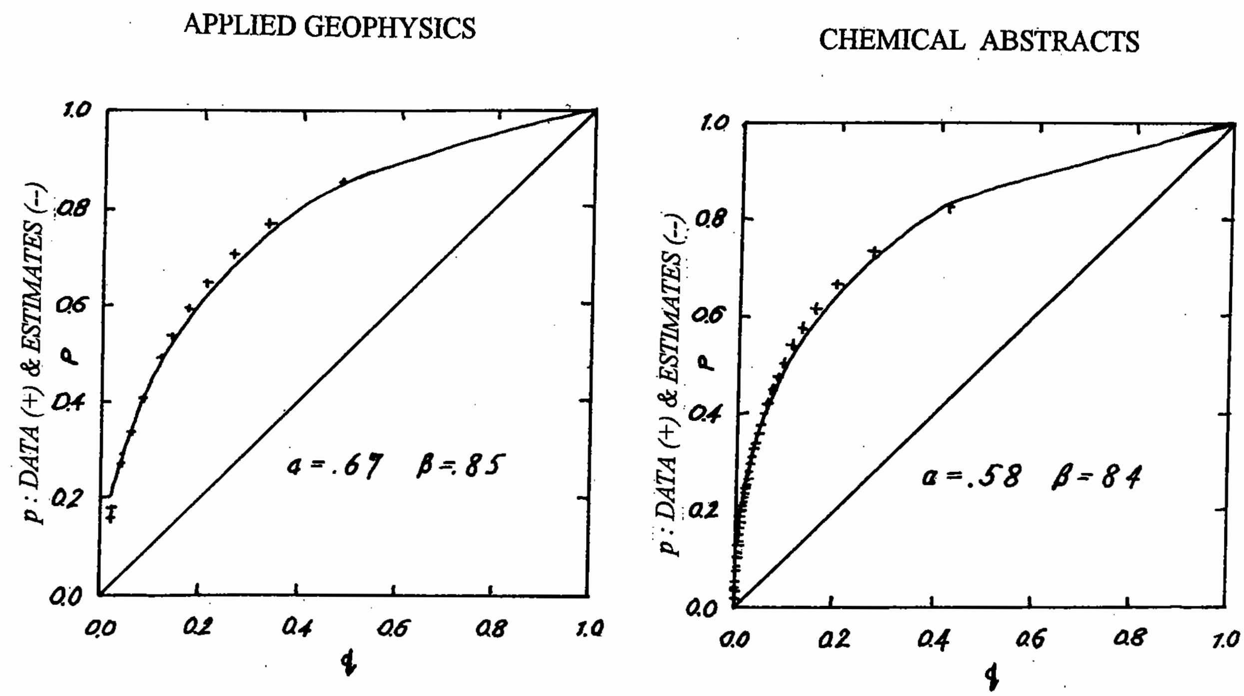

APPLIED GEOPHYSICS
p : DATA (+) & ESTIMATES (--)
q
a = .67 β = .85
CHEMICAL ABSTRACTS
p : DATA (+) & ESTIMATES (--)
q
a = .58 β = 84

Fig. 3b

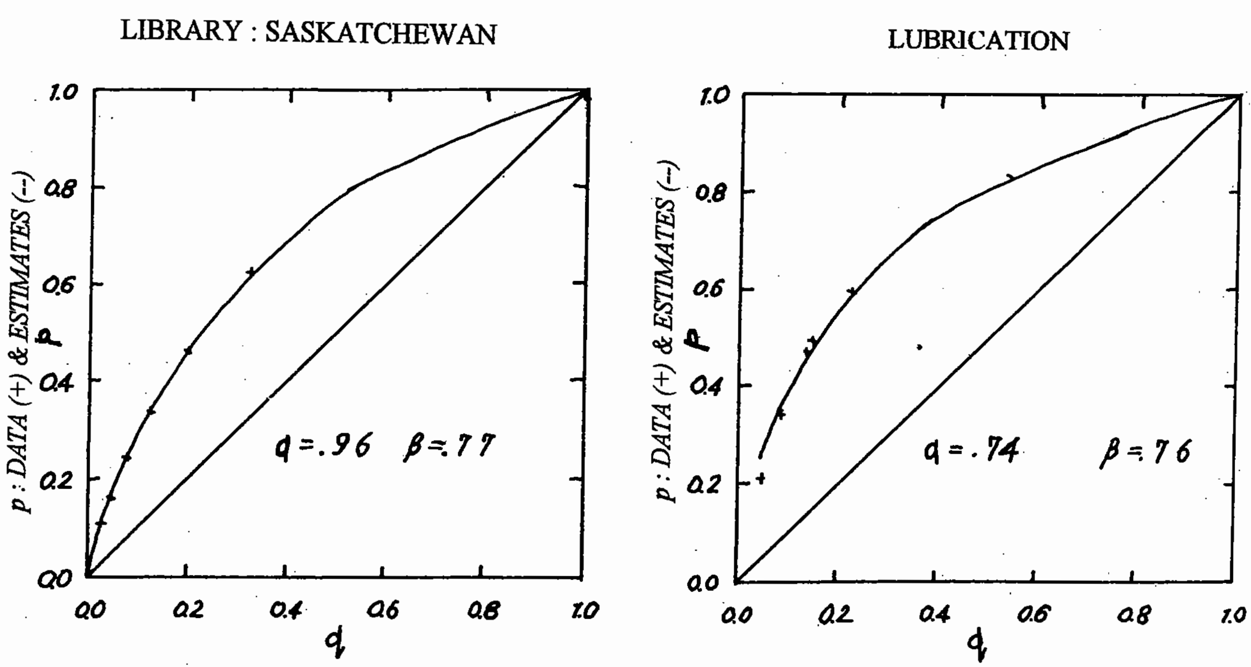

LIBRARY : SASKATCHEWAN
p : DATA (+) & ESTIMATES (--)
q
a = .96 β = .77
LUBRICATION
p : DATA (+) & ESTIMATES (--)
q
a = .74 β = .76

Fig. 3c

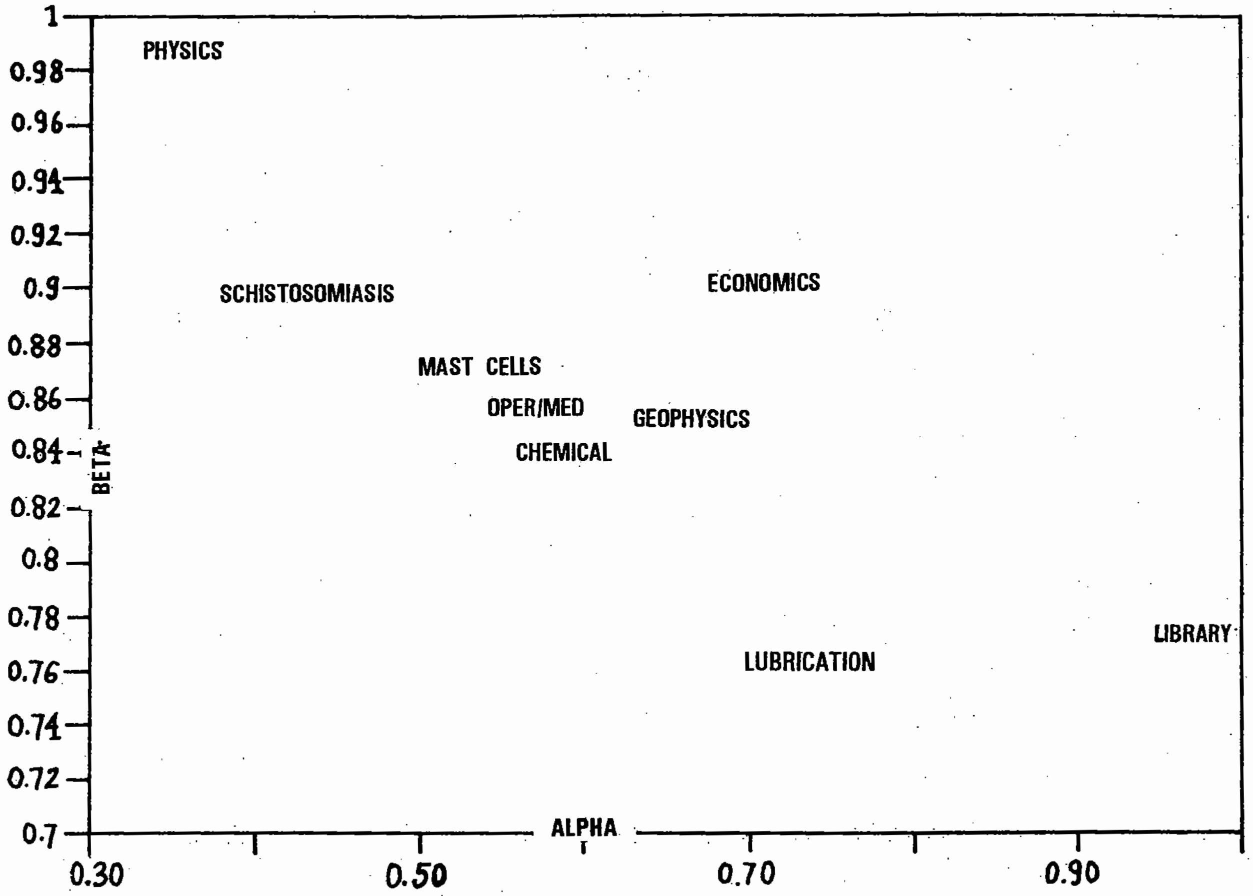


Fig. 4 : Data Position in Alpha-Beta Space

Information Measures in the Electronic Library

Yves F. Le Coadic
CNAM-ICST, 2 RUE Cont, 75141 Paris Cedex 03, France

Thierry Lafouge
E.N.S.S.I.B, 17/21 Bd du 11 novembre, 69623 Villeurbanne Cedex, France.

The introduction of measurement in libraries and information and documentation centres does not appear to be a common practice. With the present important use of electronic and photonic information techniques and the general digitization of information flows, it is possible to develop a measurement system based on well-defined quantities with proper measurement units.
After defining the electronic library and presenting the present state of technological developments in this sector, we try to define a measurement system adapted to this particular information system i.e. fundamental properties, principal and derived units and following the mathematical measurement theory.
Then, we use some of these measures in two practical applications namely a documentation center and an electronic hypertext system. These results can be obtained after having done a mathematical modelization of the corresponding informational spaces.

1. Introduction

All type of information, whether it be in the form of a text, sound, image or data can be processed and coded by the same procedures, stored on the same media, on a unique quantitative scale, in bits. It is clear that this evolutionary tendency is unavoidable and that by the end of the century digital technologies will have partly replaced the other technologies, and what is now a notion, the electronic library [3] will become reality. From his/her workstation, the user will thus be able to browse online and to transfer information sources on support of the same type.

Informetrics, which aims at developing tools for the management of information flows, will become a necessity. It will be of primary importance to develop measurement methods to control this flow. In electronic information, signal, text, time, author-reader populations, utility are measurable properties. In this article, using some of these properties, we attempt to define a simple mathematical model of the electronic library. First, we will define the properties and unit applicable to the electronic library and then, using empirical and mathematical approaches, we will propose a model based on mathematical functions, such as the *'use density function.'*

2. The Electronic Library

(a) Information techniques in the electronic library

Electronic or photonic information techniques, whether for transmission or storage, involve magnetic, magneto-optical or optical phenomena and can be distinguished according to the analog or digital coding used.

1) Nowadays, in a **proto-electronic library,** some of these techniques coexist : for example, videodiscs, audiocassette, computer, network, CD-ROM, etc.
2) Tomorrow, in a**semi-electronic library,** the digital techniques will replace the analog techniques. We list some of the techniques which will be used in the future. Some are already in use in certain centers: audiotex, image banks, CD-X, electronic conference, DAT, DVI, EDI, electronic book, laser card, OCR, ISDN, hypermedia systems, videotex, TVHD, etc.

3) At the end of this century, we can predict that the user will obtain information from an **electronic library,** a true information hypermarket which will offer on-line consultation and supply of information from different sources. The users will be able to build their own personalized "reference manual"; and thereby obtain audio and video information, either to view immediately, or to store electronically for use at a later date. The pay-per-view usage will become very common.

(b) The electronic library concept

The expression "electronic library" refers to a set of virtual objects that we call *information carrying objects* (ICO) [3]. The signs which appear in text, image and sound forms are decomposed into discontinuous signals and stored using a unique quantitative scale. These objects can then be structured in various ways into real documents which can take the form of volumes, periodicals, journals, etc......... . After such processing, the distinction is no longer made between a *'document object'* (DO) and an *information carrying object* (ICO) i.e. an document object will be considered as an information carrying object and vice versa[1].

3. Measurement and Measurable Properties

A property can be defined by its essential attributes : an attribute can either increase or decrease whatever human sense is involved. Among all these attributes, we will distinguish between measurable and non measurable attributes. Obviously, only the former can be used in informetrics. A measurable attribute is one which, when compared to an attribute of the same type, can be expressed as a number[2]. So the measurement of an attribute is the determination of the number which expresses the comparison of this attribute with another attribute of the same type, thereby acting as a term of comparison and called a unit. If it is to be measured, a property must satisfy the following two conditions :

— it must comply with the notion of equality
— it must comply with the notion of addition.

The theory of measurement took a long time to develop because humanity was incapable to distinguish clearly between what was to be measured and with what units to measure it. It was only with the appearance of the theory of real numbers and the notion of set, that the present day theory of measurement could be developed. Here are some of the primary definitions and propositions.

(a) Measurement : definitions and propositions

1 - definitions : here are some important definitions pertaining to measurement.

Definition 3.1

Let E be a set which we call **a clan** of parts of E, any subset T of P (E) (set of subsets of E) verifying:

(1) $\varnothing \in T$
(2) If $A \in$ at T its complement $\in T$
(3) If the sequence (A_n) verifies, $A_n \in A \forall n$, the union of $(A_n) \in T$. (E,T) is called a measurable set.

Definition 3.2

For a set E, the largest clan is P(E). We call measurement defined on a measurable space (E,T) any function μ of T in R^+ verifying:

(1) $\mu(\varnothing) = 0$
(2) if (A_n) is a sequence of disjointed elements T, then :

$$\mu(\cup A_n) = \sum \mu(A_n)$$

A measurement μ is finite, if μ verify : $\mu(E) < \infty$.

1. The concept of electronic library takes into account the traditional model of the documentary function where the document specialist, librarian, documentalist, archivist, acts on the document support (document object) for its conversion and on the document content (information carrying object) for its communication.

2 The method of measurement applicable to any property "G" (additive or not) lies in evaluating the intensity "i" of a phenomenon which can easily be determined. "i" is invariably linked to the value of the component G by the following relationship : G = f(i)

Definition 3.3

Let (E_1 T_1) and (E_2 T_2) be two measurable spaces and f a function of E_1 in E_2. We say that f is a measurable function if f verify :

$$A \in T_2 \Rightarrow f^{-1}(A) \in T_1$$

Some examples of measurement :

(1) Let E be a countable space and T= P(E) the application μ of T in R^+ defined by μ(A)= cardinal of A if A is a finite set, infinite, if A is an infinite set, is a measurement called a countable measurement.

(2) Dirac's measurement at any point x of a measurable set, denoted by δ_x is

$$\delta_x(A) = 1 \text{ if } x \in A, \ 0 \text{ if not.}$$

(3) A discrete measurement on (E P(E)) is defined by a set of real positive numbers p_i (the weights) and of a set x_i of elements of E. The measurement on E can then be defined by :

$$A \in P(E) \quad \mu(A) = \Sigma p_i \delta_{xi}(A)$$

2- propositions : two important propositions are the followings :

Proposition 3.1

Let (E_1 T_1) and (E_2 T_2) be two measurable spaces and f a measurable function of E_1 in E_2, if E_1 has a measurement μ, we define a measurement ν on E_2 called an image measurement of μ by f with :

$$\nu(A) = \mu(f^{-1}(A))$$

and we can write: $\int_B V dv = \int_{f^{-1}(B)} V_0 f d\mu$ where V is a function of E_2 in R^+, the integral defined here is to be taken in the sense according to Lebesgue [cf 4].

Proposition 3.2

Given a measurable space containing a measurement, (E μ), and U a function of E in R^+ defined by :

$$m_u(B) = \int_B U d\mu, \qquad \forall_B \in P(E)$$

is a measurement on E : this measurement is defined by the use density function U. The integral defined here is to be taken in the sense according to Lebesgue [cf 4]. If μ is a discrete measurement (cf Example 3.3), m_u can be written as :

$$m_u(B) = \Sigma_{pi} U(x_i)$$

The proofs of these basic propositions are given by ROGER[3a] and RUDIN[3b]. In this article, when speaking of a measurable space for a set E, we will always consider as a clan, the set P(E).

b) Measurable properties

In the field of electronic information, the possible measurable properties are the signal, the text, time, the population and the usability.

The corresponding units are the bit, the word, the second, the author and/or the reader, the uses (lending and/or quotation).

These units are considered as the fundamental units of the unit system for information measurement, from which other units can be derived such as :

- information flow expressed in number of bits per second ,
- use flux expressed in number of uses per unit of time,
- author concentration expressed in number of authors per paper, etc...

In the case of a text, these measurements have no meaning unless the measured entity i.e. the text, contains a unit : article, paragraph of an article, page of an hyperdocument, summary of an article. The measurements which can then be used are :

- the number of words in a text
- the number of key words indexed, in the case of automatic indexing.

4. Mathematical Modelling of the Electronic Library

a) The basic mathematical model

The simplest formalism to translate the electronic library is to represent it by a sequence of triplets :

$$\{(X_i, g_i, X_{i+1}) \quad i = 0,N\}$$

Where X_{i+1} and X_i represent the same structured space in a different manner at a given time t. and g_i is a surjective application of X_i in X_{i+1}.

X_{i+1} will be called the information carrying object (ICO) space and X_i, the document object (DO) space.

The application g_i means that any information carrying object belongs to a single document object and that each document object contains at least one information carrying object at a given time.

Example : Electronic representation of a classical documentation center :

X_4 : volumes space
X_3 : articles space
X_2 : words space
X_1 : characters space
X_0 : bits space

g_3 : any article belongs to a single volume and each volume contains at least one article
g_2 : any word belongs to a single article and each article contains at least one word
g_1 : any character belongs to a single word and each word contains at least one charcter
g_0 : any bit belongs to a single character and each character contains at least one bit.

Two observations can be made concerning this very simple formalism :

- no unique sequence exists, the choice of sequence depends on the object under study
- this formalism supposes that the relationship between X_i and X_{i+1} is functional and not purely relational.

b) Definition of a measurement

In the following, we will consider any given triplet : (X_i, g_i, X_{i+1})

We will designate X_{i+1} by I (ICO) and X_i by D (DO).

The following hypotheses are put forward :

D is countable.

I is a measured space (I, μ) where μ is a finite measurement (definition 3.2)

g is a surjective application

$\mu(g^{-1}(d)) \neq 0 \quad \forall d \in D$.

It can be deduced that I is a countable union of subsets of D of finite measurement.

1) *Measurement on the ICO*

Of course, this measurement depends on the nature of the space studied. We choice a discrete measurement on (I P(I)). The measurement on I can then be defined by :

$A \in P(I)\mu(A) = \sum p_i \delta_{xi}(A)$ (cf Example 3.3)

The number p_i (the weights) characterizes the element x_i of I. This weight p_i can be calculated from measurement on text, signal, etc... If no particular point of the ICO is chosen, the above formula can be applied to the whole space. The unit of measurement of μ depends on the unit of measurement of the weights.

(2) *Measurement of the use density function on the ICO space.*

Let U be a positive function (the use density function) of I in E (E = states space (in general E = N). U represents the number of uses on a given ICO, over a given period. This function is very often related to information flow [1]. The measurement of the use density function can therefore be defined as :

$m_u(B)=\int_B U d\mu \quad B \in P(I)$ (cf Proposition 3.2)

The use density function measurement so defined depends both on the use density function U and the measurement μ defined on the ICO.

3) *Definition of the use density function on the document space*

On the document space D, we define a measurement v as being the image measurement of μ by g (cf Proposition 3.1) The measurement of a DO is the measurement of the ICOs contained in this DO.

We attempt to calculate the use density function V verifying the equation [a] :

$$[a] \qquad m_u(g^{-1}(A)) = \int_A V dv = m_v(A), \quad A \subset D$$

Defined in this way, m_v represents a use density measurement on D (cf proposition 3.2).

Lemma 4.1

Let f be a measurable and bounded function in measurable set E of finite measure μ in R^+; if E_i is a family of measurable sets in E, then :

$$\int_E f d\mu = \sum \int_{Ei} f d\mu$$

We can see the proof in (4 9a) p 108).

Lemma 4.2

Let f be a measurable and bound function in measurable set E of finite measure μ in R^+, if

$\int_E U d\mu = 0$ then $f = 0$ almost everywhere. We can see the proof in (4 (a) p109).

Theorem 4.1

The application V of D in E :

$V(d) = 1/(\mu(g^{-1}(d))). \int_{g^{-1}(d)} U d\mu$

verifies the equation [a] below and is unique:

[a] $A \subset D \quad \int_{g(A)} U d\mu = \int_A V \, d\nu$

Proof : A being countable we can write,

$\int_A V d\nu = \sum \int_{a_i} V d\nu$ where $a_i \in A$.

So by using (Lemma 4.1), it is sufficient to establish the equation for an element a of A. By definition of the image measurement (cf Proposition 3.1), we can write :

$$\int_a V \, d\nu = \int_{g^{-1}(a)} V_o g \, d\mu$$

Let j be an element of I verifying the equation $g^{-1}(a) = g^{-1}(g(j))$. The application V_{og} being constant on $g^{-1}(g(j))$, we can deduce :

$\int_a V d\nu = \int_{g^{-1}(g(j))} U d\mu$

$= \mu(g^{-1}(g(j)))*V(g(j))$

(by definition of [a] and of V)

$= \int_{g^{-1}(a)} U d\mu$

Now we prove the unicity : let us suppose that two applications exist V_1 and V_2 of D in E verifying the equations :

$\int_{g^{-1}(A)} U d\mu = \int_A V_1 \, d\nu,$

$\int_{g^{-1}(A)} U d\mu = \int_A V_2 \, d\nu$

It can be deduced that

$\int_A V_1 \, d\nu = \int_A V_2 \, d\nu \quad \forall A \subset D.$

We can write the equalities

$\int_{A+} (V_1 - V_2) \, d\nu = 0, \int_{A-} (V_2 - V_1) \, d\nu = 0$, where we have :

$A^+ = \{ d \,/\, V_1 - V_2 \geq 0 \},$

$A^- = \{ d \,/\, V_2 - V_1 > 0 \}.$

Then we can therefore deduce that almost everywhere $V_2 = V_1$ (cf Lemma 4.2).

We have shown that the application V of D in E defined by :

$V(d) = 1/(\mu(g^{-1}(d))).\int_{g^{-1}(d)} U \, d\mu$

verifies the preceding equation and is unique. So we choose the use density function V as a function on the space D (DO).

We can summarize the above results by the following diagram :

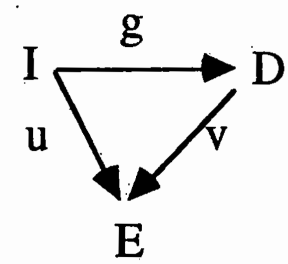


Figure 1 : Diagram of an ICO and a DO

- I : ICO space with a measurement μ
- g: surjective application
- D: DO space with the image measurement ν
- U: use density function of I
- M_u: use density measurement on I
- M_v: use density measurement on D
- V: use density function of D (function of μ, U & g)

Important remark : this diagram is not commutative. However, there is an equality between :

$$\int_I U d\mu = m_u(I),$$

quantity of relevant information carriers in I and $\int_D V d\nu = m_v(D)$, quantity of relevant in D.

5. Practical Applications

1) A classical documentation center

We will again take the example of the previousl mentioned documentation center where:

- I ; articles space
- D : volumes (containing the articles) space
- g : any article belongs to a single volume and each volume contains at least one article.
- U : application representing the number of uses for an article over a given period.

If we chose the countable measurement on I (cf example 3.1), the use density measurement on I is simply defined by the equation :

$$m_u(B) = \sum U(i) \qquad B \in P(I)$$

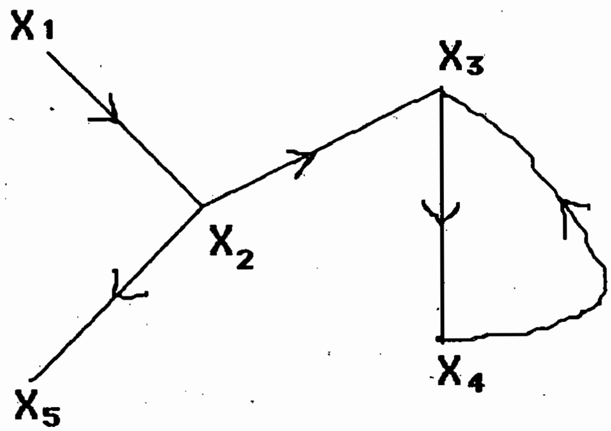


Figure 2 : Diagram of an hyperdocument

$m_u(B)$ is therefor the number of uses over a given period of the set of articles of B. If d designates a volume of k articles, the *'use density function'* V (cf Theorem 4.1) on D is defined by:

$$V(d) = H/k$$

where H is the number of uses of the volume d. V is an average value and fits very well with a "natural" use density function. A probabilistic approach is therefore necessary to build a model when the value of k is not known (2).

2)An hyperdocument :
(a)Definition of ICO and the DO

An hyperdocument consists of two elements :

- a collection of nodes (d_1, d_2, d_3....) which can contain all sorts of information : texts, graphs, images...
- a network of links ((x_1x_2)(x_2 x_5)....) allowing navigation from one node to another.

This definition allows us to write in a more formal fashion that an hyperdocument is a graph constituted of the above two elements. The ICO space I of an hyperdocument can be defined as being the set of pathways. A pathway (I) is a sequence of links I_1 , I_2 , ...I_i..I_q Where for any link l_i (i<q) the initial extremity of l_{i+1} coincides with the terminal extremity of l_i. A point of I is a pathway i.e. a possible route within the hyperdocument. The document space D associated to I will be the union of the supports of the different nodes. The DO associated to an ICO will be represented by the union of nodes belonging to a considered pathway (see figure 2):

I = pathway space
{ (x_1 x_2) (x_2 x_5) (x_2 x_3) (x_3.x_4) (x_4 x_3)
(x_1 x_2 x_5) (x_1 x_2 x_3) (x_2 x_3 x_4)
(x_1 x_2 x_3 x_4 x_5) }

D = document space
g (x_1 x_2 x_5) = d_{125}
(d_{125} is the union of d_1, d_2, d_5)
g (x_4 x_3) = g (x_3 x_4) = d_{34}
(d_{34} is the union of d_3,, d_4)

U = application representing the number of times a user has taken a given pathway of I over a given period.

If we choose the countable measurement on I, the measurement of the use density function on I is simply defined by :

$$m_u\ (L) = \Sigma\ U(I) \qquad L \in P(I)$$

If d designates a DO, the use density function V on D is defined by :

$$V(d) = H/K$$

K is the number of pathways having for DO,d
(if d = d_{34}, k = 2).
H designates the number of uses of pathways having for DO, d.

The countable measurement on I does not take into account the following two parameters

- the length (number of links) of a pathway
- the different nodes which are crossed by the pathway .

We can associate a weight to each pathway, equal to the length of the pathway :

$$p\ (x_1,\ x_2,\ x_3) = 2, \quad p(x_1,\ x_2) = 1...$$

In this way, we obtain the measurement of the use density function

$$m_u\ (L) = \Sigma\ p\ (I)\ U(I), \quad L \in P\ (I)$$

6. Conclusion

The introduction of measurement in libraries and information and documentation centers is possible now, with the important use of electronic and photonic information techniques and the general digitization of information flows.

After defining the electronic library concept and presenting some fundamental definitions and propositions of the theory of measurement, we have tried to define fundamental properties and principal and derived units of the system for information measurement. Then, after having done a mathematical modelization of the electronic library followed by a definition of measurement on informational spaces, namely the information carrying object space and the document object space, we applied the results, for the use density function, in a documentation center and an electronic hypertext system with hyperdocuments. Experiments are on their way now to verify these assertions and to compare the empirical and mathematical approaches.

References

1. Lafouge, T. Problématique de la circulation de l' information - Documentaliste Sciences de l'information - 28(3), 1991, 132-134.
2. Lafouge, T. - Stochastic information Field - 4th International Conference on bibliometrics, informetrics and scientometrics - Berlin, 11-15 September 1993.
3. Coadic, Y.Le. R. Bertrand, S Chambaud - Bibliothèque électronique et infométrie.- Revue de la Société Francaise de Bibliométrie Appliquée - 1991.
4. (a) J. Roger., - Mesure et Intégration - Les presses de l'université du québec (b) Rudin - Real and Complex Analysis - Mc Graw Hill.

Stochastic Information Field

T. Lafouge
E.N.S.S.I.B.-17/21 Bd du 11 novembre 1918, 69623 Villeurbanne Cedex , France.

In this article we propose a mathematical model using the probability formalism in order to represent documental information spaces,) documentation centres, libraries, data banks...) based on observation of the information flows. We will define one of the possible models corresponding to a documentation centre and test these model with real life situation.

1. Introduction

All information (test, image, data) can be processed and coded using the same procedures, stocked on the same media, on a unique quantitative scale, in bits,. It is clear that by the end of the nineties, digital technologies will have partly replaced all other techonologies. In this article, we propose a mathematical model using probability formalism to represent documental information spaces (documentation centre, library, data banks...) based on observation of the information flows. This article follows on from a previous study [6] in which a simple mathematical model of the electronic library was proposed.

2. Mathematical Model : Determinist Information Field

In [6], the following sequence of mathematical objects is defined :

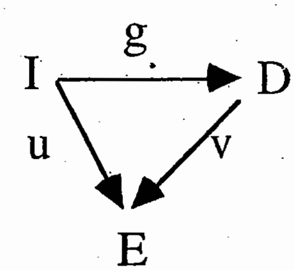
(I D E U V g) which are designated by a determinist information field in which each object represents the following elements of the documental information system.

- **I :** Information Carrying Object space (ICO),
- **D :** Document Object space (DO),
- **g :** surjective application of **I** in **D** which expresses the fact that all information belongs to a DO and that each DO contains at least one article of information,
- **E :** states (in general **E** = N),
- **U and V :** respective functions of **I** in **E** and of **D** in **E** called utility functions : they represent the number of times an ICO (respectively DO) is used over a given period.

In this defintion **U** and **V** are independent. The most natural way to define **V** would be as follows :

$$V(d)=\sum U(i) \text{ where } i \text{ belongs to } g^{-1}(d).$$

V(d) represents the number of ICOs belonging to the document **d** which have been requested.



In [6], this model is used to define the measurement of the utility functions on the spaces I and **D**. The utility function V on **D** is therefore defined according to μ, **U** and **g** where μ designates a given measurement on **I**.

More precisely, **V** is defined so as to give the relationship :

[a] $\int_I U d\mu = \int_D V d\varpi$

where $\varpi = \mu(g^{-1})$: image measurement of μ.

Example : Documentation Centre

We will define one of the possible determinist information fields corresponding to a documentation centre :

I: article space,
D: document space containing these articles,
g: each article belongs to a single document and each document comprises at least one article,
U: application representing the number of requests for an article over a given period.

This determinist information field is equivalent to a three-dimensional IPP ("information production process") defined in [1].

Journals ⟶ Papers ⟶ Occasions when they are borrowed

We demonstrate [6], that the utility function **V** of **D** in **E,** defined by the above equation **[a]** is as follows : **V(d) = H/k,** where H designates the number of requests for articles in the document **d,** and **k** is the number of articles contained in the document. In is often difficult to calculate this function as the values of **k** are not known for each document.

Several reservations can be made concerning this model :

a) It assumes that the function **g** is known, which implies that the structure of the space **I** is perfectly well defined.

b) In general the relationship between **I** and **D** is not of a functional nature but simply relational, for example as in a data bank of a bibliographic type where **I** could designate the key-words space and **D** the set of references.

c) The definition of a measurement on **I** is problematic. For example, if the measurement on the article space is known (cf previous example) we may be led to count objects of a different nature : fundamental articles, summaries, notes. Many precautions must therefore be taken in order to use such a measure.

d) The unused stock of articles called *"No Use"* * in the anglo-saxon school, is not accounted for, which is translated by the fact that **U** is defined on a subset of **I** : of course **U** can be extended to **I,** using the formula **U(i) = 0** in every case where **U** in not defined. Taking into account the *"No use of ICOs"* in this way seems difficult to us for two reasons :

knowledge of *" No use ICOs"* assumes that the space **I** is perfectly well defined, which is not always the case, it is difficult to interpret its significance. It is clear that an ICO may be of no use for totally different reasons : it may be of no use at a given moment in a given context, it may not be found i.e. not retrievable by the user, or it may be unknown to the system itself.

3. Mathematical Model : Stochastic Information Field

The above reservations lead us to define another model using probability formalism. Before defining such a model, we must recall that a stochastic system is a system in which the determinism is not elucidated. It is replaced by a statistical hypothesis on noise, which is generally difficult to justify as in the present case. We adopt the reasoning of the Bayes' school of thought which estimates the probability of everything uncertain, even non random phenomena. A stochastic information field is defined by a sequence of mathematical objects **(I D E p_u p_v p)** where the spaces **I, D** and **E** have the same signification as in the previous model. In this case p_u, p_v, and **p** are probability distributions which represent the following phenomena in relation to the deter-

* The term "No use" is applied to the case when the utility function is zero : a scientific journal which does not contain any article on a given subject for example

minist model :

p_u: probability distribution of the function U,
p_v: probability distribution of the function V,
p: probability distribution of the function V of D in R or N which to all elementd associates its *"counting measure μ"*

by using the preceding notations :

$$\varpi(d) = \mu(g^{-1}(s)).$$

In the stochastic case, knowledge of **g** and μ is replaced by the existence of a distribution **p**. We assume, **I** and **D** being fixed, that the information measure of a document is a quantity, and can be represented by a distribution, which will enable us to calculate p_v according to p_u.

p_u, p_v are what are generally called bibliometric distributions [7].

We assume that these probability distributions are of a discrete nature (this hypothesis is restrictive for the distribution **p** because it assumes that the function **v** only takes discrete values) and also that they are positive values. In [2] L. Egghe carries out a continuous approach of an equivalent problem based on convolution theory.

The case **i = 0** is not considered for the moment. This is very important as taking this case into account often totally changes the adjustment of the distribution [4]. These curves are generally adjusted using standard statistical distributions. The analytical hypotheses of their functions are presupposed. According to the previous considerations made concerning *"No Use"* it would seem misleading to us to consider this case on the prolongation of the same axis of strictly positive frequencies when an adjustment is made. Therefore, no adjustment will be made of p_u or p_v in this case. We also suppose that these distributions allow an expectation. This hypothesis is of importance as such bibliometric distributions are often characterized by the name Zipfian [3] as opposed to Gaussian due to the fact that they do not necessarily have an expectation.

The following notations are adopted :

$p_u(i)$: probability that **U** takes the value **i** : **i = 1,2,3......., ,**
$p_v(i)$: probability that **V** takes the value **i** : **i = 1,2,3......., ,**
p(i): probability that **M** takes the value **i** : **i = 1,2,3......., ,**
$E(p_u)$: expectation of p_u,
$E(p_v)$: expectation of p_v,
E(p): expectation of **p.**

There are no mathematical means to calculate p_v in relation to p_u as there is no functional relationship between **I** and **D**. However, it is easy to understand that in practise p_v depends on p_u and **p.** In this case, the hypothesis is put forward that the DO flow naturally depends on ICO consultations, which are partly conditioned by their *"counting measure μ"*. The following conditions **(h_1)** and **(h_2)** can be expressed by the following equation **(a)** :

$$\text{(a)} \quad E(p_v) = E(p_u).x$$
$$(h_1) \quad x \geq 1$$
$$(h_2) \quad x = 1 \Rightarrow p_u = p_v$$

The conditions **(h_1)** and **(h_2)** signify that the mean flow of a DO is greater than that of an ICO. If the values are identical, this implies that the DO and ICO have the same flow distribution.

x will be calculated for each model, and the conditions **(h_1)** and **(h_2)** verified.

Model A

First of all we assume that **p** is a uniform distribution i.e. that each document has the same *"counting measure"*. Let **j** be this measure. We then define a sequence of discrete distributions P^j by the following equation **(b)** :

$$\text{(b)} \quad P^j(i) = \sum_{i_1 + i_2 + \dots i_j = i} p_u(i_1) \,.\, p_u(i_2) \dots\dots p_u(i_j), \quad i = 1,2,3\dots$$

It is obvious that for any known value of **j : j = 1,2,3 ...** P^j defines a probability distribution whose expectation is defined by the following equation **(c)** :

$$(c) \quad E(P^j) = j.E(p_u).$$

This distribution corresponds simply to the convolution of **j** independent variables having the same probability law [8]. This model corresponds to the case where each DO has the same *"counting measure"*. The model assumes that **p** is a uniform distribution with expectation **j**, (positive integer) which gives the equation $p_v = P^j$. It is clear that this model has little chance of being verified in reality. The equations **(a)** and **(c)** imply that $x = j$, and therefore that **x** is an integer, which of course has no reason for being so.

In the case where $j = 1$ the two distributions p_u and p_v are identical. The conditions (h_1) and (h_2) are verified.

Model B

In this model, the probability distribution **P***(cf Appendix Lemma 1) is defined by weighing the probability P^j using the values P^k to define the probabilty **(d)** :

$$(d) \quad P^*(i) = \sum_k P^k(i)\, p(k) \qquad i = 1,2,3\ldots \quad k = 1,2 \ldots j$$

If **j** is finite, **P*** is the centre of gravity of the probabilites P^k with the values $p(k) : k = 1..j$. The expectation of **P*** (*cf Appendix Lemma 1* and *equation (c)*) is defined by the equation **(e)**:

$$(e) \quad E(P^*) = E(p_u).E(p)$$

This model assumes that the distribution **p** verifies the equation $p_v = P^*$. The equations **(a)** and **(e)** imply the equation $x = E(p)$. **x** is therefore greater of equal to**1**. $X = 1$ signifies that **p** is a uniform distribution where $p(1) = 1$ that is to say, where p_u and p_v are identical. The conditions $(h_1)$and (h_2) are therefore also verified. Model **B** generalizes the preceding one it suffices to define **p** as follows :

$$p(k) = 1 \text{ when } k = j, \text{ o when } k \neq j.$$

Model C

The question of flow is related to *"No Use"* Therefore it only seems natural to develop a model which takes this parameter into account. Models A and B are not able to take into account ICOs which are not consulted and which belong to documents of **D**,documents which are requested (containing ICOs which are consulted). The calculation of p_v is only carried out according to **p** and p_u. Model **C** assumes that p_v depends on **p**, p_u and p_o where $p_o$expresses the proportion of *"No Use"* of the ICOs which belong to the requested DOs. However, if the content of each DO is unknown, as is the case here, in opposition to the determinist information field model, the observation of p_u and p_v is insufficient to identify the requested DOs which contain the ICOs which are not requested. P_o is not an observable value in a stochastic information field as previously defined. However, we will study two cases relative to the observation of P_o.

a) p_o ***is not observable.***

Therefore any value is given to $p_o$and then the distribution probability p_{uo} is defined by the equation **(e)** :

$$p_{uo}(o) = p_o$$
$$p_{uo}(i) = (1 - p_o)\, p_u(i) \quad i = 1, 2, 3 \ldots\ldots\ldots \text{(e)}$$

p_{uo} represents and possible distribution of requests for ICOs contained in the DOs if such an observation were possible. $p_{uo}(i)$ $i = 1,2,3$ ...... may be interpreted as the probability of not being of the type *"No Use"* and of being requested **i** times (the two states being independent). In this case there is total separation between the nature of the two states : being requested, being unrequested.

In the same way as before, the distribution probability P^{*o} is defined by the equation **(f)** :

$$(f) \quad P^{*o}(i) = \sum_j P^{jo}(i)\, p(j) \qquad i = 0,1,2,3 \ldots\ldots$$

where

$$P^{jo}(i) = \sum p_{uo}(i_1).\ p_{uo}(i_2)\ldots p_{uo}(i_j) \quad i = 0,1,2,3\ldots$$
$$i_1 + i_2 + \ldots i_j = i$$

The calculation of $P^{*o}(0)$ gives :

$$P^{*o}(0) = \sum_j (p_o)\ jp(j)$$

However, according to the hypothesis, p_v is not defined for the value **0**, only DOs which are requested at least once are taken into account by the model. This is why the distribution probability P^{**} is defined by the following equation **(g)** :

$$\text{(g)}\quad P^{**}(i) = \frac{P^{*o}(i)}{1-P^{*o}(0)} \qquad i = 1,2,3\ldots\ldots$$

The model assumes that a distribution **p** exists and a number p_o so verifying the following equation $p_v = P^{**}$.

As before, the expectation of P^{**} is expressed by the following equation **(h)** :

$$\text{(h)}\quad E(P^{**}) = \frac{E(p).E(p_{uo})}{1-P^{*o}(0)}$$

The equations **(a)** and **(h)** enable us to establish the following equation **(i)** :

$$\text{(i)}\quad x.E(p_u) = \frac{E(p).E(p_{uo})}{1-P^{*o}(0)}$$

By definition of p_{uo} we have : $E(p_{uo}) = (1-p_0)\ E(p_u)$, which enables us to establish the equation **(j)** :

$$\text{(j)}\quad x = \frac{E(p).(1-p_o)}{1-P^{*o}(0)}$$

We demonstrate *(cf Appendix Lemma 2)* the following equation **(k)** :

$$\text{(k)}\quad \frac{E(p)}{1-P^{*o}(0)} > \frac{1}{1-p_0}$$

(j) and **(k)** allow us to verify the hypotheses **(h_1)** and **(h_2)**.

b) p_o is observable.

This means that p_u is defined for **i = 0,1,3......** Therefore $p_u(o) = p_o$. As before, the distribution P^{**} is defined :

$$P^{**}(i) = \frac{\sum_j P^j(i)\ p(j)}{1-\sum_j (p_o) jp(j)}$$

where we have :

$$P^j(i) = \sum p_u(i_1).\ P_u(i_2)\ldots\ldots p_u(i_j) \quad i = 0,1,2,3\ldots\ldots$$
$$i_1 + i_2 + \ldots i_j = i$$

The model assumes that **p** verifies the equation $p_v = P^{**}$. Calculation of **x** gives the following result :

$$x = \frac{E(p)}{1-\sum (p_o)jp(j)}$$

Therefore, according to **(k)** we have the equation **(h_3)** :

$$(h_3):\qquad x > \frac{1}{1-p_o}$$

If the condition **(h_3)** is verified, this automatically implies that the condition **(h_1)** is verified.

Observations

The case **(a)** does not assume any explicit relationship between **I** and **D** .However, in case **(b)** if p_o is said to be observable and corresponds to the proportion of *"No Use"* of ICOs which belong to the DOs requested, that means in practise that **I** represents the space of ICOs contained in the set of DOs identified by the fact that they have been requested at least once. The distinction between these two cases is very important. In the first case we will speak of an **Open Stochastic Information Field (O.S.I.F)**, in the other of a **Closed Stochastic Information Field (C.S.I.F.)**. In both cases p_o never designates the *"No Use"* stock of I entirely. When $p_o = 0$, we find model **B**.

4. Example

We take as an example the documentation centre defined before. We suppose that that the number of unrequested articles belonging to requested documents is not known.

- **I:** articles space
- **D:** document space
- **U(i):** number of articles requested **i** times : **i = 1,2..n,**
- **V(i):** number of documents requested i times : **i = 1,2..m,**
- $p_u(i)$: probability that an articles be requested **i** items : **i = 1,2..n,**
- $p_v(i)$: probability that a document be requested **i** times : **i = 1,2..m,**
- **p(i):** probability that a document contains i times : **i = 1,2...k,**

where **n** and **m** designate the maximum number of requests for an articles or a document respectively. **k** represents the maximum number of articles contained in a document. We assume of course that the number of requested articles and requested documents is finite; **n** and **m** are known values which can easily be observed, which is not always the case for **k**. It is known that : $1 \leq n \leq m \leq k.n$. It is clear that very often for a type of periodical, the number of articles in a document varies very little and that they are of a different nature. Therefore the signification of the distribution **p** must be extended. It would be of interest to compare the distribution **p** observed with that which "best adjusts the model"* . With the interpretation of **p** given here, the formula $P^*(i) = \sum P^j(i)p(j)$ can easily be interpreted.

The obvious relationships follow :

$$p_u(i) = \frac{U(i)}{\sum U(i)}$$

($\sum U(i)$: number of requested articles)

$$p_v(i) = \frac{V(i)}{\sum V(i)}$$

($\sum V(i)$: number of requested documents)

$$\text{(m)} \qquad D = \sum iU(i) = \sum iV(i)$$

$$\text{(n)} \qquad \sum U(i) \geq \sum V(i)$$

a) Description of data

The data used come from a multidisciplinary documentation centre (INIST of the CNRS)** with a biliographic data bank also available as a publication, PASCAL. User requests for photocopied articles may either be carried out directly on-line by simple interrogation of the data bank, or by post after consultation of the publication or reading of a scientific article quoting certain references.

The following information was available for 1985. Each request is characterized by :

- the **index reference** of the journal
- the **year** of the volume
- the **month** and/or **volume** of the journal;
- the **page** of the article;
- the data of the **request**

This set containing 315 titles of chemical journals was divided into 7 subsets representing the **classification system** according to theme adopted by the establishment. The classification of the journals was carried out by qualified scientific specialists within the documentation centre. The distribution is as follows :

* There may be more than a single solution.
** INIST : Institut National de l'Information Scientifique et Technique du CNRS Nancy.

Fundamental chemistry	171 titles
Chemical engineering	42 titles
Polymers	51 titles
Construction materials	22 titles
Paints and varnishes	10 titles
Wood and paper	6 titles
Perfumes, cosmetics, fats	13 titles
Total	**315 titles**

Let us consider the case of *"Paints and varnishes"*

b) Results

There are several methods for calculation. In [5] we develop hypotheses concerning the nature of the distribution of **p.** In the following table **U** and **V** represent the respective distributions of articles and documents containing these articles in the collection *"paints and varnishes"*

i	U (i)	V (i)	W (i)
i = 1	289	184	194
i = 2	33	31	34
i = 3	9	13	11
i = 4	6	8	5
i = 5		1	3
i = 6		2	0
i = 7		4	1
i = 8		3	
i = 9		1	
i = 10		0	
i = 11		1	
	Σ iU (i) = 406	Σ iV (i) = 406	Σ iW (i) = 337
	Σ U (i) = 337	Σ V (i) = 248	Σ W (i) = 248

We obtain again the previous relation ships **(m)** and **(n)** :

$$\Sigma\, iU(i) = 406 = \Sigma\, iV(i),$$
$$\Sigma\, U(i) = 337 > \Sigma\, V(i) = 248$$

W is a relative distribution to the document in the same way as **V** : we are only interested in the request for an article and not in its frequency; this explains the equality :
$\Sigma\, iW(i) = 337 = \Sigma\, U(i) = 337$ (number of articles requested) : in this case p_u takes two values :

$1-p_o$ an article is requested at least once
p_o an article is not requested

In this case the two models **O.S.I.F** and **C.S.I.F** converge.
The formula **(g)** can be written;

$$P^{*o}(i) = \frac{\Sigma\, Cij(1-p_o)^{j-i}\, (p_o)j.p(j)}{1 - \Sigma\, (p_o)j.p(j)} = Pw(i)$$

where **i = 1,2......j<i** and **Cij** are the number of combinations of **i** elements in a set of **j** elements. An algorithm for calculation gives us the following results :

i	w(i)	P^{*0} (i)*248
i=1	194	194
i=2	34	34
i=3	11	11
i=4	5	6
i=5	3	4
i=6	0	
i=7	1	

où po: = 0.25 and p(1) = 0.75, p(2) = 0.15, p(3) = 0.05, p(4) = 0, p(5) = 0.05

The X_2 calculated is 0.31 with four degrees of freedom.

5. Conclusion

The mathematical model developed in this article does not pretend to be able to represent all documental information spaces (a notion which must be more strictly defined in the introduction). Once the ICO and DO spaces have been identified, it simply allows us to construct tools to observe this space, through observation of the distributions p_u on I, and p_v on **D** i.e. according to utility functions. It is clear in practice that p_u and p_v can be approximated by frequencies. The calculations used in the example give satisfactory results for adjusting **W** according to the formula **(O)**. The calculation of the distribution **p** allows us to classify [5] the journals according to their theoretical or technical nature.

For the time being, the algorithms used for the calculations are rather long.

APPENDIX

Lemma 1

Let **Pj** be a sequence of expectation probabilities **Ej,** which verify the inequality **Ej < k.j, p** a probability of expectation **E,** therefore the sequence defined by :

$P^*(i) = \sum_j P^j(j_i)\, p(j)$ is a probability with expectation $\sum_j p(j)Ej$.

Proof

$$\sum_i P^*(i) = \sum_i \sum_j P^j(i)p(j)$$

$$= \sum_j \left(\sum_i P^j(i) \right) p(j) = \sum_j p(i) = 1$$

$$\sum_i iP^*(i) = \sum_i i \sum_j P^j(i)p(j) = \sum_j \left(\sum_i iP^j(i) \right) p(j).$$

$$= \sum_j Ej \cdot p(j) < k \sum_j jp(j) = k.E$$

Lemma 2

We have the following equation :

$$\frac{E(p)}{1 - \sum (p_o)jp(j)} \geq \frac{1}{1 - p_o}$$

Proof

If **E(p)** is replaced by its value, this equation is equivalent to :

$\sum_i (i - ip_o + (p_o)^i).\ p(i) \geq 1.$

This equation is true, as for any positive integer **i** we have :

$1 \leq i - ip_o + (p_o) \leq i.$

Reference

1. Egghe, L. The duality of informetric system with application to the empirical laws, Journal of Information science, 16, 1990, p 17-27.
2. Egghe. L., Special features of the author-publication relationship and a new explanation of Lotka's law based on convolution theory, to appear in *J.A.S.I.S.*
3. Haitum. S.D. Stationary scientometric distributions Scientometrics 1982, vol 4, n°1, p. 5-25; n°2, p 89-104 n°3, p 181-194.
4. Jain. A.K. A Statistical Study of Book use. - Purdue University Ph. D 1967.
5. Lafouge, T. Utilisation des distributions pour modéliser la circulation de l'information. La veille technologique sous la direction de Hélène Desvals et Henri Dou-Dunod, 1992, p 332-346.
6. LeCoadic, Y. Lafouge, T. Information measures in electronic library, (Fourth International Conference on Bibliometrics, Informetrics and Scientometrics 1993 Berlin).
7. Quoniam, L. Lafouge, T. Les distributions bibliometriques. Revue française de Bibliométrie, N°9, 1991, p 128-138.
8. Saporta, Probabilités Analyse des données et Statistique. Editions Technip chapitre 2.
9. Sichel, H.S. The generalized inverse Gaussian Poisson distribution, Information Processing and Management, Volume 28 N° 1, 1992 p 5-17.

A Dynamic Model of the Growth of Scientific Knowledge*

Karmeshu
School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi-110 067, India

V. P.Jain
Department of Economics, Correspondence Courses, University of Delhi, Delhi-110 007, India

The well-known Kuhn's paradigm of scientific evolution is examined as a self organising system in which the emergence of a new paradigm is treated as a 'mutant'. By treating the emergent paradigm as fluctuational instability, conditions have been worked out to predict trends in the evolution of scientific knowledge.

1. Introduction

The old notion that science is a logical, rational enterprise continually adding to the stockpile of knowledge has been challenged; many now recognize that the evolution of science is punctuated by violent upheavals, (Sterman, 1985). The proponents of the new idea, notably Kuhn (1970) argues that development of scientific knowledge follows an internal logic and a structure of its own.

Science develops in a cyclic manner through scientific revolutions followed by periods of what he terms as 'normal science'. During the period of normal science, the researchers obtain results within the framework of a given paradigm. The basic thrust of such a scientific enquiry is essentially in the nature of 'puzzle solving'. The inability of a given paradigm to solve certain problems and chance discoveries leads to creation of anomalies. As the explanatory power of the paradigm gets weakened, anomalies accumulate leading to a situation of 'crisis'. Efforts are then made to look for solutions outside the framework of the paradigm resulting in the multiplicity of competing hypotheses. Eventually a new paradigm is discovered which has greater explanatory potential and the crisis situation is diffused. Usually a new paradigm emerges with the work of a single person or at the most several persons (Sterman, 1985). If the paradigm proves greater explanatory power, it increasingly draws researchers from other schools or 'networks' associated with particular problem areas (Mulkay, 1975). Gradually the growth rate peters out when the explanatory power of the paradigms reaches saturation level. A new chapter in the history of scientific evolution begins once again to repeat the phases of normal science, crisis and revolution. The four phases of evolution of scientific knowledge is shown in Fig. 1 as adapted from Sterman (1985).

The evolution of scientific knowledge can be studied as a self-organizing complex system (Nicolis and Prigogine 1977; Haken 1977). Complex systems are described by nonlinear interactions among macroscopic variables where fluctuations play a crucial role in bringing about the phenomenon of order through fluctuations. The similarity between Kuhn's visualization of the evolution of science and Prigogine's framework for open systems driven far from equilibrium is particularly striking. One can easily identify the periods of normal science with steady states and the paradigms can be treated as characteristics of these states. The emergent paradigm would then correspond to transition from one steady state to another as fluctuational instability. The emergent paradigm with its expanded ability would lead to the evolution of science which means that the system gets organized at a higher level with lower entropy (Yablonsky, 1988).

A scientific community has its parallel in an ecosystem in so far as it produces structures and exchanges information as an ecosystem produces structures and exchanges biomass (Blackburn 1977). According to Allen (1976) : 'Biological evolution due to selective advantage can be described by just such a mechanism, whereby the equations describing the populations of interacting genotypes in an ecosystem must be unstable to the appearance of a genotype of species if the evolution is to occur. The mean value of each population is described by dynamics involving such concepts as birth rate and death rate, which will be characteristic of competition for vital factors, the reproductive mechanism and the form of the trophic network describing the ecosystem. An individual event, however, such as the spontaneous mutation of a single individual corresponds to a different level of description from this average density dynamics'.

The purpose of the present paper is to offer an analytical model with a view to gaining insight into the dynamics of the growth of scientific knowledge as an evolutionary system. The framework we propose is essentially based on the model advanced by Allen (1976) in the context of evolution of ecosystems where the birth of an emergent new paradigm is treated akin to mutation in biological evolution. The conditions under which the emergent paradigm will survive, co-exist or replace the existing one are examined and the evolutionary dynamics underlined.

2. Formulation of the model

The model is structured as a birth and death process in terms of the variables representing the number of researchers pursuing a given paradigm. When a new paradigm emerges its potentialities are rather hazy even to its creators as initially they are largely untested. There may be even conflicts and differences of opinions regarding the underlying foundations. This renders the initial growth of the paradigm rather slow. But as the new paradigm gains a foothold due to its enhanced explanatory power, a band-wagon like effect ensues and attracts researchers at a fast rate accelerating its growth.. The growth rate ultimately levels off as nearly all members adopt the new framework (see fig. 1). This is the well-known logistic growth pattern (Crane, 1972; Sterman, 1985; Jain and Garg, 1992). The growth equation of a given paradigm in terms of the number of researchers embracing it is given by

$$\frac{dx_1}{dt} = \alpha_1 x_1 (N_1 - x_1) - d_1 x_1 \qquad (1)$$

where x_1 is the number[1] of the adopters pursuing a given paradigm at time t, N_1 is the total number of researchers as potential adopters of a paradigm in a given field. The parameter α_1 is the rate at which researchers adopt a new paradigm and the parameter d_1 represents the rate at which researchers desert the paradigm and abandon it due to the disillusionment. As such α_1 and d_1 correspond to birth and death rates respectively.

Equation(1) has a stable steady state value

$$x_1^* = N_1 - \frac{d_1}{\alpha_1} \qquad (2)$$

at which the explanatory power of paradigm is saturated.

The scientific evolution has its parallel in biological growth behaviour. Allen (1976) has examined the conditions for growth of new species as mutants which prove more promising for survival. The emergence of a new paradigm can be similarly regarded as a 'mutant'. The Darwin's theory of evolution is a classic example of survival of the fittest. Species who fail to meet the hazards of the environment face extinction unless mutate to be transformed into more robust species. Similarly paradigms are threatened and face oblivion if they cannot meet the expectations of science and must be re-

[1] The number of puzzles which can be solved with the help of a given paradigm is assumed to be proportional to the number of researchers who have embraced the paradigm.

placed by new more promising ones.

Following the framework of Allen (1976), the growth equation for the new paradigm which appears as a mutant can be written as :

$$\frac{dx_2}{dt} = \alpha_2 x_2 (N_2 - x_2 - \beta x_1) - d_2 x_2 \qquad (3)$$

where x_2 is the number of researchers, who adopt the new paradigm, which is symptomatic of its explanatory potential and is described by the growth eq (3). The parameter β refers to overlap of 'networks' associated with a field of research between the existing and the new (transformed) paradigm. As observed by Mulkay (1975), 'research community is composed of an increasing number of relatively small scale networks which cut across the formal boundaries dividing science into disciplines and specialities. There is clear evidence that the membership of these networks overlap considerably, that participants continually move from one problem area and its associated network to another and that research networks undergo a continuous process of growth, decline and dissolution'. Accordingly b can assume values $0 \le \beta \le 1$. When $\beta = 0$ it means that the two paradigms do not share common networks at all (no overlap) and $b = 1$ implies identical 'networks' (complete overlap). Any value of β between 0 and 1 signifies partial overlap.

The emergence of new paradigm as a mutant in the course of scientific evolution modifies eq. (1) and the growth equations corresponding to the two paradigms (existing and emergent) can be written as

$$\frac{dx_1}{dt} = \alpha_1 x_1 (N_1 - x_1 - \beta x_2) - d_1 x_1$$

$$\frac{dx_2}{dt} = \alpha_2 x_2 (N_2 - x_2 - \beta x_1) - d_2 x_2 \qquad (4)$$

This set of equations structures the dynamics of the growth of scientific knowledge.

3. Analysis of the Model

Once the existing paradigm reaches a saturation level, doubts regarding its efficacy begin to emerge on the scene. Many researchers instead of puzzle solving with the help of the existing paradigm concentrate on anomalies and begin to think in terms of a new framework. As the confidence in the explanatory power of the paradigm erodes, efforts to look for alternative hypothesis gains momentum (see fig 1). Somewhere in the process a new hypothesis emerges as a mutant and the state existing at that moment can be expressed as

$$x_1^* = N_1 - \frac{d_1}{\alpha_1}, \; x_2^* = 0 \qquad (5)$$

where x^*_1 is the steady state value corresponding to the number of researches in the existing paradigm. At this stage x^*_2 is the number of researchers following the mutant paradigm which is obviously zero.

The mutant paradigm comes as a structural fluctuation which will spread only when one root of the characteristic equation has a positive real part. The condition for a root of the characteristic equation (Nicolis and Prigogine, 1977) to have a positive real part is given by

$$N_2 - \frac{d_2}{\alpha_2} > \beta\left(N_1 - \frac{d_1}{\alpha_1}\right) \qquad (6)$$

which leads to the emergence and evolution of the new paradigm. The new paradigm will then grow and would gain a foothold in the sense that it would come to occupy a 'niche' within the scientific community working in the field.

The interesting question then emerges is what happens to x_1 i.e. whether the existing paradigm survives. Various situations can emerge depending on the value of the parameter β.

Case 1 : $\beta = 1$: If the emergent paradigm comes to occupy the same niche in the scientific community as the existing paradigm then it would mean that $\beta = 1$. The new transformed paradigm will grow if

$$N_2 - \frac{d_2}{\alpha_2} > N_1 - \frac{d_1}{\alpha_1} \qquad (7)$$

Once the paradigm starts growing it will eventually replace the existing paradigm which would mean that all the researchers in a field shift to the new paradigm and the system will acquire steady state given by

$$x_1^* = 0,\ x_2^* = N_2 - \frac{d_2}{\alpha_2} \qquad (8)$$

This would correspond to the extinction of the existing paradigm.

It is pertinent to point out that successive transformation of paradigms will succeed and replace the preceding ones if their explanatory power increases. This can be realized when for the new transformed paradigm N - (d/α) is larger than the preceding ones, otherwise the emergent paradigm will get extinct. Accordingly the evolution of the scientific knowledge characterized by emergence of successive paradigms can be depicted as in Fig. 2.

Case $\beta \neq 1$: Membership of research networks may overlap but need not be identical. Depending on the degree of overlap, the parameter β can assume the value between 0 and 1. We wish to examine in this section the conditions for the success of the new transformed (emergent) paradigm when there is only partial overlap.

Following Allen (1976), it can be easily shown using linear stability analysis, that given the inequalities

$$N_2 - \frac{d_2}{\alpha_2} > \beta\left(N_1 - \frac{d_1}{d_1}\right)$$

and

$$\beta\left(N_2 - \frac{d_2}{\alpha_2}\right) > \left(N_1 - \frac{d_1}{\alpha_1}\right), \qquad (9)$$

the emergent paradigm will still completely replace the existing paradigm. In such an eventuality the final number of researchers pursuing the emergent paradigm would be greater than initial number of researchers following the existing paradigm, implying an evolutionary growth of scientific knowledge.

The second case arises when the following conditions obtain, viz.,

$$N_2 - \frac{d_2}{\alpha_2} > \beta\left(N_1 - \frac{d_1}{d_1}\right)$$

and

$$\beta\left(N_2 - \frac{d_2}{\alpha_2}\right) < \left(N_1 - \frac{d_1}{\alpha_1}\right) \qquad (10)$$

Here the emergent paradigm will grow but will not acquire all embracing explanatory power to replace the existing paradigm and therefore will coexist with it. The final outcome will be depicted by :

$$x_1^* = \left[N_1 - \frac{d_1}{\alpha_1} - \beta\left(N_2 - \frac{d_2}{\alpha_2}\right)\right] \Big/ \left(1-\beta^2\right)$$

$$x_2^* = \left[N_2 - \frac{d_2}{\alpha_2} - \beta\left(N_1 - \frac{d_1}{\alpha_1}\right)\right] \Big/ \left(1-\beta^2\right) \qquad (11)$$

In the extreme case of $\beta = 0$ (no overlap), x_2 will grow if

$$N_2 - \frac{d_2}{\alpha_2} > 0 \qquad (12)$$

and will attain steady state value

$$x_2^* = N_2 - \frac{d_2}{\alpha_2}$$

In this case the new paradigm will grow independently of the existing paradigm and will coexist to serve diffferent research area networks.

4. Replacement Dynamics

When one paradigm replaces another paradigm several patterms emerge. Fisher and Pry (1971) have examined such replacement patterns in the contest of technology replacement. They find what when f (t) corresponds to the market share of new technology and [1-f(t)] that of the old technology, then the replacement of the old technology by the new follows the equation.

$$ln\frac{f(t)}{1-f(t)} = ln\frac{f_0}{1-f_0} + at \qquad (13)$$

where $f_o = f(0)$ and a (>0) is the constant replacement parameter.

Recently Karmeshu, Bhargava and Jain (1985) provided rationale for this replacement hypotheses based on the model by Batten (1982) in the context of industrial evolution. The model

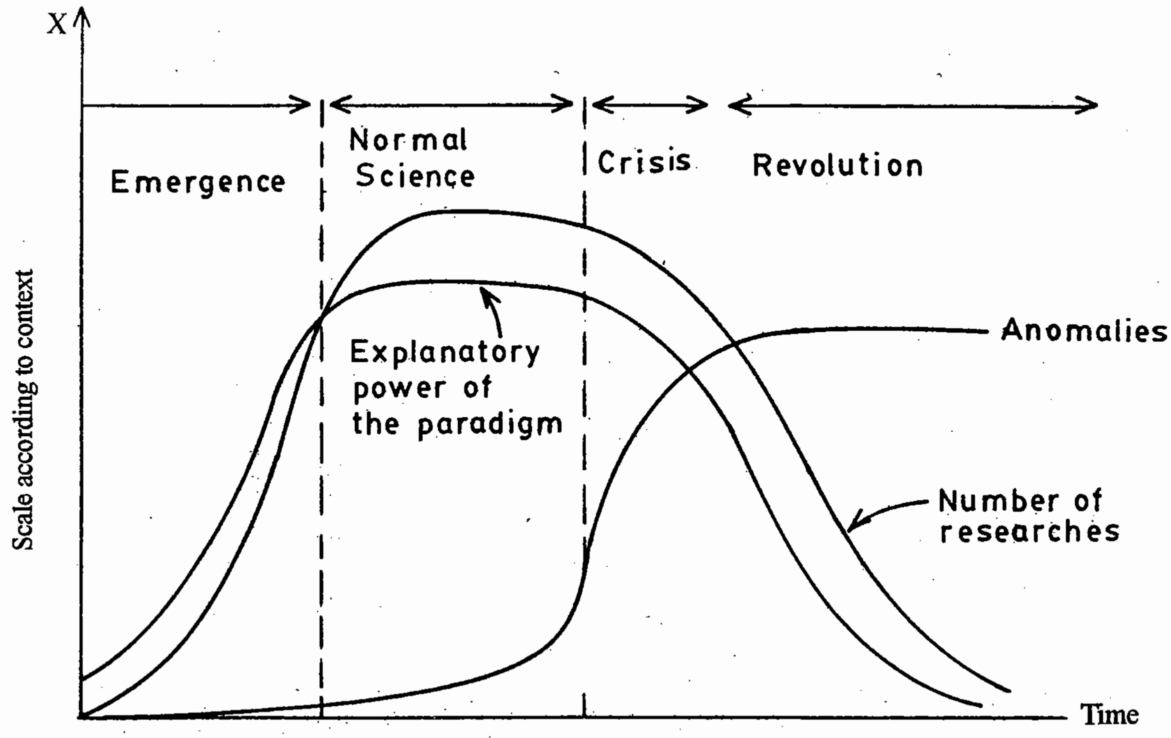


Fig. 1 : Dynamic evolution of a paradigm

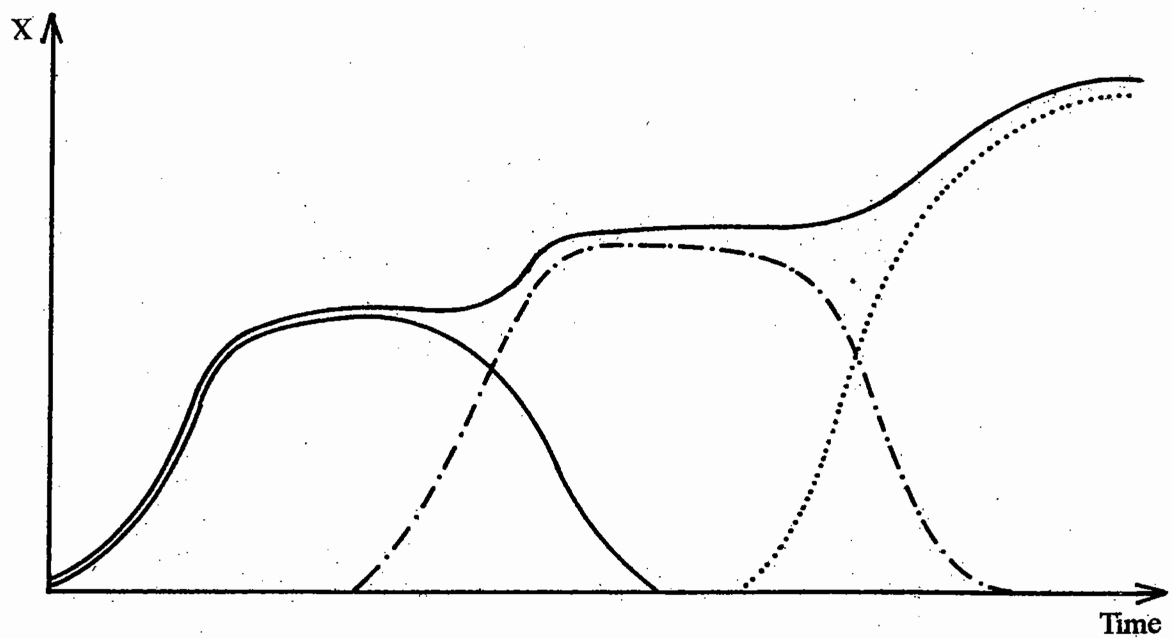


Fig. 2 : Emergence of successive paradigms during the evolution of scientific knowledge

of replacement of one paradigm by another examined in this paper also follows similar patterns under certain conditions.

Writing $\phi = x_2/x_1$ and $f = x_2/(x_1+x_2)$ we find

$$\frac{\dot{\phi}}{\phi} = \frac{\dot{x}_2}{x_2} - \frac{\dot{x}_1}{x_1} \tag{14a}$$

and

$$f(t) = \frac{\phi(t)}{1+\phi(t)} \tag{14b}$$

Assuming $\alpha_1 = \alpha_2 = \alpha$ and $\beta = 1$, we find on using eq. (4)

$$\frac{\dot{\phi}}{\phi} = \alpha\left[\left(N_2 - \frac{d_2}{\alpha}\right) - \left(N_1 - \frac{d_1}{\alpha}\right)\right] = a$$

The rhs corresponds to the constant replacement parameter $a > 0$ and eq (15) reduces to eq (13). This yields the well known Fisher-Pry law. It may be worthwhile to point out that other replacement patterns observed in the context of technology replacement can also be obtained here in case of paradigm replacement when $\alpha_1 \neq \alpha_2$ (Linstone and Sahal, 1976).

5. Concluding Remarks

The model due to Yablonsky (1988) offers a rationale for Kuhn's paradigm in terms of competing biological species. However, the equations describing such a competition are statistical averaged equations. These equations fail to capture 'mutations'. To quote Allen and Lesser (1993) : 'The consequences of viewing evolution as a dialogue between average processes and non average details has led to the concept of evolutionary drive which shows that evolution selects for population with the ability to learn rather than for populations exhibiting alleged optional behaviour'. In this paper we have broadened the scope of enquiry by focusing on the evolutionary phenomenon of scientific growth and associated dynamics.

The 'replacement dynamics' of the new paradigm replacing the existing one has also been shown to follow, under certain conditions, the well-known Fisher-Pry law in technology replacement dynamics. It would be of interest to empirically observe the patterns of paradigm replacement. It would also be of interest to take loss parameters d_1 amd d_2 as state-dependent to capture their varying functional relationship with the changing explanatory potential of the paradigm. Both these issues can form part of future enquiry.

Acknowledgement :

The authors wish to thank Dr. A.Jain and Dr. S. Kaicker for useful discussions.

Reference

1. Allen, P.M. (1976) 'Evolution, population dynamics and stability', *Proc. National Acad. Sci., USA* 73, 665.
2. Allen, P.M. and Lesser, M. J (1993) 'Evolution : Cognition, ignorance and selection' in The Evolution of Cognitive Maps, E. Lazlo, I. Masulli, R Artigiani and V. Csanyi (editors), Gordon and Breach, New York.
3. Batten D. (1982) 'On the industrial evolution', *Reg. Sci. Urban Eco.*, 12, 449
4. Blackburn, T.R. (1973) 'Information and the ecology of scholars', *Science*, 181, 1141
5. Crane, D. (1969)'Social structure in a group of scientists : a test of the "invisible college" hypothesis', *Am. Social Rev.*, 34, 335.
6. Fisher, J.C. and Pry, R.H. (1971) A simple substitution model *Technology Forecast Soc. Change* 3, 75.
7. Haken, H (1977) Synergetics, Springer, Berlin.
8. Jain, A and Garg, K.C. (1992) 'Laser Research in India : Scientometric Study and Model Projection', *Scientometrics* 23, 395.
9. Karmeshu, Bhargava, S.C. and Jain, V.P. (1985), 'A rationale for law of technological substitution', *Reg. Sci. Urban Eco.* 15, 137.
10. Kuhn, T.S. (1970) The structure of scientific revolutions (2nd ed.), Univ. of Chicago Press, Chicago.
11. Linstone, H.A. and Sahal, D. (1976) Technological Substitution, Elsevier, New York Mulkay, M.J. (1975) ' Three models of scientific development' *Socialog. Rev.* 23, 509.
12. Nicolis, G. and Prigogine, I. (1977) Self-organization in non-equilibrium systems, Wiley, New York.
13. Sterman, J.D (1985) 'The growth of knowledge : Testing a theory of scientific revolutions with a formal model', *Tech. Forecast. Soc. Change.* 28, 93
14. Yablonsky A.I. (1988) The development of science as an open system in Systems Research, translated E.L. Nappelbaum Yu A. Yaroshevskii, D. G. Zaydin, Pergamon Press, Oxford.

Time Dependencies in Bibliometric Distributions : A Comparative Empirical Study of Two Research Specialities

H. - J. Czerwon
KAI e.v., Quantitative Science Studies, Schiffbauerdamm 19,10117, Berlin Germany

In this study the stability of bibliometric distributions for two research front specialities in physics (quantum Hall effect) and computer simulations (molecular dynamics method), respectively, is tested under time interval change. Especially we consider the cumulative distributions from year to year in both cases within a space of seven years. It turns out that the distributions of publications in journals are characterized by an increasing data concentration (success-breeds-success phenomenon). This fact is combined with the appearance of S-shaped curves in Bradford diagrams. According to that the slopes of approximated straight lines in rank-frequency distributions (Zipf-Pareto form) become steeper. Beyond it the applicability of the 80/20 rule is discussed in detail. By contrast the cumulative distributions of individual productivity (Lotka's law) are essentially time-independent (scaling). However the annual Lotka distributions are not Levy stable.

Introduction

Since the pioneering works done by Lotka [33] and Bradford [5] bibliometric distribution functions are subjects of a vast number of detailed scientific studies. In this connection especially the influencing factors which determine the shape of these distributions have played a central role. Concerning the distribution of scientific papers on a specific topic to periodicals, many problems have been intensively discussed such as the effect of a possible incompleteness and/or the size of a bibliography as well as the subject field under consideration (e.g. interdisciplinarity of a field) on the shape of Bradford curves. Besides a large number of empirical studies on this subject matter (e.g. [7,12,17.24-26,29,38,39,41] to quote in this context there are also many theoretical papers (e.g. [1,6,8,10,11, 18,20,34;37,42,44]).

Specific problems exist in the investigation of distributions of individual scientific productivity (Lotka distribution) as well : In which way is to count the number of authorships for multiple authorized papers, and are the most prolific authors on the well-known log-log scale for representation of Lotka's law to ignore ? The latter problem is closely related to the effect of finite research lifetime on author productivity.

Surprisingly apart from a few exceptions only little attention has been paid to a possible time dependence of bibliometric distributions. On the one hand this omission is caused by the fact that empirical studies take as a starting point completed bibliographies which can be considered per se as time independent. On the other hand theoretical approaches have given the impression that bibliometric regularities are always stationary, i.e. they do not contain explicitly the time. As a result it was certainly without intention suggested that investigations of time dependence therefore would be unnecessary. At this point one must refer to papers of Haitun, Yablonsky and Bookstein [2-4,27,47]. In contrast to such theoretical approaches Kunz gave in his controversy with Haitun [28] on the basis of a simplified Lotka distribution a plausible explanation that negative power distributions are time dependent: "If a long observation period is splitted into its parts, we get, if events are distributed evenly in time, distributions with shorter spans and steeper slopes" [30]. Also in case of Lotka's law M.L. Pao found "the most surprising fact that the data compiled from a short period, of a single year, had larger values of n" [40] (n stands here for the Lotka exponent, H.J.C.).

Beyond all theoretical considerations it would seem advisable as a result of more recent findings to Bradford's law by Oluic-Vukovic [38,39] to extend the empirical basis of our knowledge of bibliometric distributions in order to come to a better understanding of the time factor in bibliometric regularities. Whereas Oluic-Vukovic analysed in her paper [39] the research output in chemistry and physics of authors from Croatia for a ten- year period with regard to its journal distribution, it is the primary aim of the present study to analyse the possible time dependence of bibliometric regularities (distributions of journals and individual research performance) in the light of evolution of two different research front specialties. The research specialties under consideration are the quantum Hall effect (QHE) on the one hand and the interdisciplinary field of applications of molecular dynamics (MD) simulations on the other hand.

Starting from the seminal paper by K.von Klitzing et al. in 1980 for which he was awarded the Nobel Prize for physics in 1985 we analyse the time development of the publication activity in the research specialty of quantized Hall effect for the decade 1980-1989. By contrast we study the specialty of molecular dynamics method for a publication window 1982-1988, i.e. for a period of time more than 25 years after the first molecular dynamics algorithm was developed (cf. [15]). Apart from the different temporal development stages the comparison of both research specialties give the chance to verify (or falsify) assumptions to the influence of content and characteristics of research fields on bibliometric distribution functions. The specialty of quantum Hall effect has well defined boundaries to other research specialties in physics, e.g. in solid state physics. In other words, the QHE bibliography can be characterized as "homogeneous" [12]. Compared with this the MD method is distinguished by manifold uses ranging from the theory of simple liquids to biophysics. In this latter case we are dealing with an interdisciplinary research specialty, and we can for instance expect that the publications are widely scattered among periodicals. The question in what extent the above-mentioned differences of specialties determine really the time dependence of bibliometric distributions is open so far and therefore subject of our study.

Some additional remarks on time dependencies in bibliometric distributions

As mentioned already there are only a few studies with clear indications that the consideration of the time factor influences the shape of bibliometric distributions. In this context we refer to the fundamental paper of Price in which he proposed a stochastic model for bibliometric and other cumulative advantage processes [43]. This model is derived without explicit reference to time as a variable, but involves implicitly the time dependence of such processes: "It is common in bibliometric matters and in many diverse social phenomena, that success seems to breed success. A paper which has been cited many times is more likely to be cited again than one which has been little cited. An author of many papers is more likely to publish again than one who has been less prolific. A journal which has been frequently consulted for some purpose is more likely to be turned to again than one of previously infrequent use. Words become common or remain rare. A millionaire gets extra income faster and easier than a beggar."

It is immediately evident that the formation of a hierarchy in a social system happens in time and that such a process of stratification is quantitatively reflected by characteristic parameters of bibliometric, income and other distributions. However, within the scope of the cumulative advantage concept and other models no attempt has been made to explain the underlying causes of the detailed, time dependent processes by which the "fine structure" of bibliometric distributions establishes itself. What is observed in most cases as a bibliometric regularity " is an equilibrium or stationary state of a dynamic process"[1] i.e. bibliometric laws give synoptic views of consolidated data.

As pointed out by Bookstein [2-4], a variety of phenomena taken from different fields of human activity and exhibiting a variety of forms can be conceptualized as versions of a single regularity, so that one can properly speak of the informetric law and its manifestations. In most cases it is enough to approximate the empirical manifestations of this informetric law by hyperbolic distributions (inverse power laws: Zipf, Pareto, Lotka, Bradford, and others) for both discrete and continuous variables (cf. the excellent survey given by Fairthorne [22]). These hyperbolic distributions belong to a class of formal statistical distributions first studied in detail by Paul Levy [31]. On the other hand Mandelbrot has shown that hyperbolic probability distributions are most closely related to fractals [35], i.e. such distributions are self-similar in that their parts differ from the whole only in scale (scaling invariance). Especially the exponents in inverse power distributions (Pareto, Zipf, Lotka etc.) appear as so-called fractal dimensions.

In this context a possible time dependence of bibliometric distributions would imply obviously the violation of scaling invariance reflected by the variation of the exponents in inverse power distributions. Therefore we will pay our special attention to this problem.

Data sources and methodology

All publications data on the quantum Hall effect for the decade 1980-1989 are based on computer-assisted searches in the Abstracts. Since the controlled term "quantum Hall effect" is not taken into consideration in the INSPEC thesaurus until 1985, the searches for the period of time 1980-1984 were very time-consuming. (Bibliometric indicators for 1980-1985 were published by Czerwon et al. [13].)

For the decade 1980-1989 (actual publication years) the bibliography of the quantized Hall effect contains 1275 documents, among them are 1062 contributions in journals. One can assume that this bibliography reflects the publication activities in the research specialty under discussion almost completely because the INSPEC database collects publication data not only from international key journals but also from national journals, important conference proceedings, monographs etc. During 1980-1989, 1128 scientists published on the quantum Hall effect.

The bibliography of the second, with regard to applications, more interdisciplinary research specialty, the molecular dynamics method, was primarily compiled by on-line searches in INSPEC and complemented by data from Physics Abstracts, Biological Abstracts and the Science Citation Index. It contains 2164 publications for the seven-year period 1982-1988; the huge majority of them (2017 papers) was published in journals. It should be noted that at first in 1982 the term "molecular dynamics method" became a controlled term in the INSPEC thesaurus, i.e., 25 years after the first MD paper by Alder and Wainwright [15]. In the bibliography of MD simulations were altogether collected 1905 authors.

In accordance with the aim of our study we subdivide the publication periods, ten and seven years respectively, into one-year intervals. Then we calculate the one-year and in particular the cumulative distributions of individual productivity on the one hand and the corresponding distributions of publications in periodicals on the other hand. In this context it should be mentioned that in the case of the specialty "quantum Hall effect" at first in the period 1980-1983 there is a sufficiently large number of publications in order to draw statistically significant conclusions. This is the reason why we analyse preferably the possible influence of the time component on the curve shape of cumulative bibliometric distributions for both research specialties successively in seven subperiods: 1980-1983, 1980-1984, ..., 1980-1989 (QHE) and 1982, 1982-1983,..., 1982-1988 (MD method).

For representation of the dispersion of publications over periodicals we have chosen both the well-known cumulative rank-frequency distribution according to Bradford and the rank-frequency distribution exactly in the form used by Zipf to rank words in order of their frequency of occurrence in texts. The distribution of individual publication productivity is described in the size-frequency form used in the standard formulation of Lotka's law.

It must be emphasized that we take into account for both research specialties the complete data sets to fit regression lines on double logarithmic scale for rank-frequency distributions of journals as well as size-frequency distributions of authors' productivity. As we shall see in the following this procedure is appropriate because nevertheless a high correlation of data is ensured. More often the distributions of Lotka and Zipf (or rather of Pareto) describe only the tails of such non-Gaussian distributions, i.e., there is a concentration of highly-productive scientists and journals respectively in long tails of extremely skew distributions characterized by inverse power laws.

Results

Distribution of journals

The best-known presentation in order to rank journals with regard to their productivity on a particular scientific subject is the Bradfordian one. If the journals are arranged in order of decreasing productivity, a long-linear plot of the cumulative total of papers in journals of a certain rank or better versus that journal rank r yields often a characteristic stretched S-shaped curve. In that case for journals of low productivity (high values of r) there is a downward deviation, called the Groos droop [26]. The interpretation of the Groos droop has been a point of controversy: Among other causes the incompleteness of data due to selectivity or omission has been usually quoted, while other consider it to be an integral part of the scatter process.

As can be seen from Figure 1 and Figure 2 the curved shape of the cumulative journal distributions for both research specialities changes over time in the like manner : The larger the data sets the more pronounced is the Groos droop in the upper section of the Bradford curve, i.e. the journals of low productivity lie more and more below a straight line.

In order to describe the journal distribution in another way we choose the rank-frequency form of Pareto and Zipf, respectively:

$$P = \frac{\text{const}}{r^k} \quad (1)$$

where P - number of relevant publications in a journal, r - rank of journal, k - Zipf exponent. We used the linear least square method to fit regression lines on log-log scale. Then const. is equal to the y intercept of the regression line.

The most important parameter in this version of Bradford's law is the Zipf exponent which entirely describes the scatter of articles in journals. Table 1 and Table 2 list the Zipf exponents

Table 1 : Quantum Hall effect

Numerical values of the exponent in cumulative rank-frequency distributions of publications in journals

Time period	Zipf exponent	Correlation coefficient
1980-1983	1.07	-0.971
1980-1984	1.18	-0.984
1980-1985	1.16	-0.986
1980-1986	1.18	-0.987
1980-1987	1.19	-0.988
1980-1988	1.24	-0.986
1980-1989	1.26	-0.989

including the corresponding correlation coefficients for evolving cumulative distribution functions. It is obvious that the numerical value of the Zipf exponent for both research fields increases continuously with

Table 2 : Molecular dynamics method

Numerical values of the exponent in cumulative rank-frequency distributions of publications in journals

Time period	Zipf exponent	Correlation coefficient
1982	0.86	-0.974
1982-1983	1.05	-0.981
1982-1984	1.10	-0.983
1982-1985	1.16	-0.981
1982-1986	1.22	-0.981
1982-1987	1.22	-0.983
1982-1988	1.25	-0.982

elapsed time, i.e. the slope of the regression line becomes steeper. This fact is accompanied by an increasing concentration of papers in periodicals as it is shown in detail below. The cumulative rank-frequency distributions (Pareto-Zipf form) of 1989 (QHE) and 1988 (MD method) are presented in Figure 3 and Figure 4, respectively.

The time dependence of the Zipf exponent means in the language of fractals that the cumulative distribution of papers to journals is not scaling under time transformation.

The 80/20 Rule

There are different formulations of the 80/20 rule [9, 19]. Concerning the scattering of papers in a subject area over journals the classical 80/20 rule states that only 20% of the most productive periodicals in a bibliography are needed to produce 80% of the relevant publications. This regularity is exactly fulfilled only in exceptional circumstances. Therefore it is appropriate to generalize this rule which was done recently by Egghe [21].

In the same theoretical paper Egghe analysed, picking up the thread of Burrell [9], the intuitively clear property that in a bibliography with a higher average items per source μ we need a smaller fraction of the most productive sources in order to have the same fraction of the items yielded by these sources and vice versa. On the assumption that a bibliography satisfies the Lotka distribution (with Lotka exponents >2), Egghe came to the conclusion that the 80/20 rule is exactly valid only if μ is a little bit larger than 7. Table 3 and Table 4 list the results of testing the 80/20 rule for our two research specialities and the seven time intervals under consideration.

Table 3 :Quantum Hall effect : Test of the 80 / 20 rule

Time Period	No. of items in journals	No. of Journals	Average No. of items per journal	% of items in 20% of most produc tive journals
1980-1983	128	35	3.7	67.2
1980-1984	271	52	5.2	72.7
1980-1985	407	71	5.7	74.9
1980-1986	600	96	6.3	77.2
1980-1987	732	107	6.8	77.6
1980-1988	896	115	7.8	78.8
1980-1989	1062	129	8.2	79.5

Table 4: Molucular dynamics method : Test of the 80/20 rule

Time Period	No. of items in journals	No. of Journals	Average No. of items per journal	% of items in 20% of most produc-tive journals
1982	182	67	2.7	59.3
1982-1983	449	99	4.5	67.9
1982-1984	689	128	5.4	71.3
1982-1985	978	154	6.4	72.8
1982-1986	1270	175	7.3	76.1
1982-1987	1631	211	7.7	77.5
1982-1988	2017	242	8.3	79.0

Note that when we observe a subfield over an increasing length of time we find that the average number of items per source is continuously growing combined with a gradual fulfilment of the 80/20 rule. In the present empirical study this rule proves to be correct for μ values a little bit larger than 8. The comparison of numerical values for μ shows that this parameter is very sensitive to variation of the time span. Generally our result confirms Burrell's finding for library loans that the length of the time period observed is of crucial importance to verify the 80/20 rule [9].

Distribution of individual productivity

As in case of the journal distribution we consider both the cumulative individual publication performance for seven time intervals and the annual productivity. In order to measure the contribution of a single scientist to a multiple authorized paper we choose the so-called normal or complete count procedure, i.e. if more than one scientist contributes to the authorship of a single paper, each collaborating author is credited with a full contribution [32]. In this case are, as comparative studies show [14,46], the slope values of regression lines in Lotka diagrams lower than in diagrams which so-called fractional authorships take into account. This statement contradicts Rousseau's finding that fractional counting of authors does not lead to a Lotka distribution [45].

From Table 5 and Table 6 it can be seen that for the specialty of quantized Hall effect the exponent m in Lotka's law,

$$A = \frac{\text{const}}{P^m} \qquad (2)$$

(where A — number of authors and
P — number of publications)

varies between 1.73 and 1.84 for cumulative distributions, while it ranges from 1.99 to 2.74 for annual productivity distributions. This exponent has for MD

Table 5 : Quantum Hall effect

Numerical values of the exponent in Lotka's law (cumulative distributions)

Time period	Qumulative No. of authors	Lotka exponent	Correlation coefficient
1980-1983	157	1.73	-0.993
1980-1984	303	1.74	-0.961
1980-1985	421	1.79	-0.965
1980-1986	629	1.84	-0.971
1980-1987	757	1.82	-0.971
1980-1988	961	1.79	-0.963
1980-1989	1128	1.74	-0.948

Table 6 : Quantum Hall effect

Numerical values of the exponent in Lotka's law 1983—1989

Year	Number of authors	Lotka exponent	Correlation coefficient
1983	132	2.62	-0.977
1984	237	1.99	-0.967
1985	245	2.74	-0.976
1986	382	2.37	-0.961
1987	339	2.44	-0.981
1988	409	2.42	-0.966
1989	405	2.71	-0.949

Table 7 : Molucular dynamics method

Numerical values of the exponent in Lotka's law (cumulative distributions)

Time period	Qumulative No. of authors	Lotka exponent	Correlation coefficient
1982-1983	485	1.89	-0.958
1982-1984	694	1.85	-0.960
1982-1985	923	1.83	-0.970
1982-1986	1195	1.91	-0.972
1982-1987	1520	1.90	-0.967
1982-1988	1905	1.87	-0.970

Table 8 : Molecular dynamics method

Numerical values of the exponent in Lotha's law 1982-1988

Year	Number of authors	Lotka exponent	Correlation coefficient
1982	268	2.28	-0.955
1983	352	1.93	-0.969
1984	393	2.37	-0.988
1985	427	2.72	-0.986
1986	537	2.55	-0.978
1987	628	2.60	-0.949
1988	725	2.46	-0.978

simulations a similar tendency: It varies between 1.83 and 1.91 for cumulative productivity distributions and between 1.93 and 2.72 for annual Lotka distributions (cf. Table 7 and Table 8). The cumulative distributions for 1989 (QHE) and 1988 (MD method) are illustrated in Figure 5 and Figure 6, respectively. It is noticeable that the time fluctuations of the Lotka exponent for both research fields are statistically irregular, a concentration effect in cumulative distributions cannot be observed. However, there is such a phenomenon if we compare one-year distributions with cumulative ones (cf. [46]). The Lotka exponents of cumulative distributions are relatively time independent, whereas the sum of two one-year distributions yields a distribution with a lower slope of the regression line. As mentioned above, this modification of Lotka distribution is caused by a short-time concentration of papers by highly productive scientists.

It should be noted that the property that the sum of two Lotka distributions does not yield in general a self-similar distribution is called violation of the Levy stability [31,35,36].

Discussion and Summary

The result of a computer simulation study by Qiu and Tague [44] indicated that the hypothesis that the Groos droop is primarily caused by incomplete data sets must be rejected. This finding is on the whole confirmed by the results of the present paper. Another attempt to give an explanation of the Groos droop by Egghe and Rousseau is based on the observation that a merging of pure Bradfordian bibliographies, i.e. such bibliographies the Bradford curves of which have for high values of r approximately the form of a straight line, could also yield bibliography with a Groos droop [20]. Egghe and Rousseau concluded from this fact that a Groos droop can always be expected in interdisciplinary bibliographies.

Compared with this our results show that bibliographies of the exactly defined specialty of quantized Hall effect as well as the multidisciplinary specialty of applications of molecular dynamics simulations tend to exhibit for ever-increasing collections of data more and more a characteristic stretched S-shaped Bradford curve. Apart from that Rao's assumption (cited by Egghe [20]) cannot be corroborated that a Groos droop can be found in case that a subject is new and the scientific articles in this early state are scattered over a lot of journals which are not directly devoted to the new subject. As can be seen from the results for the rapidly expanding subject field of Quantum Hall effect, from the start a continuously increasing concentration of publications in journals is observed (success-breeds-success phenomenon).

In this context one of the hypotheses formulated by Oluic-Vukovic [39], that the appearance of the S-shaped Bradford curve is always associated with a higher data concentration, was verified on the basis of the generalized 80/20 rule. This fact is also reflected by increasing values of the slope parameter (k-exponent) in the Pareto-Zipf power law version of journal distributions, i.e. the corresponding hyperbolic curves are not semi-similar. Generally it seems to be that the Groos droop does not result from incomplete data gathering, but is influenced first of all by intrinsic probabilistic processes [17].

Concerning the evolving distributions of individual publication productivity we ascertained that the cumulative distributions for both specialties are essentially time independent, excepting we consider sums of distributions for shorter time periods (one year or so). In this case we observe a concentration of articles by more prolific authors ,too.

In summary it may be said that the dynamics of the stratification of journals with regard to their publication productivity on a subject field is comparatively much more obvious than changes in the hierarchical structure of individual performance. Because of this finding we have also to modify some of our earlier conclusions [16]. On the other hand we cannot confirm former results that the shape of bibliometric distribution functions is fundamentally determined by the subject matter under consideration or other extrinsic characteristics of data sets. Rather it looks as if the shape parameters of such distributions are more and more similar for sufficiently large bibliographic data collections, i.e. the distributions exhibit fractal properties.

It will be the aim of forthcoming research on the time factor in bibliometric distribution functions to test the validity and generality of corresponding regularities for emerging scientific fields and especially to ascertain the time periods in which variations of characteristic parameters in bibliometric regularities play a significant role. In this context it is indispensable to use statistical tests in order to fit theoretical models to empirical data sets. In particular, there is a need to include refined theoretical models in our considerations.

Note added in proof:

After completion of this paper a study by K.C. Garg et al. [23] was published which reveals that a journal distribution curve similar to Bradford's curve is obtained when emerging research fields mature.

Acknowledgement

The author would like to thank Matthias Winterhager (Institute for Science and Technology Studies, University of Bielefeld) for the partial preparation of the bibliography of molecular dynamics simulations.

Reference

1. Basu, A. Hierarchical distributions and Bradford's law, Journal of the American Society for Information Science, 43 (1992) 494-500.
2. Bookstein, A. The bibliometric distributions, Library Quarterly, 46(1976) 416-423.
3. Bookstein, A. Patterns of scientific productivity and social change: A discussion of Lotka's law and bibliometric symmetry, Journal of the American Society for Information Science, 28(1977) 206-210.
4. Bookstein, A. Informetric distributions, Part I : Unified Overview, Journal of the American Society for Information Science, 41(1990) 368-375. Part II : Resilience to ambiguity, Journal of the American Society for Information Science, 41(1990) 376-386.
5. Bradford, S.C. Sources of information on specific subjects, Engineering, 137(1934) 85-86.
6. Brookes, B.C. The derivation and application of the Bradford-Zipf distribution, Journal of Documentation, 24(1968) 247-265.
7. Brookes, B. C. Bradford's law and the bibliography of science, Nature, 224(1969) 953-956.
8. Brookes, B.C. Theory of the Bradford law, Journal of Documentation, 33(1977) 180-209.

9. Burrell, Q.L. The 80/20 rule: Library lore or statistical law ? Journal of Documentation, 41(1985) 24-39.
10. Burrell, Q.L. Modelling the Bradford phenomenon, Journal of Documentation, 44(1988) 1-18.
11. Chen, Y. S.and F.F. Leimkuhler, Bradford's law : An index approach, Scientometrics, 11(1987) 183-198.
12. Coleman, S.R. Bradford distributions of social-science bibliographies varying in definitional homogeneity, Scientometrics, 27(1993) 75-991.
13. Czerwon, H.-J., J. Vlachy, Quantized Hall effect. Publication and citation follow-ups of a Nobel prize discovery, Czechoslovak Journal of Physics B, 36(1986) 1101-1106.
14. Czerwon, H.-J. Scientometric indicators for a specialty in theoretical high-energy physics : Monte Carlo methods in lattice field theory, Scientometrics, 18(1990) 5-20.
15. Czerwon,H.-J.,G. PEINEL.Molecular dynamics method : Publication and citation patterns of a research specialty of computer simulations,Information Quarterly for Compu ter Simulation of Condensed Phases, No.32 (1990) 51-67.
16. Czerwon, H.-J. Bibliometric distributions and the evolution of research front specialties in basic research (in German), In: W. NEUBAUER, K.-H. MEIER (Eds), Proceedings of the Deutscher Dokumentartag 1992, September 22-25, 1992, Berlin, Deutsche Gesellschaft für Dokumentation, Frankfurt a.M., 1993, pp. 625-639.
17. Drott, M. C. and B. C. Griffith, An empirical examination of Bradford's law and the scattering of scientific literature, Journal of the American Society for Information Science, 29(1978) 238-246.
18. Egghe, L. Consequences of Lotka's law for the law of Bradford, Journal of Documentation, 41(1985) 173-189.
19. Egghe, L. on the 80/20 rule, Scientometrics, 10(1986) 55-68.
20. Egghe, L.and R. Rousseau, Reflection on a deflection : A note on different causes of the Groos droop, Scientometrics, 14(1988) 493-511.
21. Egghe, L. Exact probabilistic and mathematical proofs of the relation between the mean μ and the generalized 80/20- rule, Journal of the American Society for Information Science, 44(1993) 369-375.
22. Fairthrone, R.A. Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction, Journal of Documentation, 25(1969) 319-343.
23. Garg, K.C., P. Sharma,and L. Sharma, Bradford's law in relation to the evaluation of a field. A case study of solar power research, Scientometrics, 27(1993) 145-156.
24. Goffman, W., K.S. Warren, Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures, Nature, 221(1969) 1205-1207.
25. Goffman, W.and T.G. Morris, Bradford's law and library acquisitions, Nature, 226(1970) 922-923.
26. Groos, O.V. Bradford's law and the Keenan-Atherton data, American Documentation, 18(1967) 46.
27. Haitun, S.D. Stationary scientometric distributions. Part I. Different approximations, Scientometrics, 4(1982) 5-25. Part II. Non-Gaussian nature of scientific activities, Scientometrics, 4(1982) 89-104. Part III. The role of the Zipf distribution, Scientometrics, 4(1982) 181-194.
28. Haitun, S.D. On M. Kunz's article " A case study against Haitun's conjectures", Scientometrics, 13(1988) 35-44.
29. Kendall, M.G. The bibliography of operational research, Operational Research Quarterly, 11(1960) 31-36.
30. Kunz, M. A case study against Haitun's conjectures, Scientometrics, 13(1988) 25-33.
31. Levy, P. Calcul des probabilites, Gauthier-Villars, Paris, 1925.
32. Lindsey, D. Production and citation measures in the sociology of science : The problem of multiple authorship, Social Studies of Science, 10(1980) 145-162.
33. Lotka, A.J. The frequency distribution of scientific productivity, Journal of the Washington Academy of Science, 16(1926) 317-323.
34. Maia, M.J.F.and M.D. Maia, On the unity of Bradford's law, Journal of Documentation, 40(1984) 206-216.
35. Mandelbrot, B.B. The fractal geometry of nature, W.H. Freeman, New York, 1983 (cf. especially chapter 38).
36. Montroll, E.W. and M.F. Shlesinger, Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise : A tale of tails, Journal of Statistical Physics, 32(1983) 209-230.
37. Naranan S. " Power law " version of Bradford's law: Statistical tests and methods of estimation, Scientometrics, 17(1989) 211-226.
38. Oluic-vukovic, V. Impact of productivity increase on the distribution pattern of journals, Scientometrics, 17(1989) 97-109.
39. Oluic-vukovic, V. Journal productivity distribution: Quantitative study of dynamic behaviour, Journal productivity distribution : Quantitative study of dynamic behaviour, Journal of the American Society for Information Science, 43(1992) 412-421.
40. Pao, M.L. An empirical examination of Lotka's law, Journal of the American Society for Information Science, 37(1986) 26-33.
41. Peritz, B.C. A Bradford distribution for bibliometrics, Scientometrics, 18(1990) 323-329.
42. Praunlich, P. and M. Kroll, Bradford's distribution : A new formulation, Journal of the American Society for Information Science, 29 (1978) 51-55.
43. Solla Price, D. De J , A general theory of bibliometric and other cumulative advantage processes, Journal of the American Society for Information Science, 27(1976) 292-306.
44. Qiu, L. and J. Tague, Complete or incomplete data sets. The Groos droop investigated, Scientometrics, 19(1990) 223-237.
45. Rousseau, R. Breakdown of the robustness property of Lotka's law : The case of adjusted counts for multiauthorship attribution, Journal of the American Society for Information Science, 43(1992) 645-647.
46. Vlachy, J., Evaluating the distribution of individual performance, Scientia Yugoslavica, 6(1980) 155-189.
47. Yablonsky, A.I., Stable non-Gaussian distributions in scientometrics, Scientometrics, 7(1985) 459-470.

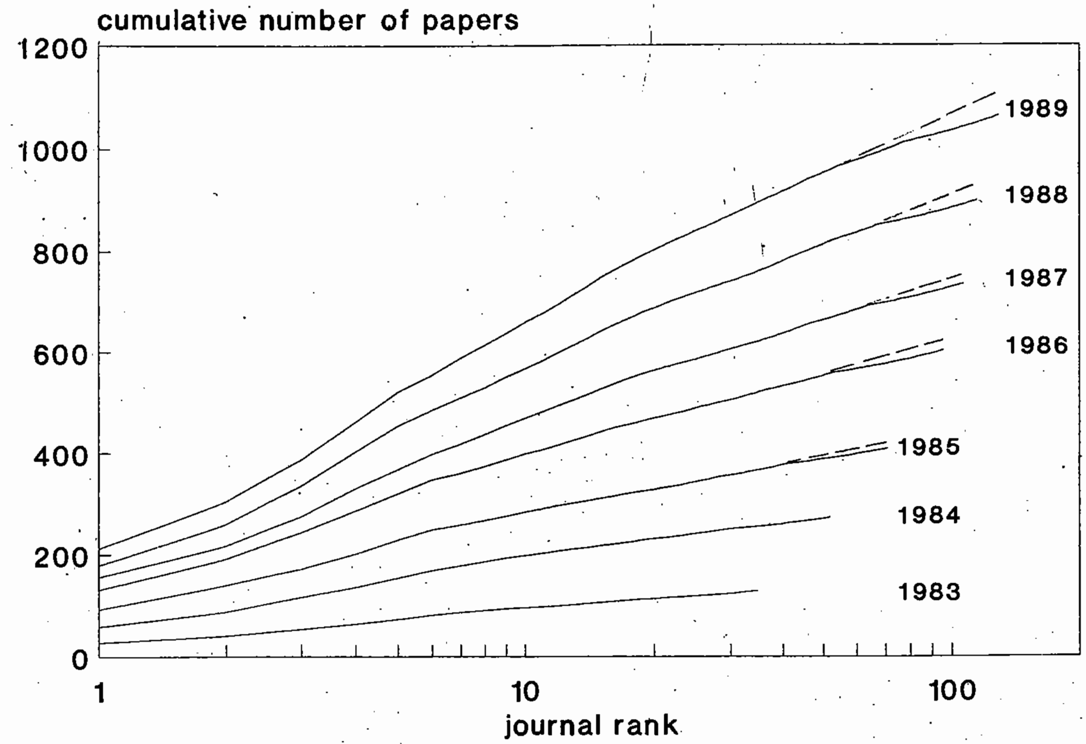


Fig. 1 : Distribution of QHE papers published in journals during seven time intervals (Bradford's diagram)

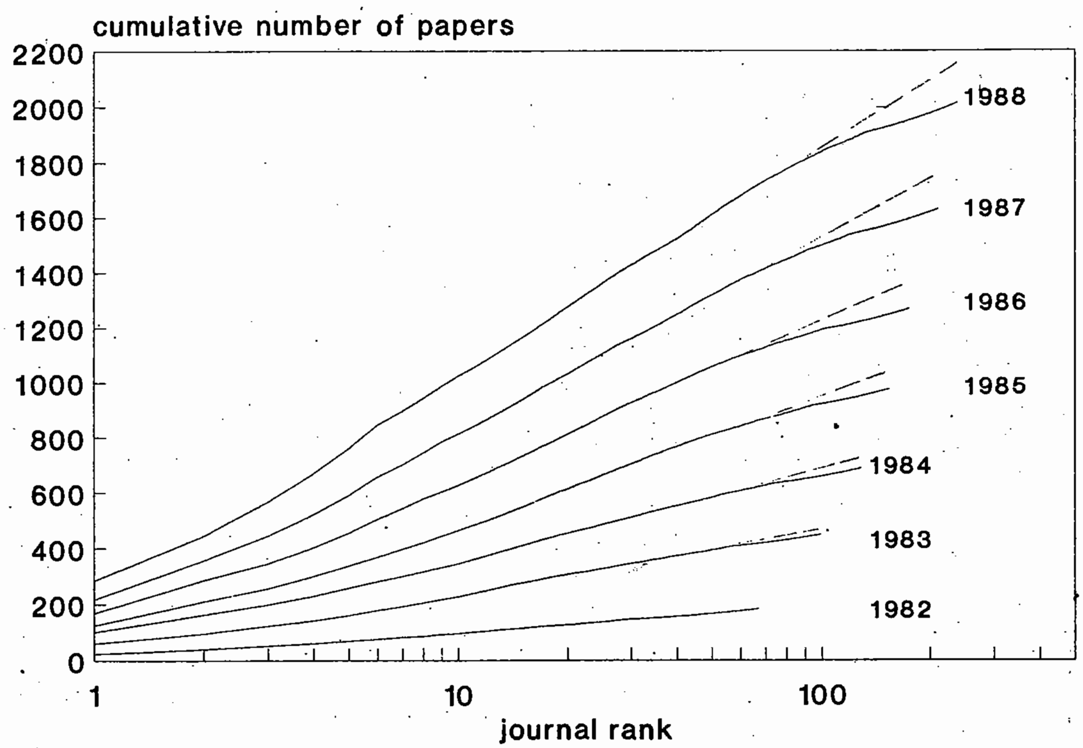


Fig. 2 : Distribution of MD papers published in journals during seven time intervals (Bradford's diagram)

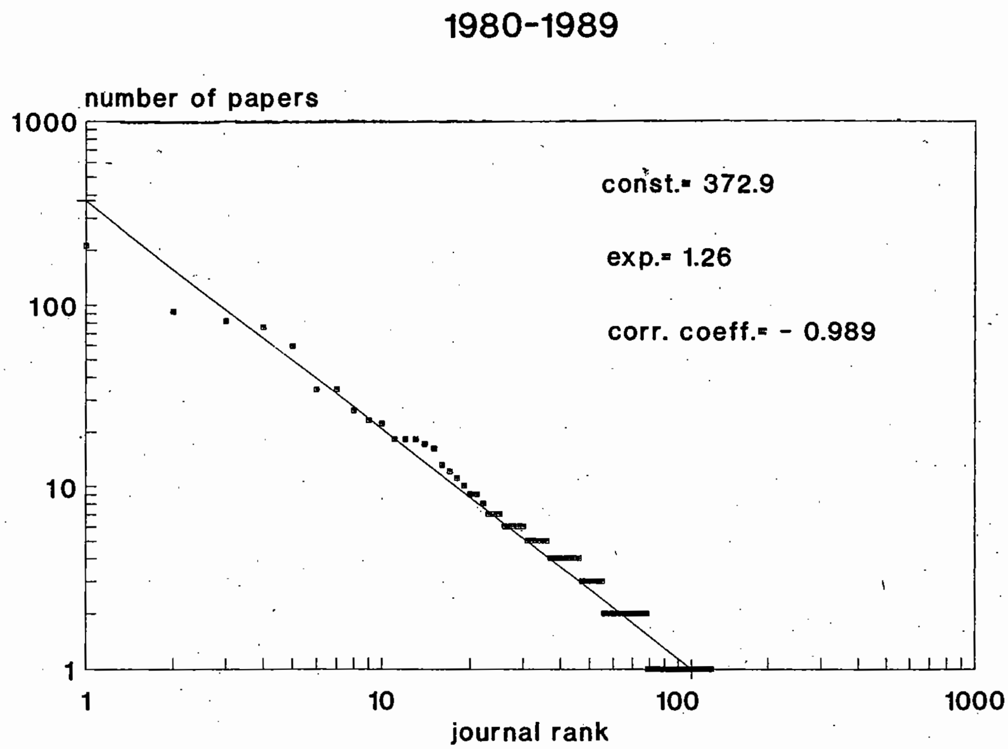


Fig. 3 : Rank-frequency distribution of QHE papers published in journals (Zipf-Pareto form)

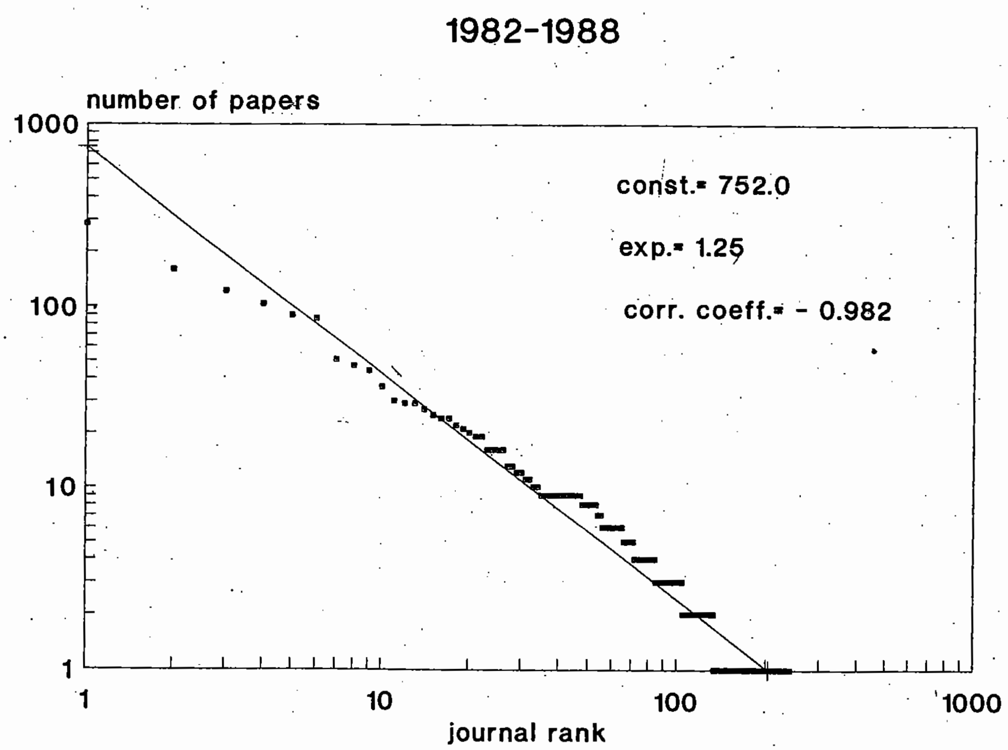


Fig. 4 : Rank-frequency distribution of MD papers published in journals (Zipf-Pareto form)

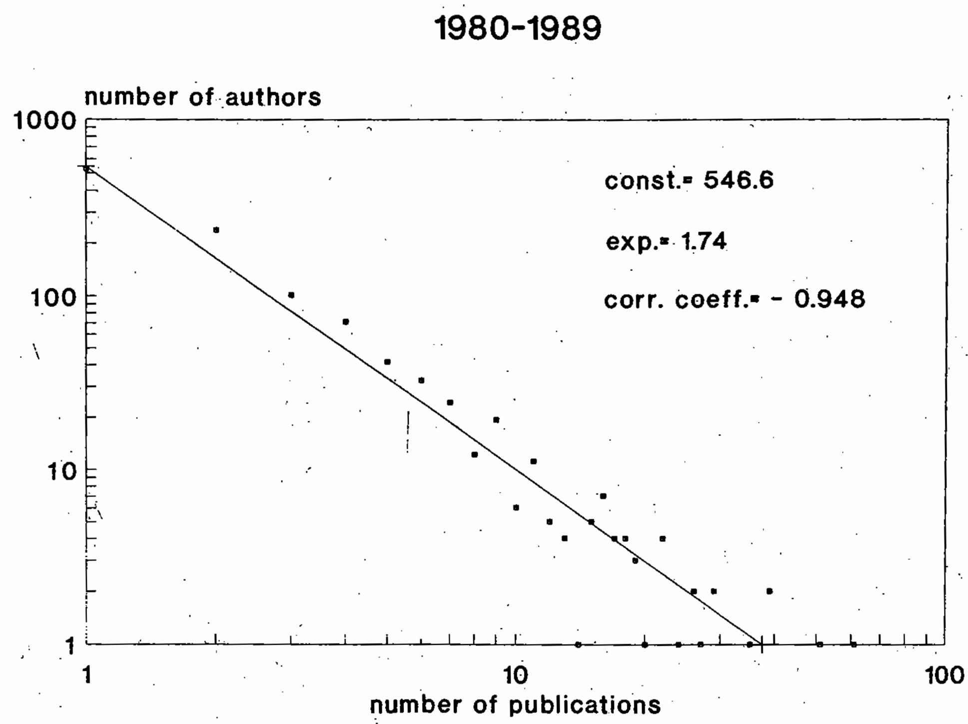


Fig. 5 : Distribution of individual productivity in the speciality of quantum Hall effect (Lotka's plot)

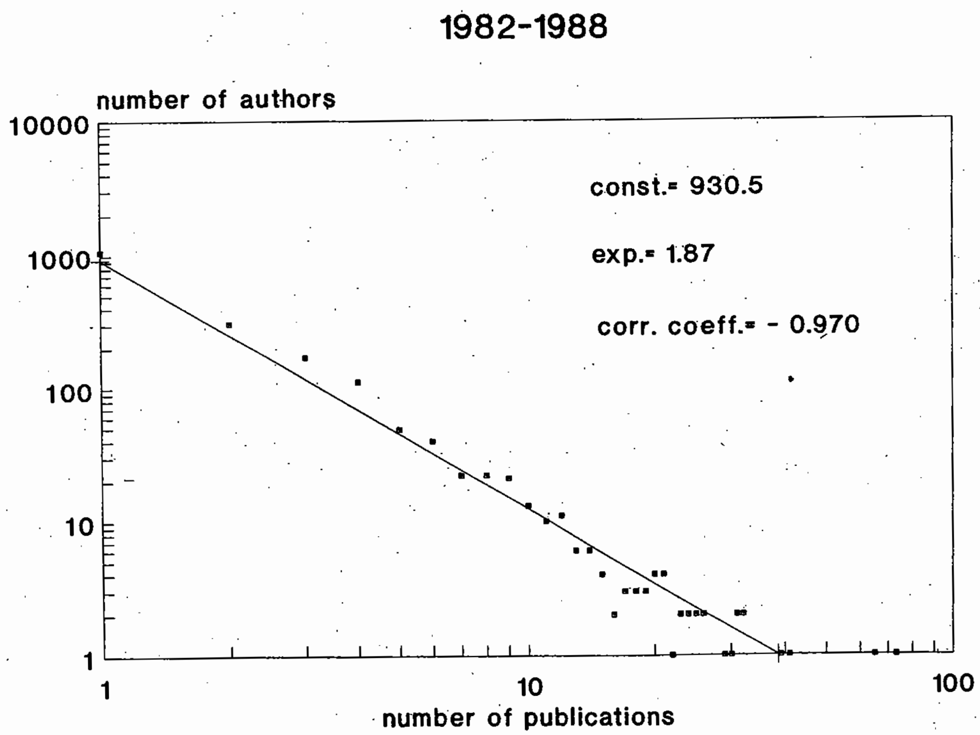


Fig. 6 : Distribution of individual productivity in the speciality of molecular dynamics simulations (Lotka's plot)

Core Journals in Immunology : Correlation Analysis : Rank v/s Rank and Rank v/s Impact Factor

Jagdish Arora
Indian Institute of Technology, New Delhi -110016. India
Usha Pawan
Deptt. of Library & Information Science, University of Rajasthan, Jaipur, India

Spearman's Ranks order coefficient correlation has been used in this study to measure the degree of association or agreement periodicals in immunology. Three correlation studies have been conducted . In the first study, the correlation coefficient between ranks of 48 core journals common to the two independent ranked lists of journals derived from citations collected from reviewing serials each containing 66 core journals, computes to 0.766 which is rated very high and indicate near perfect and positive rank correlation. In the second correlation study, computation of Spearman's rank correlation coefficient suggest positive but low degree of correlation between ranks assigned according to citation received and the rank assigned according to the Impact Factor. In the third case, the ranked lists of core journals are first divided in five major subject categories and then re-ranked according to their citation performance in each category. The ranking is then done according to the 1988 Impact Factor of journals in each category to assess the degree and direction of correlation. A marked increase in the value of correlation coefficient varying from 0.32 to 0.96 can be noted which indiate high degree of correlation between the two ranking methods.

1. Introduction

The scientific advancement is accepted only after it has been published and subject to scrutiny by the rest of the scientific community. The results of research become completely scientific only when they are published, said Ziman[11]. The primary scientific literature, thus, represent the sole legitimate base of scientific information and is a means for stamping an idea as one's possession. It is also considered as a medium for promoting self interest on the road of recognition and for enhancing professional reputation in one's field of specialization. Primary journals continue to remain the main stream of scientific communication because of the inherent need in contemporary science to preserve a permanent record of results for later reference. Nevertheless, the number of learned journals has grown so large that not all relevant papers are readily accessible. The mass of professional journals, said Quine[8], is so indigestible and so little worth digesting that the good papers, though more numerous than ever, are increasingly in danger of being overlooked. "It is easier to make a scientific discovery than to learn whether it has already been made" said Bernal[4]. Post-war proliferation of scientific literature caused by tremendous post-war expansion in scientific and technological research activities resulted in a flood of scattered information so great that the scientists are no more in a position to scan regularly and consistently all publications that contain information which may be of interest to them. A report by King[7] shows the continuing exponential growth of scientific literature published in primary journals. The report reveals that the number of journals published world wide have increased from 18,800 titles in 1960 to 49,440 titles in 1974. 28th edition of Ulrich International Periodical Directory[10] gives information on 1,11,950 serial titles. It can be observed that the number of periodical titles during ± 15 years between 1960 to 1974 and 1974 to 1989 has multiplied by ± 2.6. Likewise, the number of articles published by these journals have also increased many fold with increase in periodicity of their publication. The state of information explosion in biomedical sciences was dramatically illustrated by Durack[6] in terms of the weight of the Index Medicus measured over past 100 years. The situation, in fact, is much worse because Index Medicus includes citations of only about 3,500 periodicals from amongst the world's all biomedical periodicals which were close to 20,000 in 1978.

It is, however, evident that the problem of information overload in biomedical sciences in terms of its quantitative growth have received utmost attention during past half a century. A number of government and non-government organizations in the developed world direct their efforts towards production of large-scale computerized information systems. These efforts resulted in the production of a number of large computerized database of the Institute of Scientific Information, Philadelphia. Currently, nearly 5,000 online databases are available publicly through a number of online vendors like DIALOG, DATASTAR, ORBIT, BRS, STN IN USA; Pergamon Online, BLAISE ONLINE in UK; DMDI in Germany; Euronet and DIANA in Europe; ESAIRS in Italy and CN/OLE in Canada. Many of these databases are now available on Compact Disk format also. The modern information technology has thus been effectively used not only for collection, organization and storage of vast amount of bibliographic data into sophisticated computerized databases, but also for their remote access and efficient retrieval using sophisticated computer (software as well as hardware) and communication technology.

Bibliometric methods based on statistical analysis were used[1] to eliminate journals publishing low-quality research articles of less significance and to evolve three ranked list of core journals in the field of immunolgy. This article determines strength of correlation among following variables :

(i) Correlation/association between the ranks of two sets of core journals evolve from the citations collected from the Ann. Rev. Immunol and Immunol. Rev.;

(ii) Correlation between the ranks assigned on the basis of citation received and ranks assigned on the basiss of 1988 ISI's Impact Factors for the three sets of ranked lists of periodicals; and

(iii) Correlation / association between ranks assigned on the basis of citation received and ranks assigned on the basis of their 1988 ISI's Impact Factors when the core journals are regrouped and reranked according to their subject / discipline covered for the three sets of ranked lists of periodicals.

2. Materials and Methods

The methodolgy adopted for creation of a citation base consisting of 20833 citations to journal articles culled out from two prestigious reviewing journals, namely the Annual Review of Immunology and the Immunological Reviews for four consecutive years from 1983 to 1986 are given in an another article[2] . The criteria for selection of reviewing journals in preference to primary journals has also been eleborated in the above referred articles.

3. Correlation Analysis

The correlation analysis is a statistical technique used for measuring the strength of association between two random variables. Two sets of data are said to be correlated when any change in one set of data accompany by corresponding change in other. Croxdon and Cowden[5] defined it as "the casual relationship existing between any two variables possessing different characteristics". The correlation analysis is useful for measuring the directions as well as degree of relationship that exist between two variables. The technique contribute to understanding of pattern variables. The statistical prediction based on correlation analysis are more reliable and nearer to reality. While an XY graph or a "scatterplot" is considered first step in correlation analysis and is useful in visualizing correlation between two variables, a quantitative measure called "correlation coefficient" is derived to measure exact degree of correlation.

4. Correlation Coefficient

The relative degree of intensity of association between two variables is measured by correlation coefficient. The correlation coefficient is a ratio which express the extent to which the two variables are accompanied by the change in concerned variable. The value of correlation coefficient ranges from -1 to +1. When there is a perfect positive correlation, the value of correlation will be "0" if there is an absence of correlation. Sardana and Sehgal[9] devised the following four broad categories into which correlation coefficient can be classified:

Table 1 : Value and Degree of Correlation Coefficient

Degree	Direction	
	(-)	(+)
Perfect	-1	+1
Very High	-0.75 to -0.99	+0.75 to 0.99
High	-0.50 to -0.74	+0.50 to 0.74
Low	-0.25 to -0.49	+0.25 to 0.49

Since most of the "real world" situation do not involve perfect correlation of +1 or -1, in most cases correlation coefficient (p) is something between 0 and +1 or between 0 and -1 depending on whether the correlation is +tive or -tive.

4.1. Rank Correlation

Ranks are value of an ordinal number which indicate order or position in a series. When objects are arranged in order according to some quality which they all possess to a varying degree, they are said to be ranked with respect to that quality. The arrrangement as well as the process is called ranking. The rank of each object indicates its respective position in the ranking and represents cumulative numbers of sources when they are arranged in decreasing order by number of items. Rank correlation can thus be defined as method to measure the degree of association or agreement with respect to order or ranks between pairs of objects. As is done in the present study, the items in two series of data are ranked or scored according to an attribute common to both. The correlation between the ranks of such series is called the "Rank Correlation" and the coefficient of correlation by ranks is calculated by Spearman's Ranks order Coefficient Correlation represented by (p).

$$P = 1 - \frac{6\sum D_i^2}{n(n^2 - 1)}$$

Where D is difference between the ranked values of X and the ranked value of Y; n is the number of items in a series.

5. Corelation between the Ranked Lists of Core Journals

The two sets of ranked lists of core journals were developed from 10,024 and 10,809 citations to journal articles collected from the Ann. Rev. Immunol. And the Immunol. Rev. between 1983 and 1986 respectively[3]. Both the ranked lists were restricted to 66 journals as remaining journals in both the lists had very insignificant number of citations to their credit. The journals listed in both the ranked lists were matched to identify journals common to both the lists. Table 2 lists 48 journals common to both the lists of core journals arranged in their respective rank orders. In order to assess the exact correlation coefficient, the 48 journals common to both the core lists of journals were reranked according to citations received by them in their respective core lists. This was done to avoid gaps in the two ranked lists caused by deletion of journals that were not common in the two core lists. Retention of original ranks with such gaps in ranks would have caused distortion in the derivation of correlation coefficient. The table also shows computation of summation of square of variables in order to derive the Spearman's Rank Order Coefficient essentially to demonstrate the methodology followed. Calculation of Spearman's Rank Order Coefficient is given below :

$$p = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6 \times 4312}{48\ (48^2 - 1)}$$

$$= 1 - \frac{25872}{110544} = 1 - 0.234 = 0.766$$

Where $\sum di^2$ = 4312 and n = 48

TABLE 2

COMPUTATION OF CORRELATION COEFFICIENT THAT ARE COMMON IN THE LISTS OF CORE JOURNALS EVOLVED FROM THE ANN. REV. IMMUNOL. & THE IMMUNOL. REV. (1983 - 1986)

SR. NO.	NAME OF JOURNAL	RANK IN ANN. REV. IMMUNOL.	RANK IN [redacted] REV IMMUNOL	Di (3-4)	Di^2
1	J. Immunol.	1	2	-1	1
2	J. Exp. Med.	2	1	1	1
3.	Nature	3	3	0	0
4.	Proc. Natl. Acad, Sci.	4	4	0	0
5.	Cell	5	8	-3	9
6.	Europ. J. Immunol.	6	5	1	1
7.	Science	7	12	-5	25
8.	Immunol. Rev.*	8	7	1	1
9.	J. Biol. Chem.	9	19	-10	100
10.	Immunology	10	13	-3	9
11.	J. Clin. Invest.	11	25	-14	196
12.	Mol. Immunol.	12	28	-16	256
13.	Cell. Immunl.	13	9	4	16
14.	N. England J. Med.	14	15	-1	1
15.	Immunogenetics	15	11	4	16
16.	Biochemistry	16	40	-24	576
17.	Transplantation	17	6	11	121
18.	Lancet	18	14	4	16
19.	Adv. Immunol.	19	23	-4	16
20.	Ann. Rev. Immunol.	20	26	-6	36
21.	Fed. Proc.	21	32	-11	121
22.	Scand. J. Immunol.	22	18	4	16
23.	Clin. Exp. Immunol.	23	17	6	36
24.	Infect. Immun.	24	42	-18	324
25.	EMBO J.	25	21	4	16
26.	Nucleic Acids Res.	26	27	-1	1
27.	Biochem. J.	27	44	-17	289
28.	Cold Sprng. Harb. Symp. Quant. Biol.	28	37	-9	81
29.	Ann. N Y Acad. Sci.	29	24	5	25
30.	Arthr. Rheum.	30	30	0	0
31.	Transplant. Proc.	31	10	21	441
32.	Clin. Immunol. Immunopathol.	32	29	3	9
33.	Blood	3	16	17	289
34.	Proc. Soc. Exp. Biol. Med.	34	43	-9	81
35.	U. Natl. Cancer Inst.	35	35	0	0
36.	Ann. Rev. Biochem.	36	46	-10	100
37.	Immunol. Today	37	20	17	289
38.	Am. J. Pathol.	38	34	4	16
39.	Ann. Intern. Med.	39	39	0	0
40.	Ann. Inst. Pasteur/Immunol.	40	22	18	324
41.	Prog. Allergy	41	33	8	64
42.	Proc. Royal Soc. (London) B. Biol. Sc.	42	31	11	121
43.	Mol. Cell. Biol.	43	48	-5	25
44.	Nature-New Biol.	44	47	-3	9
45.	J. Investigative Dermatol.	45	41	.4	16
46.	Springer Sem. Immunopathol.	46	45	1	1
47.	J. Immunol. Methods	47	36	11	121
48.	Immunobiology	48	38	10	100

$\sum Di^2 = 4312$
$p = 0.765958$

The spearman's correlation coefficient p = 0.766 suggest that a very high and positive correlation exist between the ranks of two lists of core journals serve as statistical validation to the selection of rèviewing serials as source of citations in preference to the citations from primary journals. A close scrutiny of the citing articles of two reviewing journals for the period covered in the study, however, reveal that their subject coverage are quite dissimilar to each other even though 72.7% core journals cited by the articles published in them are common. It may be noted that the Annual Review of Immunology commenced in 1983 by Annual Reviews, Inc. with its first volume while Immunological Reviews was launched in 1969 as Transplantation Reviews by Munksgaard publishers, a leading publication house of medical journals. Immunological Reviews appears 6 times a year with each issue devoted to publication of articles on different aspect of research in immunology.

6. Correlation between Ranks of Core Journals and their Impact Factor

The Impact Factor is the average citation rate per published articles assigned to a journal by the Institute of Scientific Information by dividing the number of times a journal has been cited by the number of articles it has published during some specific period of time. Although the ranks in the present study is assigned to a journal on the basis of its citation performance alone, a statistical correlation, it is assumed, should exist between the ranks assigned in this study with their Impact Factors.

The value of Spearman's correlation coefficient between the ranks assigned according to citation received and the ranks assigned according to citation received and the ranks assigned according to the 1988 Impact Factor was calculated according to the Spearman's method given above. The journals whose impact factors are not available have also been included considering their Impact Factor to be zero. The Spearman's correlation coefficient for the three ranked lists are given in below in table 3.

Table 3 : Ranks vs Impact Factors : Rank Order Correlation Coefficient for three sets of Ranked lists

Ranked lists from citations collected from		ΣDi^2	n	p
1.	Ann. Rev. Immunol.	26408	66	0.45
2.	Immunol. Rev.	33370	66	0.30
3.	Ann. Rev. Immunol. & Immunol. Rev.	121438	103	0.33

The value of correlation coefficient given in the table 3, column 4 is positive and shows low degree of correlation between the ranks assigned on the basis of citation performance of a journal and ranks assigned according to their 1988 Impact Factors. The low degree of correlation is also illustrated in the scatter graphs between the rank orders and their Impact Factors in figure 1, 2 and 3 for three ranked lists of core journals. (See Appendix)

7. Subject-wise Categorization of Journals and correlation with Impact Factors

Low degree of Correlation is observed between the decreasing rank of core periodicals with their 1988 Impact Factor. A marked increase can, however, be observed in degree of correlation between the decreasing rank of journals and their impact factors when the ranked lists of core journals are re-arranged and reranked according to five major subject categories covered by them. A number of variations can however be noticed in each subject category. The possible explanation for these variations are given below :

1. Reviewing journals, owing to their very nature, receive large number of citations in primary journals although their scientific merit may not necessarily correlate with the impact factor that they receive. Moreover, review articles are not necessarily cited in subsequent reviews. Further, possibility of an early subsequent review on the same subject is a rare chance. Annual Reviews, Advances, etc. are few examples of this category.

2. Vaccination is the success story of immunology. Immunology as a science owes its existence to its clinical applications, i.e. vaccines and immunotherapy. A number of journals on immunology, thus necessarily deal with the vaccines, Immunotherapy, immunopathology , etc. The impact factors of such journals are greater than those which deal with the immunology as a pure science. Examples of such journals are : Journal of Experimental Medicine; Infection and Immunity; Clinical and Experimental Immunology.
3. Certain journals, although categorized under immunology, deals with subjects that are actually amalgamation of two or more sciences. The journals devoted to these are purely multi-disciplinary. These journals do not follow the proposed correlations for obvious reasons. Immunogenetics, Journal of Allergy and Clinical Immunology, etc. are examples of this category.

Tables 4-8 (Appendix I) list ranked list of core periodicals in Immunology arranged according to five major subject categories for three sets of data. Figure 4,5 and 6 are scatter graphs for two to categories (Immunology and Medical Sciences) plotting ranks of journals on the ordinate against their Impact Factor on abscissa for three sets of ranked lists of periodicals. (See Appendix)

Table 9 gives their correlation coefficient calculated with the Spearman's formula. Although the value of "n" is very low for some subject categories for obtaining reliable value of "p", the correlation coefficient of most of the subject categories varies between 0.31 to 0.82 (with only two exceptions marked with *). The value of "p" thus obtained is fairly high in most of the cases and is of great significance.

8. Inferences and Discussion

The near-perfect correlation between ranks or two independent sets of core journals may be interpreted as statistical consent to the selection of reviewing serials as source of citation. The citations collected from the reviewing literature have an edge over the citations collected from the primary sources. Bibliometric study of any subject based on the citations collected from the selected primary journals, suffer from a number of bias which are characteristic to the items referred in primary media of communication. Some of the typical problem relate to self-citations, implicit citations, limit in the time period covered and the citations given for reasons other than professional. On the contrary, the citations given in a review article do not suffer from such problems. A review carries a highly analytical, organized and complete synthesis of information in a given area. The process of reviewing involves great amount of time and effort on the part of experts in summarizing consensus of scientific opinion on a given topic. Experts to support the credibility of their views, cite highest quality and the most relevant literature. Morever, because of constraints of space, the number of citation is severely limited and redundancy of citations is minimized in a review article.

Table 9 : Correlation between Ranks assigned on the basis of citation received and the Ranks assigned on the basis of their 1988 Impact Factors.

Source of Citation for Ranked list	Immunology	Medical Science	Biochem. Biol.	Cell/Mol. Biol.	General Sciences
Ann. Rev.	n = 21	22	12	7	4
Immunol.	$\Sigma di^2 = 942$	941	174	32	6
	p = 0.40	0.47	0.39	0.33	0.4
Immuno.	n = 24	20	9	7	5
Rev.	$\Sigma di^2 = 1382$	569	158	32	5
	p = 0.40	0.53	-0.32*	0.54	0.75
Ann. Rev.	n = 27	40	20	11	5
Immunol.	$\Sigma di^2 = 1494$	6560	1336	148	8
&	p = 0.54	0.35	0.004*	0.327	0.96
Immunol. Rev.					

The low degree of correlation between ranks assigned on the basis of citation performance of a journal and ranks assigned according to their 1988 impact factors can be attributed to the eclectic nature of immunology as a basic science. The amalgamation of three sciences namely biochemistry, cellular biology and genetics led to the establishment of immunolgy as a scientific discipline. Immunology, thus, is an eclectic science that draws on many branches of knowledge. The pursuit of research in modern immunology requires a broad cross-disciplinary knowledge-base with strong information back-up on the protein chemistry, modern cell-culture techniques, genetic engineering, microbiology, virology, nucleic acid chemistry, molecular and cellular biology. The core list of journals in immunology, thus, constitute of journals from various other discipline beside journals that are specifically devoted to publications of research articles in immunology. "The lack of correlation in the rank assigned according to the citation performance and rank assigned according to its impact factor can be attributed to the fact that number of citations received by a given journal in the ranked list of journals accounts for citations of articles in immunology while the impact factor is derived from the total no. of citations and the total number of items published during a given year. Likewise corresponding high degree of correlation between the rank according to no. of citations and rank according to its impact factor when the journals are re-arranged and re-ranked according to the major subject categories can be explained as the journals on a specific subjects are now culled out and arranged according to their citation performance. The corresponding impact factor correlates with the rank assigned according to the citation performance as the impact factor is also specific to the given subject speciality.

Another factor that can explain the varying degree of rank correlation is the citation practices that are known to vary from one field to another. Wide citation may be necessary practice in one field, but in another it may be considered as a redundancy. Likewise half-life and obsolescence rate of journals also vary from one field to another. All these factors contribute to the citation rate and the impact factor of a journal from one field to another. While the citation rate and corresponding impact factor of journals in biochemistry and molecular biology is the highest, it is lowest in case of aerospace engineering and technology and agricultural economics and policy. Likewise citation rates and its corrresponding impact factor should differ for journals in other disciplines that publish research articles in immunology. An article published in Ann. Rev. Biochemistry or cell is likely to have higher citations than one that is published in European J. of Immunology. Though the paper might have high research value in their respective fields. It may be noted that science citation Index lists a total of 144 subject categories and highest impact factor of journals varies to a large extent from one subject field to another.

9. Conclusion

Correlation studies were conducted on the three sets of core journals in the field of immunology evolved by applications of bibliometric methods based on statistical analysis. Technique of rank correlation was used to measure the degree of association or agreement with respect to order or ranks between the two sets of core journals. The coefficient of correlation by ranks is calculated by Spearman's Ranks Order Coefficient Correlation. Three correlation studies were conducted. In the first study, the correlation coefficient between ranks of 48 core journals common to the two pairs of ranked lists, each containing 66 core journals, computes to 0.766 which is rated very high and indicate near-perfect and positive rank correlation. It is significant that 48 journals common to both the ranked lists of journals account for 17,679 citations (84.86%) as against 19,306 (92.67%) covered by 103 journals given in the ranked lists of core periodicals evolved from the combined citations from both the reviewing journals. In the second correlation study, computation of Spearman's rank correlation coefficient suggest positive but low degree of correlation between ranks assigned according to citation received and the rank assigned according to citation received and the rank assigned according to the Impact Factor for all the three sets of core journals. The values of correlation coefficient given below in is positive and ranges from 0.30 to 0.45. In the third case, the three core lists of journals are first divided in the five major subject categories and then re-ranked according to their citation performance in each category. The ranking is then done according to the 1988 Impact Factor of journals in each category to access the degree and direction of correlation. A marked increase in the value of correlation coefficient varying from 0.32 to 0.96 can be noted which indicate high degree of correlation between the two ranking methods. It may, however, be noted that the correlation coefficient calculated by Spearman's formula is less reliable than Pearson's correlation co-efficient formula. While the former is applicable in the cases where the values are available in the form of ranks, the later is applicable when the actual values are available and are distributed normally in a given population.

Correlation analysis is thus a useful statistical technique to study the direction and degree of relationship between two sets of data. The technique provides firm statistical ground to the proposed ranked lists of periodicals in immunology based on the quantitative performance of journals in terms of totality of citations obtained from the two reviewing journals. The prediction based on correlation analysis will be more reliable and near to reality.

References

1. Arora, Jagdish and Pawan, Usha (1995). Core journals in immunology : A comparative analysis of bibliometric parameters. (paper presented at the Fourth International Conference on the Bibliometrics,Informatics and Scientometrics, 11 - 15 Sept., 1993, Berlin, Germany) JISSI **1**(2) 99-122.
2. Ibid.
3. Ibid.
4. Bernal, J. D. (1939). The social function of science. London, Rutledge & Kegan Paul.
5. Croxton, F.E. and Cowden, D.J. Applied general statistics. 2nd ed. New Delhi, Prentice Hall of India, 1964, p. 766.
6. Durack, D.T.(1978), The weight of medical knowledge. N. Engl. J. Med., **298** (14), 773-775.
7. King, D.W. (1976) Statistical indicators of scientific and technical communication (1969 - 1980). Vol. II, Washington, D.C., NTIS.
8. Quine, W.V. (1974) Paradoxes of plenty. Daedalus, **103**(4), 38-40.
9. Sardana, J. L. and Sehgal, R.L. (1982) Statistical Methods for Librarians, New Delhi, Ess Ess Publications, p. 167-182.
10. Ulrich's International Periodical Directory. 28th Ed. (1989) 3 vols. New York, R.R. Bowker.
11. Ziman, J.M. (1969) Information, communication and knowledge. Nature, **224**, 318-324.

Table 4
General And Applied Immunology

ANN. REV. IMMUNOL. REV.

SR. NO. (1)	RANK NO. (ORG)	NAME OF JOURNAL (2)	IMPACT FACTOR (3)
1.	1.	J.Immunol.	6.91
2.	2.	J.Exp. Med.	11.83
3.	5	Europ. J. Immunol.	4.34
4.	7	Immunol. Rev.*	9.67
5.	10	Cell. Immunol.	1.81
6.	12	Immunogenetics	3.01
7.	13	Immunology	2.66
8.	18	Mol. Immunol %%	1.95
9.	19	Cl in. Exp. Immunol.	2.28
10.	20	Scand. J. Immunol	2.19
11.	22	Adv. Immunol.	16.40
12.	24	Ann. Rev. Immunol.	25.41
13.	27	Immunol. Today	10.65
14.	30	Tissue Antigens	1.52
15.	32	Infect. Immun.	3.21
16	33	Ann. Institut Pasteur (Immunol.)	1.38
17.	34	Cl in. Immunol. Immunopathol.	1.97
18.	42	Human Immunology	3.22
19.	48	J. Immunol. Methods	2.17
20.	50	Immunobiology	2.71
21.	57	Springer Sem. Immunopathol.	1.68
22.	58	J. Reprod. Immunol.	1.67
23.	60	Int. Arch. allergy Appl. Immunol.	1.31
24.	66	J.Mol. Cell Immunol.	2.89
25.	73	J. Allergy Clin. Immunol. !!	2.95
26.	87	J. Immunogenetics	1.12
27.	97	Am. J. Reprod. Immunol.	1.25

ANN. REV. IMMUNOL

SR NO. (1)	RANK NO. (ORG)	NAME OF JOURNAL (2)	IMPACT FACTOR (3)
1.	1	J.Immnol.	6.91
2.	2.	J.Exp. Med.	11.83
3.	6	Europ. J. Immunol.	4.34
4.	8	Immunol. Rev.*	9.67
5.	11	Immunology	2.66
6.	12	Mol. Immunol.!!	1.95
7.	13	Cell. Immunol.	1.81
8.	15	Immunmogenetics	3.01
9.	19	Ann. Rev. Immunol.	25.41
10.	20	Adv. Immunol.	16.40
11.	22	Second. J. Immunol.	2.19
12.	23	Clin. Exp. Immunol.	2.28
13.	24	Infect. Immun.	3.21
14.	39	Cl in. Immunol Immunopathol.	1.97
15.	44	Immunol. Today	10.65
16.	49	Int. Arch. Allergy Appl. Immunol.	1.31
17.	53	J Allergy Cl in. Immunol.	2.95
18.	55	Ann. Inst. Pasteur/Immunol.%%	1.38
19.	57	Springer Sem. Immunopathol.	1.68
20.	62	J. Immunol. Methods	2.17
21.	64	Immunobiology	2.71

IMMUNOL. REV.

1.	1	J. Exp. Med.	11.83
2.	2	J. Immunol.	6.91
3.	5	Europ.J.Immunol.	4.34
4.	7	Immunol. Rev.*	9.67
5.	9	Cell. Immunol.	1.81
6.	11	Immunogenetics	3.01
7.	13	Immunology	2.66
8.	17	Clin. Exp. Immunol.	2.28
9.	18	Scand. J. Immunol.	2.19
10.	19	Tissue Antigens	1.52
11.	21	Immunol. Today	10.65
12.	23	Ann. Inst. Pasteur (Immunol.)!!	1.38
13.	24	Adv. Immunol.	16.40
14.	25	Human Immunol.	3.22
15.	28	Ann. Rev. Immunol.	25.41
16.	30	Mol. Immunol.!	1.97
17.	31	Clin. Immunol. Immunopathol.	1.97
18.	38	J.Reprod. Immunol.	1.67
19.	41	J. Immunol. Meth.	2.17
20.	43	Immunobiology	2.71
21.	48	Infec. Immun.	3.21
22.	52	J. Mol. Cell.Immunol.	2.89
23.	59	Springer Sem. Immunopathol.	1.68
24.	66	Am. J. Reprod. Immunol.	1.25

Table 5 : Medical Sciences

ANN. REV. IMMUNOL. & IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	8	Transplantation	2.98
2.	14	Transplant. Proc.	1.10
3.	15	New England J. Med.	21.15
4.	16	Lancet	14.48
5.	17	J. Clin. Invest.	7.59
6.	23	Blood	6.85
7.	31	Arthr. Rheum.	4.81
8.	37	American J. Pathol.	4.36
9.	40	Prog. Allergy	3.75
10.	41	J. Natl. Cancer Inst.	2.71
11.	44	Ann. Intern. Med.	8.47
12.	46	Ann. Rheum. Dis.	2.01
13.	49	Cancer Res.	4.30
14.	51	J. Investigative Dermatol.	3.47
15.	55	Int. J. Cancer	3.16
16.	56	Exp. Hematol.	2.71
17.	62	Diabetologia	3.81
18.	64	J. Infectious Dis.	4.91
19.	65	Br. Med. J.	3.14
20.	67	Acta Pathol. Microbiol. Scand.	1.67
21.	68	J. Rheumatol.	1.47
22.	69	American J. Med.	2.73
23.	72	Am. Rev. Respiratory Dis.!	4.48
24.	76	Clin. Res.	0.12
25.	77	Am. J. Obstet. Gynecol.	1.93
26.	78	Br. J. Exp. Pathol.	0.95
27.	79	J. Clin. Endocrinol. & Metabol.	4.09
28.	80	Arch. Dermatol.	1.99
29.	82	J. Am. Med. Assoc. (JAMA)	5.28
30.	84	Neurology	2.97
31.	88	Prostaglandlins	2.12
32.	89	Thymus	0.86
33.	92	J. Physiol.	3.91
34.	93	Rev. Infect. Dis.	2.74
35.	96	Surgery	1.73
36.	99	Adv. Cancer Res.	6.54
37	100	Gastroenterology	6.13
38.	102	Br. J. Dermatol.	1.78
39.	103	Br. J. Obstet. Gynaecol.	1.63

ANN. REV. IMMUNOL.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	10	J. Clin. Invest.	7.59
2.	14	N. England J. Med.	21.15
3.	17	Transplantation	2.98
4.	18	Lancet	14.48
5.	26	Diabetes	4.42
6.	30	Arthr. Rheum.	4.81
7.	34	Transplant. Proc.	1.10
8.	37	Cancer Res.	4.30
9.	38	Blood	6.85
10.	42	Diabetelogia	3.81
11.	43	J. Natl. Cancer Inst.	2.71
12.	45	Int. J. Cancer	3.16
13.	46	Am. J. Pathol. %	4.36
14.	47	Ann. Intern. Med.	8.47
15.	48	Am. Rev. Respiratory Dis. !	4.48
16.	50	Clin. Res.	0:12
17.	52	Prog. Allergy	3.75
18.	56	J. Investigative Dermatol.	3.47
19.	59	Neurology	2.97
20.	61	J. Infectious Dis.	4.91
21.	66	Arch. dermatol.	2.00
22.	65	Prostaglandl in	2.12

IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	6	Transplantation	2.98
2.	10	Transplant. Proc.	1.10
3.	14	Lancet	14.48
4.	15	New Engl. J. Med.	21.15
5.	16	Blood	6.85
6.	26	J. Clin. Invest.	7.59
7.	32	Arthr. Rheum.	4.81
8.	33	Ann. Rheum. Dis.	2.01
9.	36	Prog. Allergy	3.75
10.	37	American J. Pathol. %	4.36
11.	39	Exp. Hematol.	2.71
12.	40	J. Nat. Cancer Inst.	2.71
13.	44	Ann. Intern. Med.	8.47
14.	46	J. Rheumatol.	1.47
15.	47	J. Invest. Dermatol.	3.47
16.	53	American J. Obstet. Gynecol.	1.93
17.	54	Acta Pathol. Microbiol. Scand.	0.92
18.	58	Br. Med. J.	3.14
19.	64	J. Am. Med. Assoc. (JAMA)	5.28
20.	65	J. Clin. Endocrinol. Metabol.	4.09

Table 6: Basic Biology and Biochemistry

ANN. REV. IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	11	J. Biol. Chem.	6.49
2.	25	Fed. Proc.	0.33
3.	26	Biochemistry	4.01
4.	38	Biochem. J.	3.94
5.	39	Proc. Royal Soc. B. Biol. Sci. (Londo	2.23
6.	43	Proc. Soc. Exp. Biol. Med.	1.33
7.	47	Ann.Rev. Biochem.	48.31
8.	52	Biochem. Biophys. Res. Commun.	3.17
9.	53	Biochem. Biophys. Acta	2.28
10.	59	Nature-New Biol.	3.81
11.	63	Australian J. Exp. Biol. Med. Sci.	0.90
12.	70	Phil. Trans. Royal Soc. B. Biol. Sci.	
13.	71	Lab. Investigation	5.14
14.	75	FEBS Letters	3.57
15.	81	J. Cell Biochem. %	4.47
16.	85	J. Reticuloend. Soc.	
17.	86	J. Am. Chem. Soc.	4.57
18.	95	Europ. J. Biochem.	3.15
19.	98	Ann. Rev. Microbiol.	8.44
20.	101	Hoppe-Seyler's Z. Physiol. Chem.**	1.93

IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	20	J. Biol. Chem	6.49
2.	34	Proc. Royal Society(London) B. Biol. Sc	2.23
3.	35	Fed. Proc.	0.33
4.	45	Biochemistry	4.01
5.	51	Proc. Soc. Exp. Biol. Med.	1.33
6.	55	Biochem. J.	3.94
7.	58	Ann. Rev. Biochem.	48.31
8.	60	Nature - New Biology	

ANN. REV. IMMUNOL

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	9	J. Biol. Chem.	6.49
2.	16	Biochemistry	4.01
3.	21	Fed. Proc.	0.33
4.	28	Biochem. J.	3.94
5.	33	Biochem. Biophys. Res. Commun.	3.17
6.	35	Biochem. biophys. Acta	2.28
7.	40	Proc. Soc. Exp. Biol. Med.	1.33
8.	41	Ann. Rev. Biochem.	48.31
9.	54	Proc. Royal Soc. (London) B.Biol.sc.	2.23
12.	58	Nature - New Biology	
10.	60	Australian J. Exp. Biol. Med. Sci.	0.90
11.	63	FEBS Letters	3.57

Table 7 : Molecular and Cellular Biology

ANN. REV. IMMUNOL. & IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	6	Cell	23.91
2.	21	EMBO J.	10.94
3.	28	Nucleic Acids Res.	4.30
4.	35	Cold Sprng. Harb. Symp. Quant. Biol.	2.25
5.	45	J. Cell Biol. $	9.75
6.	54	J. Mol. Biol.	6.56
7.	61	Mol. Cell. Biol.	7.73
8.	74	Cell Tissue Kinetics	0.75
9.	83	Am. J. Hum. Genet.	4.79
10.	91	Ann. Rev. Genet.	15.12
11.	94	J. Cell Physiol.	3.15

ANN. REV. IMMUNOL.

1.	'5	Cell	23.91
2.	25	EMBO J.	10.94
3.	27	Nucleic Acids Res.	4.30
4.	29	Cold Sprng. Harb. Symp. Quant. Biol.	2.25
5.	32	J. Cell Biol.	9.75
6.	36	J. Mol. Biol.	6.56
7.	51	Mol. Cell. Biol.	7.73

IMMUNOL. REV.

1.	8	Cell	23.91
2.	22	J. EMBO	10.94
3.	29	Nucleic Acids Res.	4.30
4.	42	Cold Spring Harbor Symp. Quant. Biol.	2.25
5.	49	Cell and Tissue Kinetics	0.75
6.	57	Am. J. Human Genet.	4.79
7.	63	Mol. Cell. Biol.	7.73

Table 8 : General Sciences

ANN. REV. IMMUNOL & IMMUNOL. REV.

SR. NO. (1)	RANK (ORG)	NAME OF JOURNAL (2)	1988 IMPACT FACTOR (3)
1.	3	Nature	15.76
2.	4	Proc. Natl. Acad. Sci	10.03
3.	9	Science	16.46
4.	29	Ann. N Y Acad. Sci.	0.75
5.	90	C.R. Acad. Sci.	0.37

ANN. REV. IMMUNOL.

1.	3	Nature	15.76
2.	4	Proc. Natl. Acad. Sci.	10.03
3.	7	Science	16.48
4.	31	Ann. N Y Acad. Sci.	0.75

IMMUNOL. REV.

1.	3	Nature	15.76
2.	4	Proc. Natl. Acad Sci.	10.03
3.	12	Science	16.46
4.	27	Ann. NY Acad. Sci. (USA)	0.75
5.	51	Phil. Trans. Royal Soc. (London)	

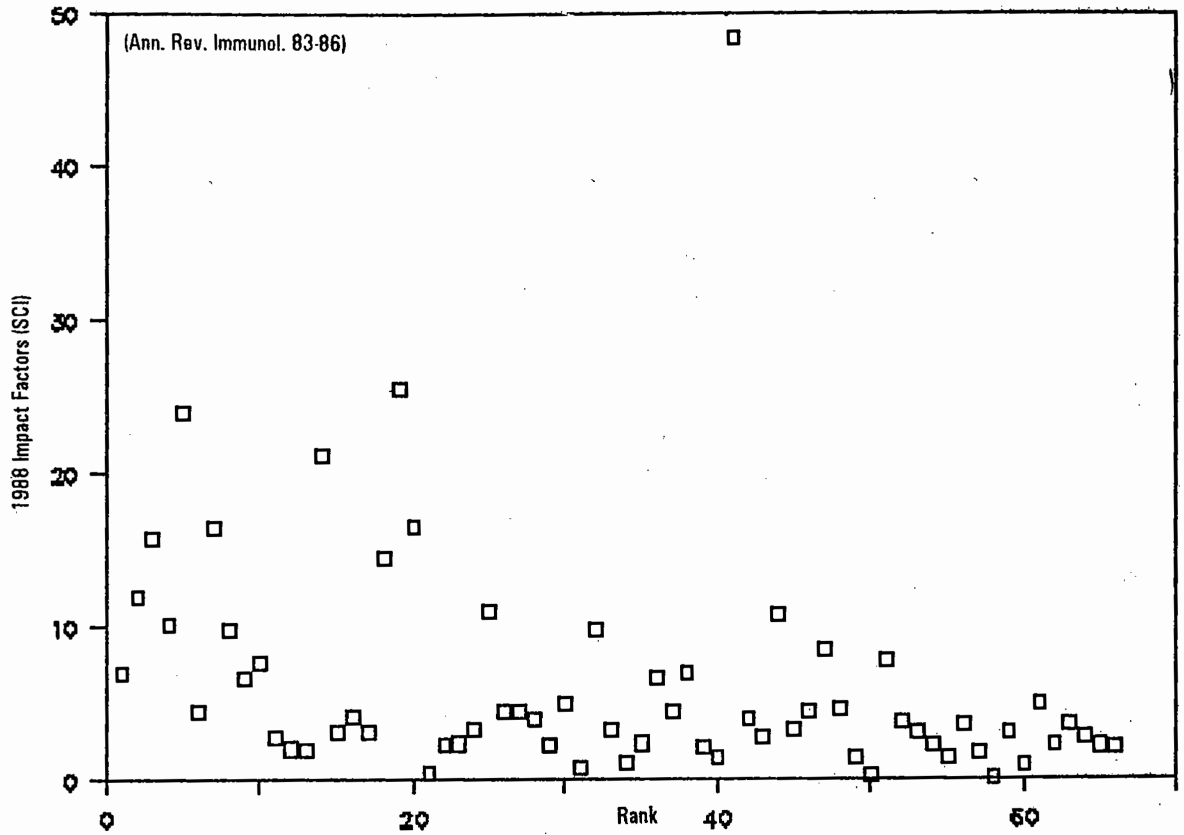


Fig. 1 : Correlations : Ranks & Impact Factors

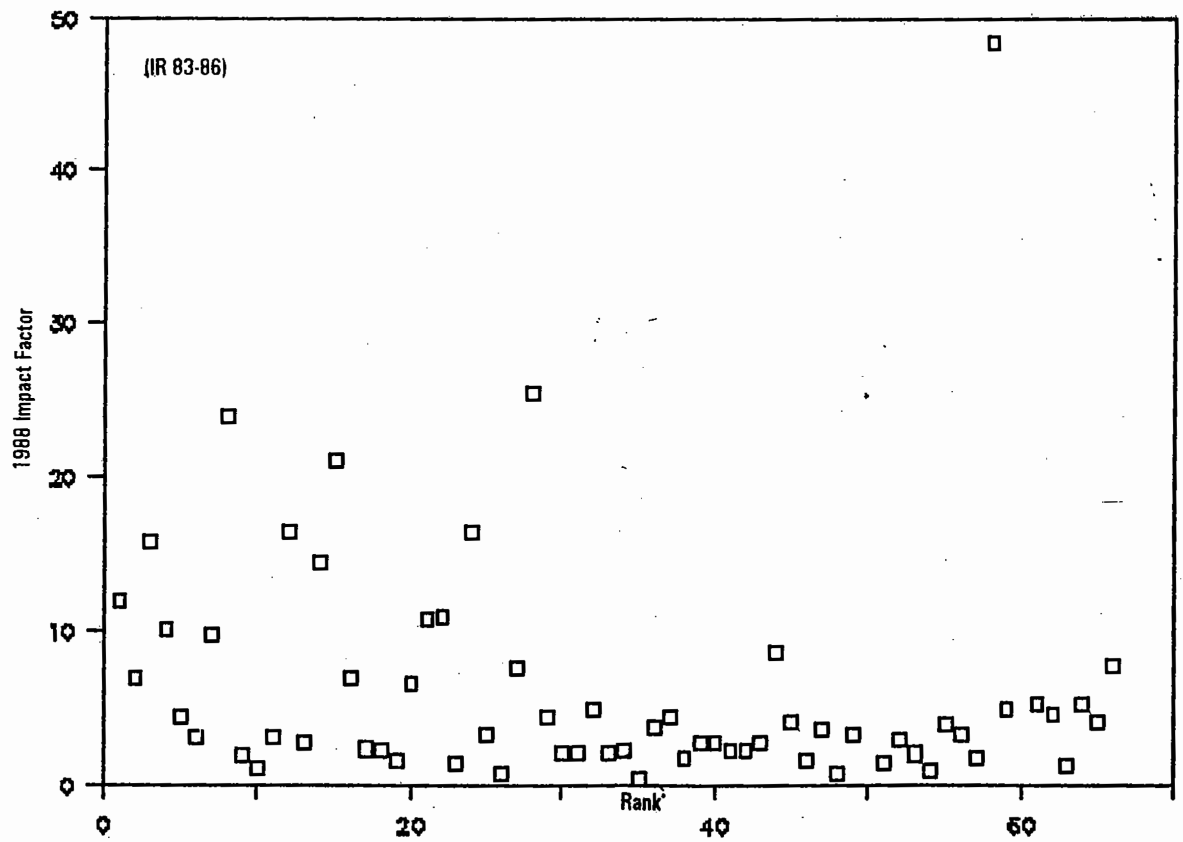


Fig. 2 : Correlations : Ranks and Impact Factors

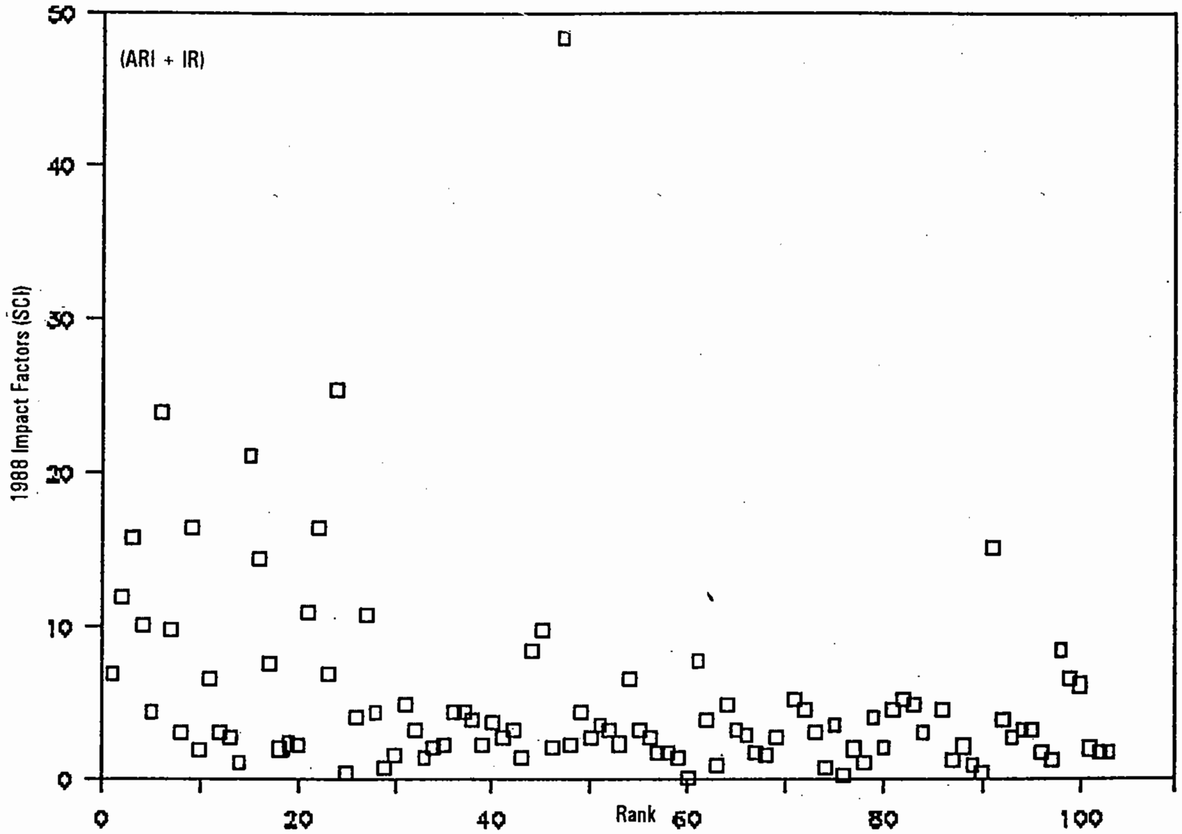


Fig. 3 : Correlations : Ranks and Impact Factors

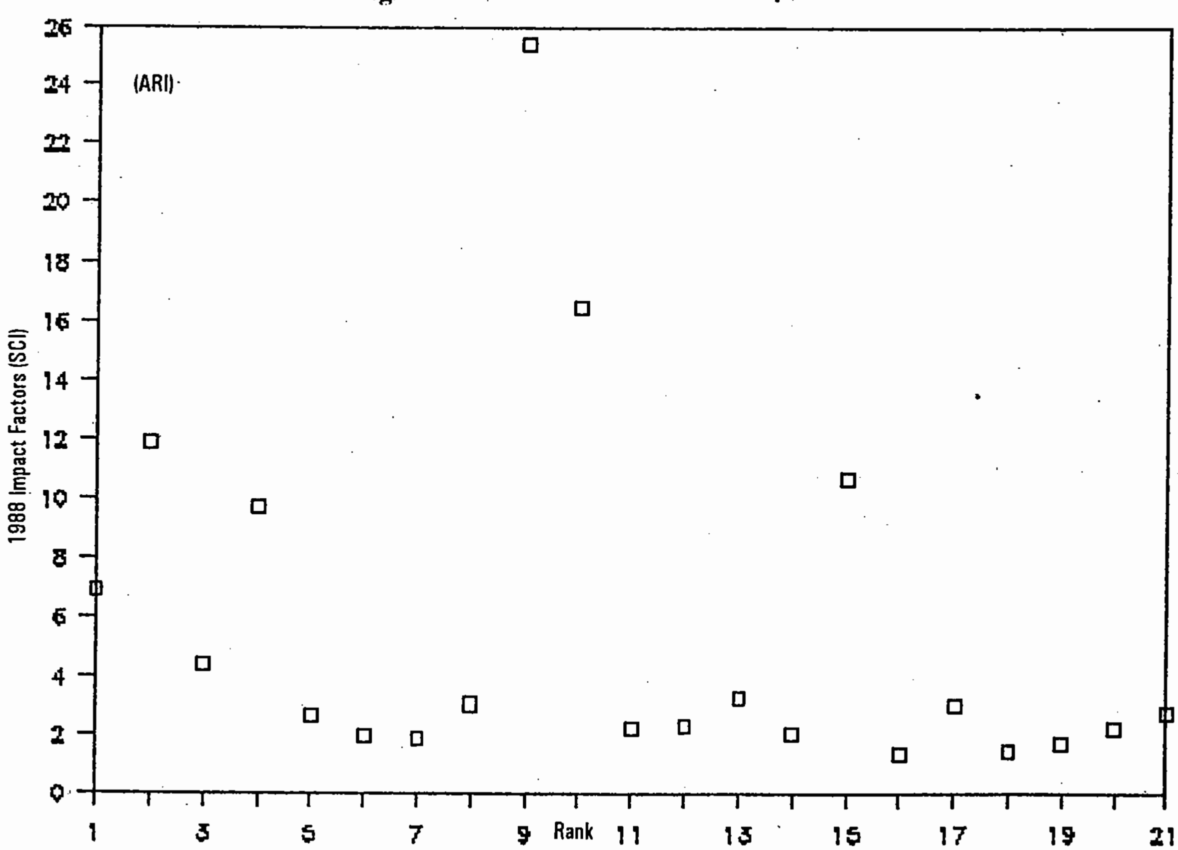


Fig. 4 : Correlations : Ranks and Impact Factors

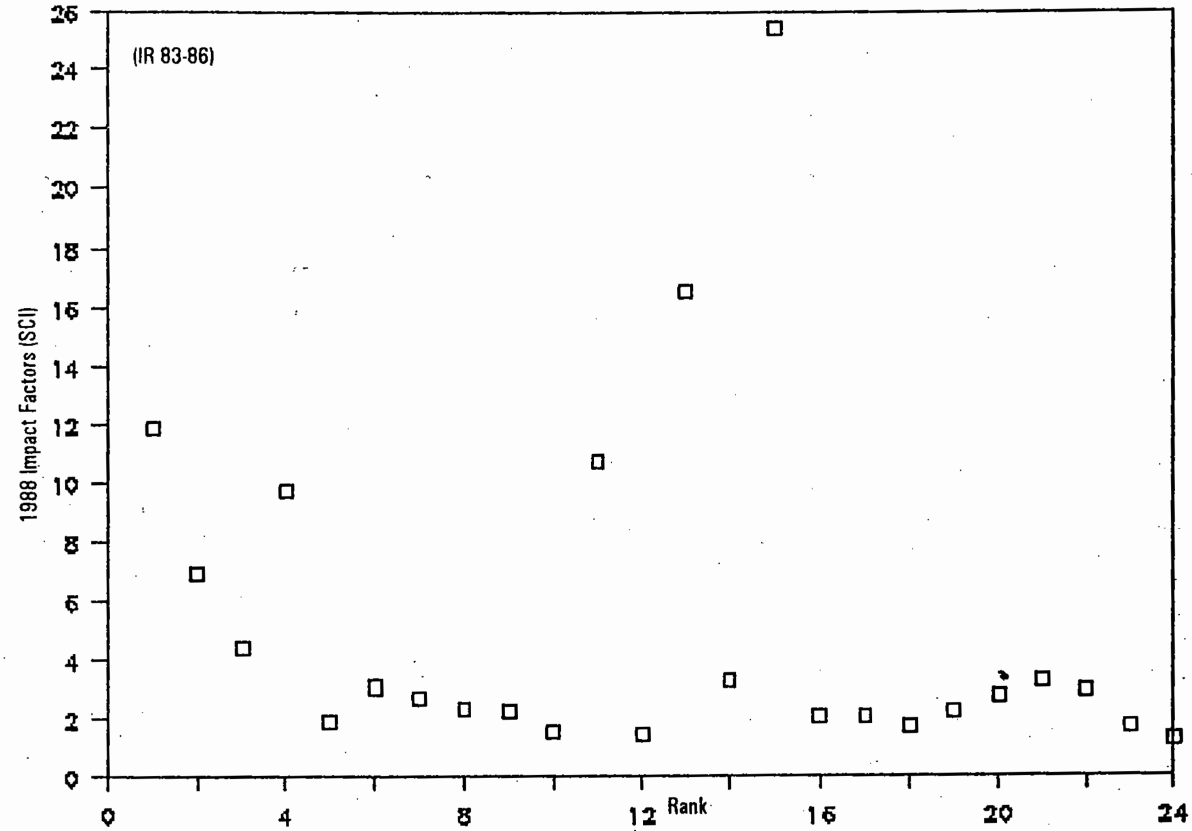


Fig. 5 : Correlations : Ranks and Impact Factors

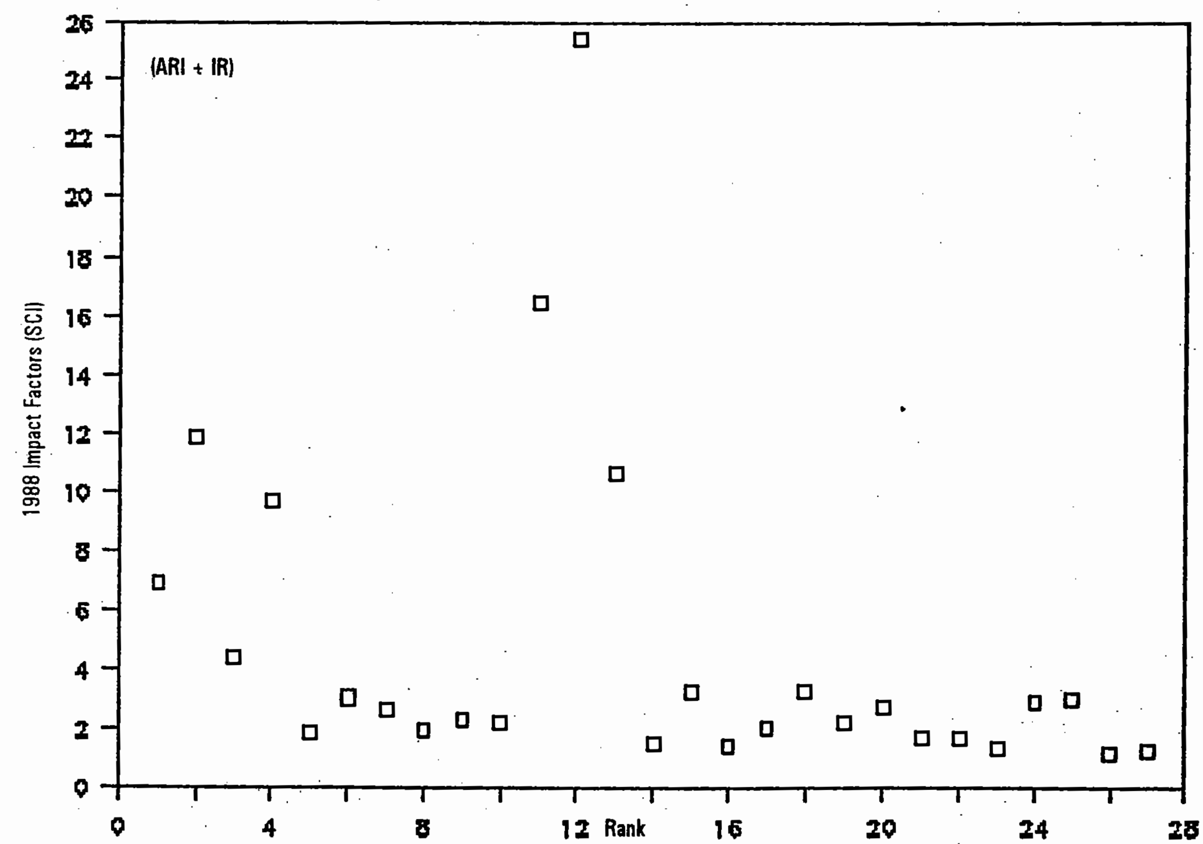


Fig. 6 : Correlations : Ranks and Impact Factors

Core Journals in Immunology
A Comparative Analysis of Bibliometric Parameters

Jagdish Arora
Library, Indian Institute of Technology, New Delhi-110 016
Usha Pawan
Associate Professor, Deptt. of Lib. & Information Science, University of Rajasthan, Jaipur (India)

The article conducts bibliometric studies on core journals in immunology derived from citations collected from two reviewing journals for four consecutive years. The bibliometric studies, conducted on the three independent sets of ranked lists of core journals, show a very high degree of correlation amongst parameters analyzed. The first four core journals, common to all the three ranked lists, contribute close to 45% of total citations and are essentially in the same rank order. The article determines country, subject, physical format, chronological and language-wise distribution of core journals for three sets of data. The half life and citations peak of journals in immunology have also been worked out. The scattering of literature in immunology is evaluated by applying Bradford's law. Lastly, the article compares the core journals in immunology with that of other discipline of biomedical sciences.

1. Introduction

Citations in a written document are empirical in nature and reflect the actual practices of the authors. Hence analysis of a wide range of citation would reveal the reading pattern of a wide range of citation would reveal the reading pattern of a community or a branch of it, depending upon the scope of the study. It is specially applicable to the periodicals because of continuity in their publication pattern. This article determines two separate sets of ranked lists of core journals in immunology by bibliometric analyses of a large number of citations collected from the Annual Review of Immunology and Immunological Reviews published from 1983 to 1986. Finally a unified ranked list of core journals in immunology is proposed, based on the combined citations from the two reviewing journals. The three ranked lists of core journals thus obtained based on the citations collected from the two reviewing journals for corresponding period were subjected to the similar bibliometric analysis essentially to examine the correlation between the three ranked lists thus evolved. This data may be used in adopting a ranked list of periodicals in immunology based on the quantitative performance of journals in terms of totality of citations obtained from the two reviewing journals. Thus qualitative journals are determined on the basis of their quantitative performance. The country, subject coverage, form and language-wise distribution for three sets of core periodicals thus obtained have also been done. It also evaluate scattering or concentration of literature in immunology by applying **Bradford's law of scattering / Garfield's law of concentration** to the citations collected for three sets of data. The article also compare the core journals in immunology with core journals in other discipline of biomedical sciences. The ranked lists of core periodicals developed from combined citations from the two reviewing journals provide most objective and unbiased data regarding the performance of periodicals and thus may be put to practical use for assessing the effectiveness of a periodical collection.

2. Creation of Citation Base

The citation for bibliometric analysis can be derived from any document or set of documents which provide a list of referred documents at the end of each article / chapter. Most of the bibliometric studies, however, derive their raw data for analysis from a set of primary journals of a chosen subject discipline. Bibliometric study of any subject based on the citations collected from selected primary journals, suffer from a number of bias which are characteristic to the items referred in primary media of communication. Some of the typical problem relate to self-citations, implicit citations, limit in the time period covered and the citations given for reasons other than professional. Further, since the process of bibliometric analysis is labour-intensive number of source journals and citations collected from them are usually insufficient. Moreover, selection of periodicals for bibliometric analysis based on the country and the languages covered by them, would effect the inferences drawn.

On the contrary, the citations given in a review article do not suffer from the problems that are characteristic to the citations obtained from primary journals. A review article is a transient miniature representing literature published in a given subject. A review provides synthesis of the proliferating fragmented knowledge appearing in a plethora of publications. It carries a highly analytical, organized and complete synthesis of information in a given area. The process of reviewing involves great amount of time and effort on the part of experts in summarizing consensus of scientific opinion on a given topic in order to simplify the task of documenting earlier research. Experts, to support the credibility of their views, cite highest quality and the most relevant literature. Moreover, because of constraints of space, the number of citation is severely limited and redundancy of citations is minimized in a review article.

While all review articles are not highly cited, the relatively high impact of review journals is well known. Journal Citation Report (1988) lists 30 reviewing serials out of the top 50 high impact serials. The reviews are used primarily as a device by researchers to help orient themselves to a new area. But for the reviews, it is felt, it is impossible for a scientist in most specialities to keep up with the primary literature in that field. The reviews play an indispensable role in connecting the individual scientist with the broader scientific culture. Citations from reviewing literature as source of citation for bibliometric analysis has thus been preferred to the citations from the primary journals since the former as source of citation,is known to have an edge over the later. Pao[11] in a study conducted in 1975 has used review articles as quality filters to identify recognized authorities in the field and has used the bibliographies of these review articles to identify relevant and influential journals and text books literature. Sengupta[12] elaborates factors favouring creation of a citation base for bibliometric analysis from citations given in the annual reviews.

The Annual Review of Immunlogy and the Immunological Reviews, the two leading reviewing journals in the field of immunology, have been selected to obtain citations for analysis in this study. The publication of Annual Review of Immunology commenced in 1983 by Annual Reviews, Inc., The Annual Reviews Inc. was founded in 1932 by James Murray Luck with successful introduction of Annual Review of Biochemistry. Today, Annual Reviews, Inc. publishes 27 different volumes covering such diverse topics as genetics, nutrition, material sciences and psychology. Annual Review of Computer Science launched in 1986, is the latest introduction from Annual Reviews, Inc. Immunological Reviews was launched in 1969 as Transplantation Reviews by Munksgaard Publshers, a leading publication house of medical journals. Immunological Reviews is published bimonthly and each issue is devoted to different aspect of research in immunology.

2.1. *Methodology*

Two sets of citation base was obtained by collecting citations from the end of each chapter of the Annual Review of Immunology (vol. 1-4) and the Immunological Reviews (vol. 69 to 94) respectively for four successive years i.e., 1983-

1986. The references given at the end of each chapter were xeroxed and each citation was then transferred in original (cut and paste) to 3" × 5" slips, eliminating the chances of error which would have occurred otherwise. Each individual citation was assigned a unique identification number to facilitate co-citation analysis. Identification of journal titles was done using 28th edition of Ulrich's International Periodical Directory (including Irregular Serials) and the World list of Scientific Peiodicals. MEDLINE, EMBASE and Compact Cambridge : Life Science Collection, databases, available on CD-ROM at the National Institute of Immunology and the Indian MEDLARS Centre was used to search for the incomplete / inadequate citations.

The citations collected from the two sources for four successive years (1983-86) totalled to 22, 753 consisting of 10830 citations from the Annual Review of Immunology and 11923 citations from the Immunological Reviews respectively. of 22,753 citations, 20833 were journal articles (10024 from the Annual Review of Immunology and 10809 from the Immunological Reviews) and 1920 were references to other forms of publication.

Two sets of core journals were obtained from the two separate sequences of citation bases by arranging them journal-wise in decreasing frequency of citation counts received by each journal. The two sequences of citation base were then merged and arranged journal-wise to adopt a single ranked list of periodicals according to their quantitative performance.

3. Bibliometric Analyses on Ranked Lists of Core Periodicals

3.1. Core Journals in Immunology : From the Citations collected from the Ann. Rev. Immunol., the Immunol. Rev. and Combined Citations from the two Reviewing Journals during 1983 - 1986

Citation analysis has its application as a device to measure the relative importance of journals. The greater the number of citations made to a journal the greater its relative importance in the discipline or specific branch of knowledge in which the study is undertaken. Table 1 and 2 (Appendix) furnish 66 core journals each in the field of immunology arranged in decreasing frequency of citations received by them in the Ann. Rev. of Immunol. and the Immunol. Rev. respectively for four consecutive years. Table 3 (Appendix) lists 103 core journals in immunology based on the combined citations from the two reviewing journals. In cases of ties, i.e. if the citations received by two or more journals are same, impact factor has been taken into consideration. The ranked lists of journals in immunology has deliberately been restricted to 66 journals each and 103 journals respectively for table 1,2 and 3. Remaining 359, 357 and 516 journals have not been included in the ranked lists of core periodicals thus developed as each of them had very insignificant number of citations to their credit. Figure 1 to 4 are the pictorial representation of comparative data for three lists of core journals.

3.2. J. Exp. Med. and J. Immunol. : Two Leading Journals in Immunology

It would be interesting to note the citation performance of these two journals in the SCI Database which is derived from a much larger citation base in comparison to citation base used in the present study. Garfield[4] in a study which compared citation patterns of the Journal of Experimental Medicine and the Journal of Immunology observed that no other journal seems to play as important role today in the transfer of information in immunology as does the Journal of Experimental Medicine. Inspite of generality of its title, the journal essentially publishes articles in immunology. The Journal of Experimental Medicine, he opined, is one of the most-cited journals in the world and it ranks highest in terms of its impact. In an other study, Garfield[7] analyzed the citation performance of 6,500 articles published in the Journal of Experimental Medicine from 1955 to 1985. The study reflected that 2,900 or 44% of total articles published in the Journal of Experimental Medicine during the period were cited over 50 times. Although, in comparison to the other journals, the Journal of Experimental Medicine has pub-

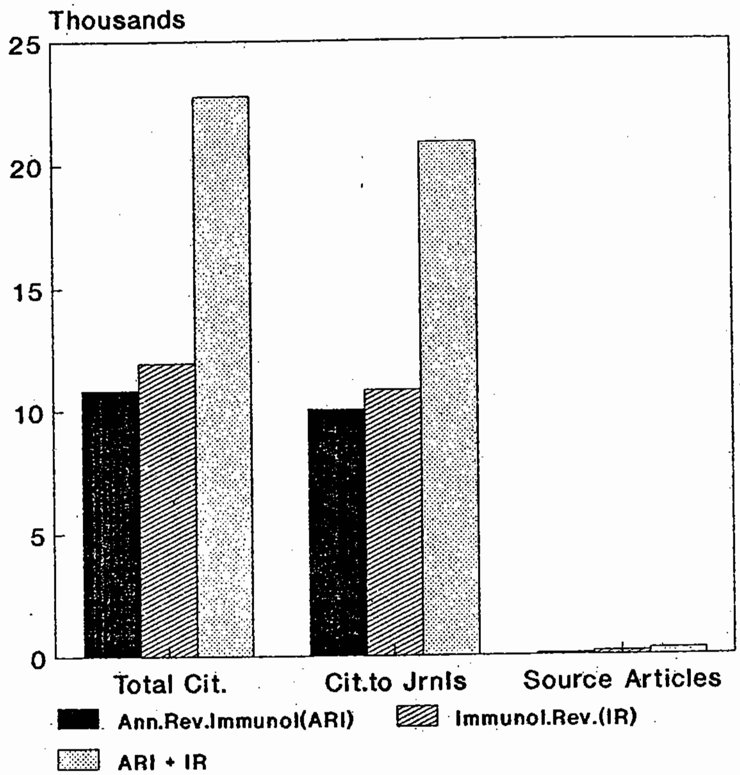


Fig. 1 : Comparative Data for Core Journals In Immunology

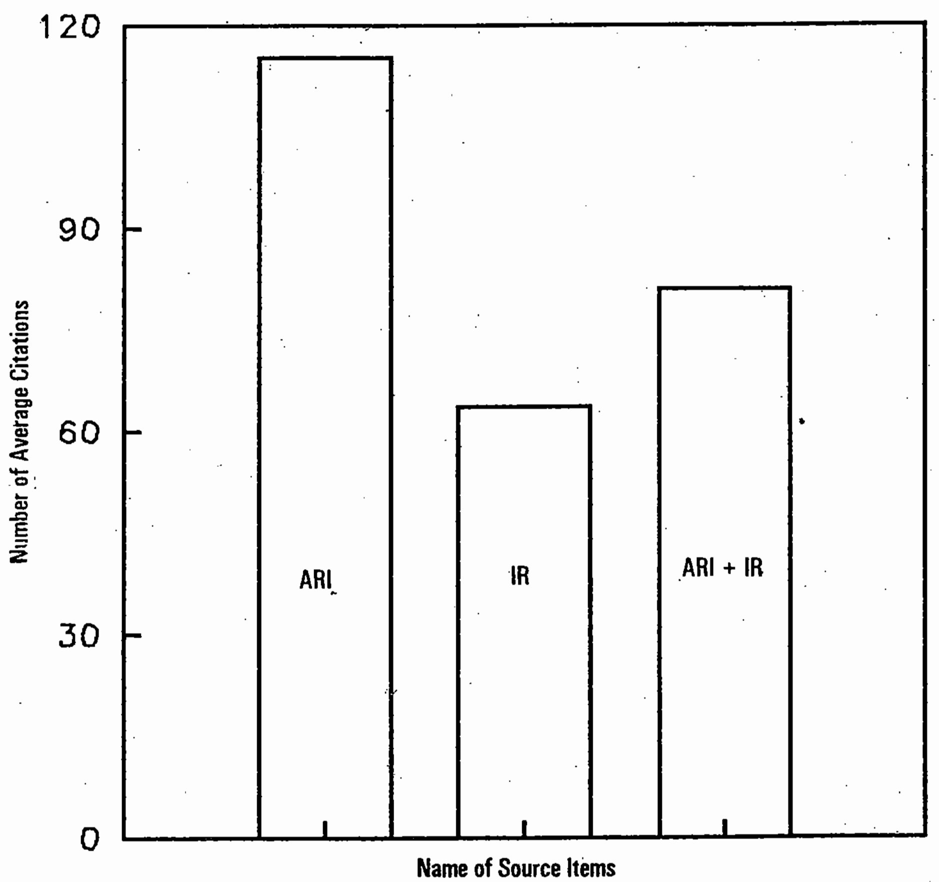


Fig. 2 : Average Citations

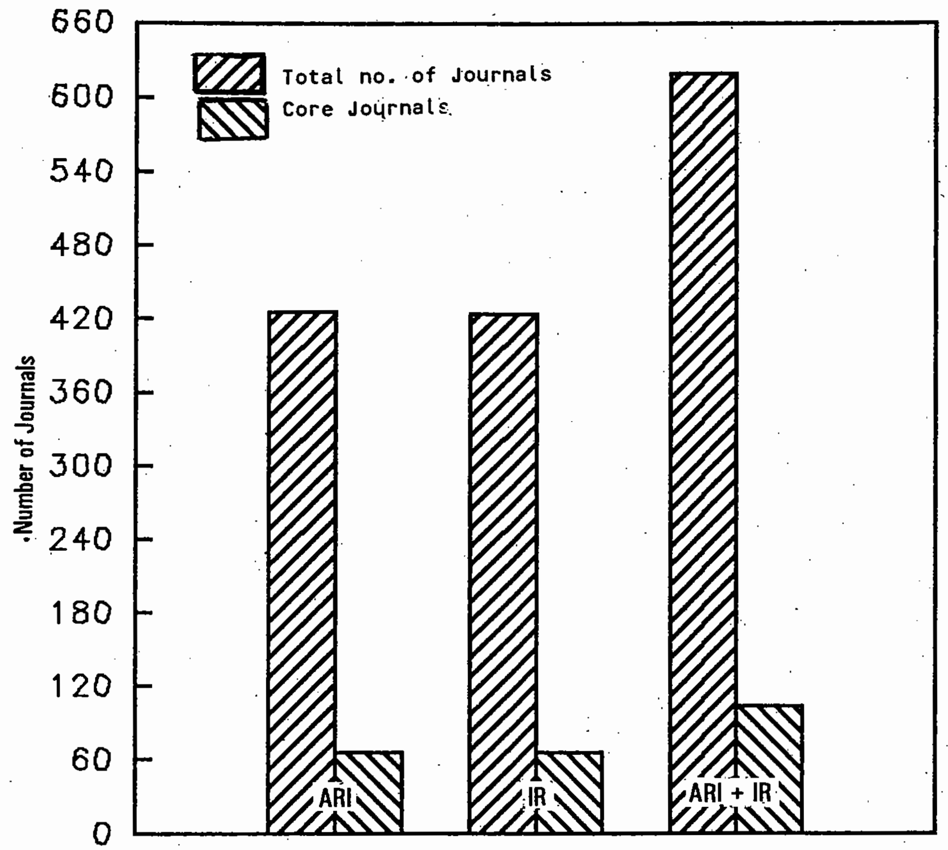


Fig. 3 : Total No. of Journals Vs Core Journals

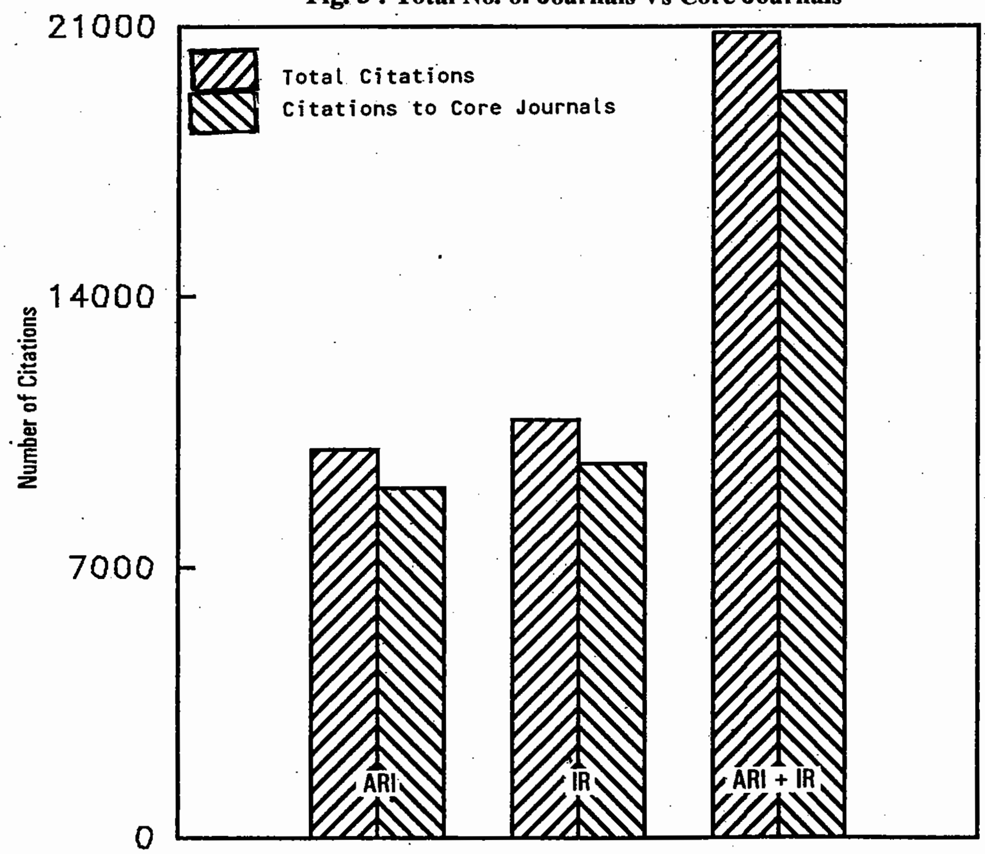


Fig. 4 : Total Citations Vs No. of Citations to Core Journals

lished far fewer articles, nevertheless its citation performance makes it a journal extraordinaire. Further, as per the Journal Citation Report, based on analysis of citations from 4,398 journals and other publications available in ISI database, the Journal of Experimental Medicine was cited over 37,079 times while the Journal of Immunology was cited 64,358 times till 1988. Comparing the citation figure of the two journals in terms of their impact as analyzed in the Journal Citation Report[9], the Journal of Experimental Medicine ranks 22nd with an impact factor of 11.83 and the Journal of Immunology ranks 53rd with an impact factor of 6.9 amongst 4,398 journals.

In the present study, the Journal of Immunology and the Journal of Experimental Medicine are two highly cited journals that occupy two top slots in all the three ranked lists of periodicals. The two journals between themselves accounts for close to 30% of total citations to the core journals in immunology. The quantitative performance of two journals in terms of citations received by them in the three sets of core periodicals are given Figure 5 and 6.

3.3. *Subject-wise Distribution of Core Journals*

The subjects or disciplines are labels assigned to certain area of knowledge and do not constitute a standard unit of size. Classifying journals is a particularly complex problem as they cover a variety of different subjects and they change their subject coverage with time. An attempt has, however, been made to categorize core journals in immunology according to the broad subject categories they cover. Subject-wise distribution of journals in immunology for the three sets of core journals is shown in Figure 7.

A sizable representation of journals of medical sciences in the core journals of immunology explain the proximity of two subjects. Immunology continues to be an important discipline taught in medical sciences and has contributed substantially in combating a number of fatal diseases of epidemic nature through vaccines which have tangible application in medical sciences. SCI's Journal Citation Report (1988) reveals substantial amount of citations by clinical journals to the Journal of Experimental Medicine and the Journal of Immunology as well as other journals on immunology which illustrates role and impact, basic research has in clinical applications. Representation of journals from various other discipline of life sciences in the core list of journals in immunology ascertain the fact that immunology is an eclectic science drawing on biochemistry, cellular and molecular biology, medicine, genetics, etc. Further, several top-ranked journals in all the three ranked list of core journals are exclusively devoted to general or applied immunology which goes on to prove the well established status of immunology in scientific research. Publication of articles in general science journals may be interpreted as indication of present role of immunology in the advancement of fundamental sciences.

3.4. *Language-wise Distribution of Core Journals in Immunology*

Language-wise distribution of literature provides major vehicles for dissemination in the subject put to the analysis. The data provide the scholars with an idea of the language he has to master, the librarian with the guidelines to strengthen his collections in certain languages, the publisher of secondary services with a guide to coverage policies, and the publisher of translating services with the necessary knowledge to detect areas that might benefit from the introduction of translation facilities.

Earlier studies of periodical production in sciences have shown that the English, Russian, German, French and Japanese dominate the publication scenario and usually in that order too. The findings of the present study, however, shows that close to 96% of core journals in immunology are published in English language while the remaining 4% are multilingual, i.e. they publish in English, French, German and Spanish. This shows that the English is the lingua-franca of communication for immunologists. Table 4 provide language-wise distribution of core journals in immunology for three sets of core journals. (Shown in Figure 8).

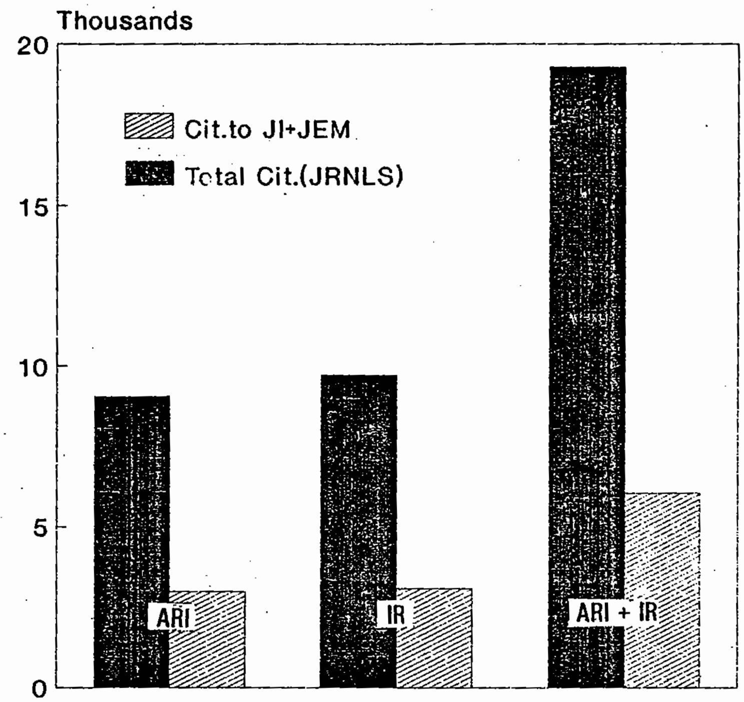


Fig. 5 : Comparative Performance of J. Immunol. & J. Exp. Med.

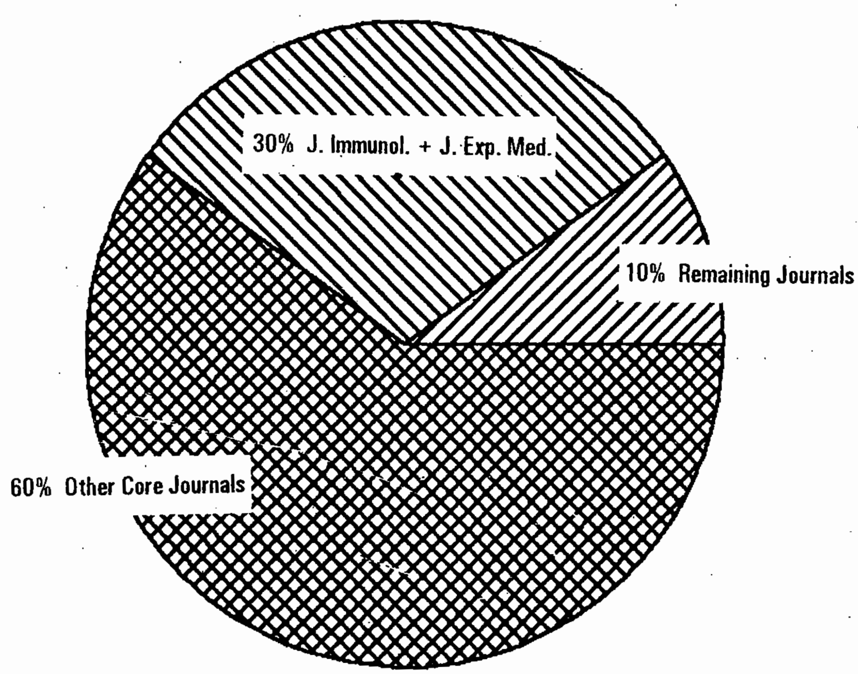


Fig. 6 : Citation Frequency of J. Immunol. and J. Exp. Med.
In Comparison to other Journals in Three Lists of Core Journals

3.5. *Country-wise Distribution of Core Journals in Immunology*

A geographic analysis of citations furnish information on the range of countries active in a subject field and the relative contribution each country makes. Bibliometrically, the value of place of publication rests very much with its ability to point to the "nationality" of a document. Place of publication is the best guide available to nationality of a document. The number of documents that any one country produces may be a pointer to its importance in the publishing industry and sphere, the financial health of publishing industry and productivity of its authors. A geographic analysis furnishes information on the range of countries active in a subject field and the relative contribution each of them makes. From this information an impression of "scatter" can be obtained, i.e. the extent to which subject publishing is the preserve of a few high volume publishers or the product of a large number of smaller publishers. Figure 9 depicts country-wise distribution of core journals in immunology for the three sets of core journals. While USA takes the top slot for all the three sets of ranked lists of core journals with its contribution hovering around 65%, the UK is second on the lists of countries contributing core journals in immunology with its contribution close to 20% in all the three cases. Remaining countries contribute insignificant number of journals ranging from 1 to 4.

3.6. *Bibliographical Forms of Cited Documents*

Each document type or form has a specific role to play in dissemination of information and it is this role that dictates variously the nature, level and currency of information carried, as well as physical format and the frequency with which it is published. A preponderance of journals suggests that the subject has a high turnover of ideas, a number of continuing developments requires current news and information and that results of research or study can be best accommodated by the article type of format of the journals.

More than 90% of total citations in all the three sets of data are to articles published in journals. Citations to books and monographs is around 4.5% for all the three sets of data. Citations to published proceedings vary from 0.905% to 2.130% for the three sets of data. Citations to irregular serials is around 1.2% and more for three sets of data. Papers in preparation, unpublished observation, etc. has its representation with around 0.7% citations for all the three sets of data. Citations to dissertation vary from 0.102% to 0.243%, while citations to reports is around 0.13% for all the three sets of data. Figure 10 depicts the distribution of bibliographical forms of cited documents for three sets of data obtained from the Ann. Rev. Immunol., the Immunol. Rev. and the combined citations from both the reviewing journals for four consecutive years, i.e. 1983 - 1986.!

Table 4

Language-wise Distribution of Core Journals in Immunology

S.No.	Source of Citations	English	%	Multi-lingual	%
1.	Ann. Rev. Immunol. (ARI)	61	92.42	5	7.58
2.	Immunol. Rev. (IR)	64	96.97	2	3.03
3.	Ann. Rev. Immunol. and Immunol. Rev. (ARI + IR)	99	96.12	4	3.88

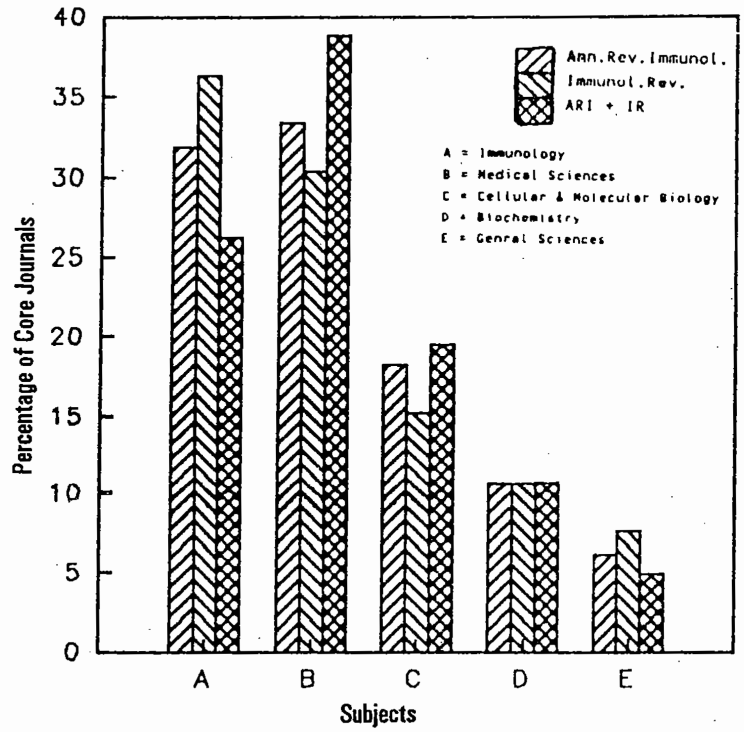


Fig. 7 : Subjectwise Distribution of Core Journals in Immunology

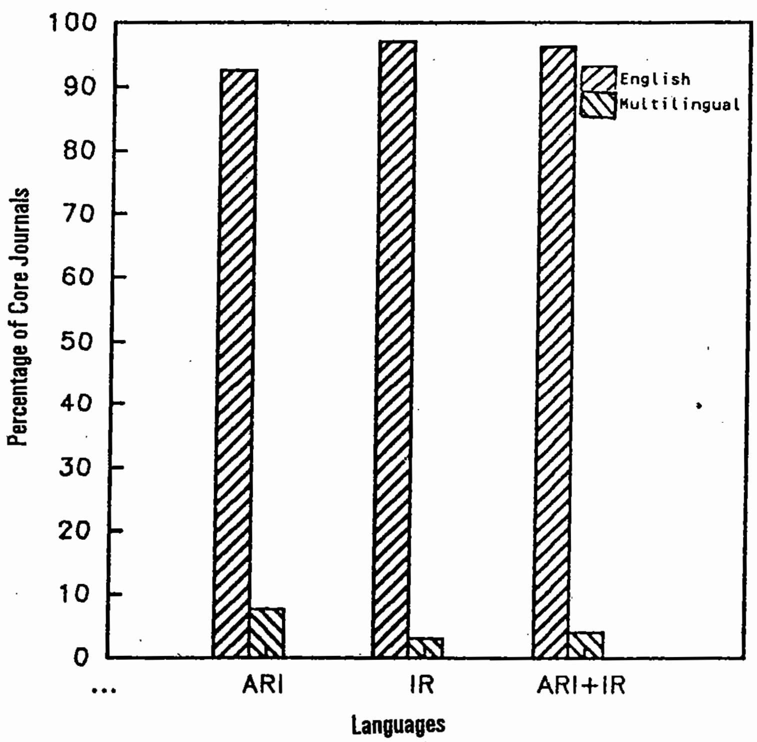


Fig. 8 : Languagewise Distribution of Core Journals in Immunology

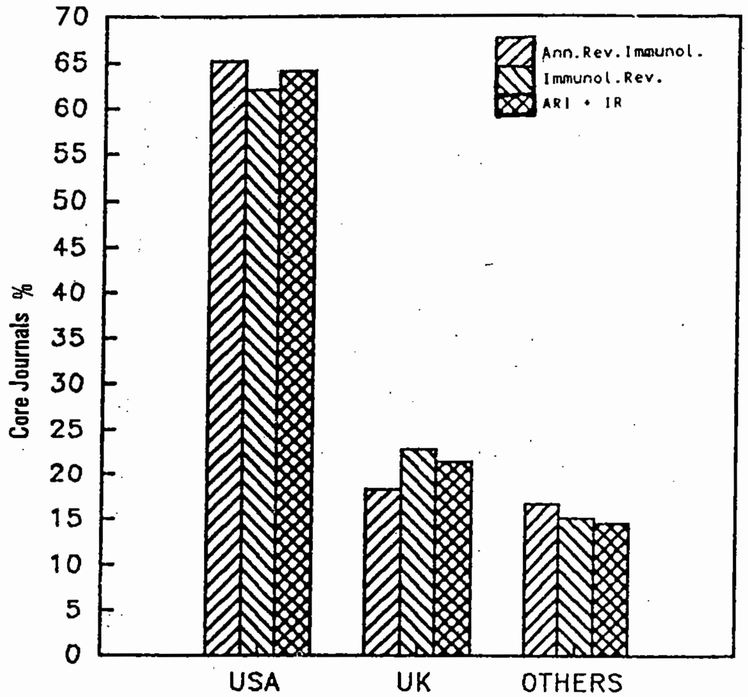


Fig. 9 : Countrywise Distribution of Core Journals in Immunology

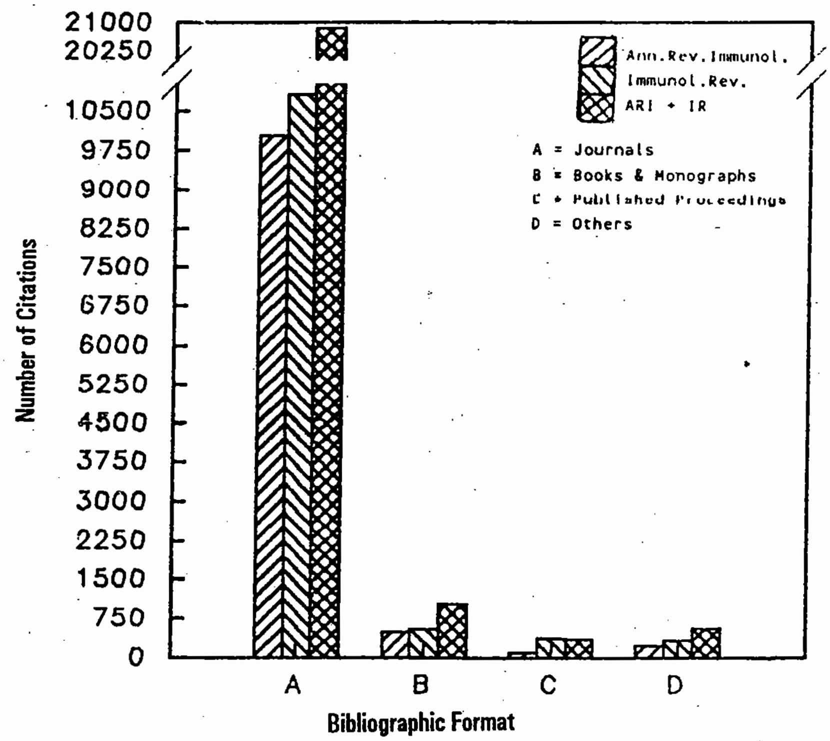


Fig. 10 : Formwise Distribution in Core Journals of Immunology

3.7. Year-wise Distribution of Citations and Ageing of Journal Collection

A number of studies have revealed that generally the recent volumes of a current journal seem to be used more often than the older ones. Several studies have been carried out to measure the use of back volumes of periodicals in order to optimize holdings of back volumes of periodicals or for discarding the old files of journals which were used sparingly in the past. The pioneering work on obsolescence or half-life of scientific literature by Line and Sandison[10] outline its relevance to the librarianship. Burton and Kebler[3] defined the half-life of periodical literature as "the time during which half of all currently active literature was published. On the basis of this definition, the half-life of journals in immunology was determined to be less than six years since more than half of the total citations (i.e. 11391 or 54.68%) belonged to 1981 - 1987 period. The year-wise distribution of citations is given in Table 5 in descending chronological order.. The data makes it obvious that the older research paper received fewer citations than the

Table 5 : Year-wise Distribution of Citation in Immunology (Combined Citations from Ann. Rev. Immunol. & Immunol. Rev., 1983-1986)

PERIOD (in years) (1)	YEAR (2)	CITATIONS RECEIVED (3)	% OF COL. 3 (4)	CUMULATIVE CITATIONS (5)	% OF COL. 5 (6)
1	1987	5	0.024	5	0.024
2	1986	132	0.634	137	0.658
3	1985	1295	6.216	1432	6.874
4	1984	2168	10.407	3600	17.280
5	1983	2721	13.061	6321	30.341
6	1982	2757	13.234	9078	43.575
7	1981	2317	11.122	11395	54.697
8	1980	1868	8.967	13263	63.663
9	1979	1323	6.351	14586	70.014
10	1978	1057	5.074	15643	75.088
11	1977	897	4.306	16540	79.393
12	1976	779	3.739	17319	83.133
13	1975	561	2.693	17880	85.825
14	1974	435	2.088	18315	87.913
15	1973	343	1.646	18658	89.560
16	1972	331	1.589	18989	91.149
17	1971	302	1.450	19291	92.598
18	1970	219	1.051	19510	93.649
19-28	1969-60	853	4.094	20363	97.744
29-38	1959-50	314	1.507	20677	99.251
39-48	1949-40	57	0.274	20734	99.525
49-58	1939-30	52	0.250	20786	99.774
59-68	1929-20	21	1.101	20807	99.875
69-78	1919-10	15	0.072	20822	99.947
79-88	1900-09	6	0.029	20828	99.976
89-98	1898-75	5	0.024	20833	100.000

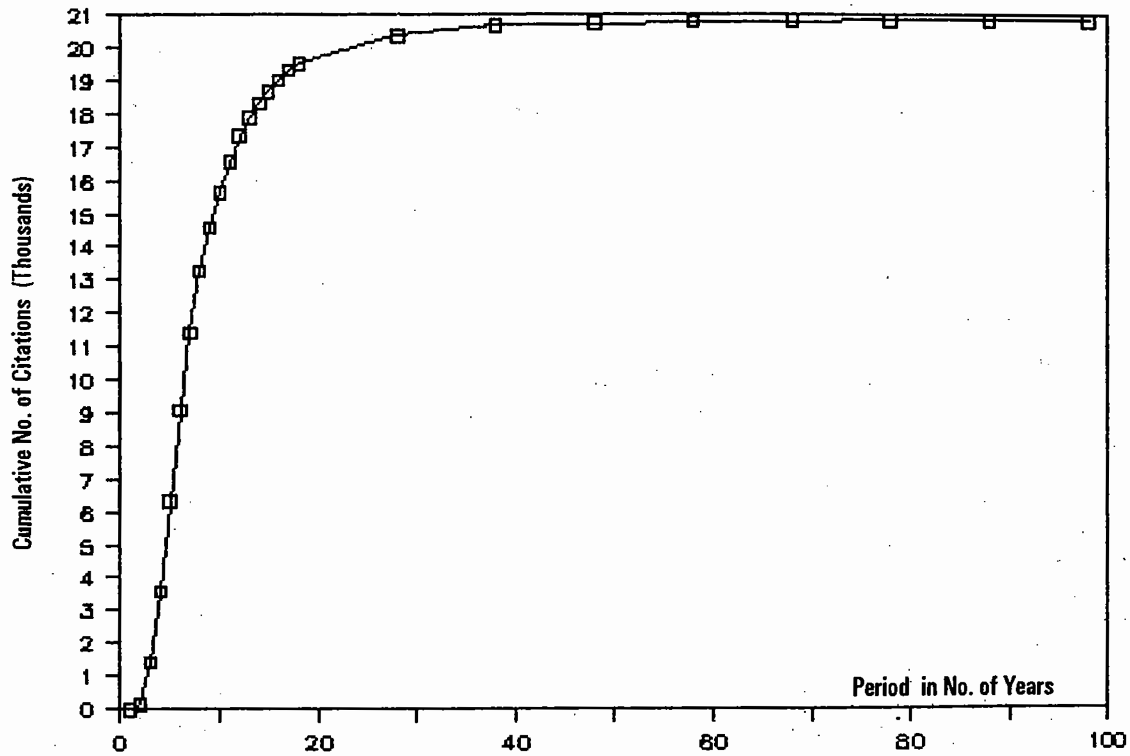


Fig. 11 : Year—Wise Distribution of Citations in Immunology

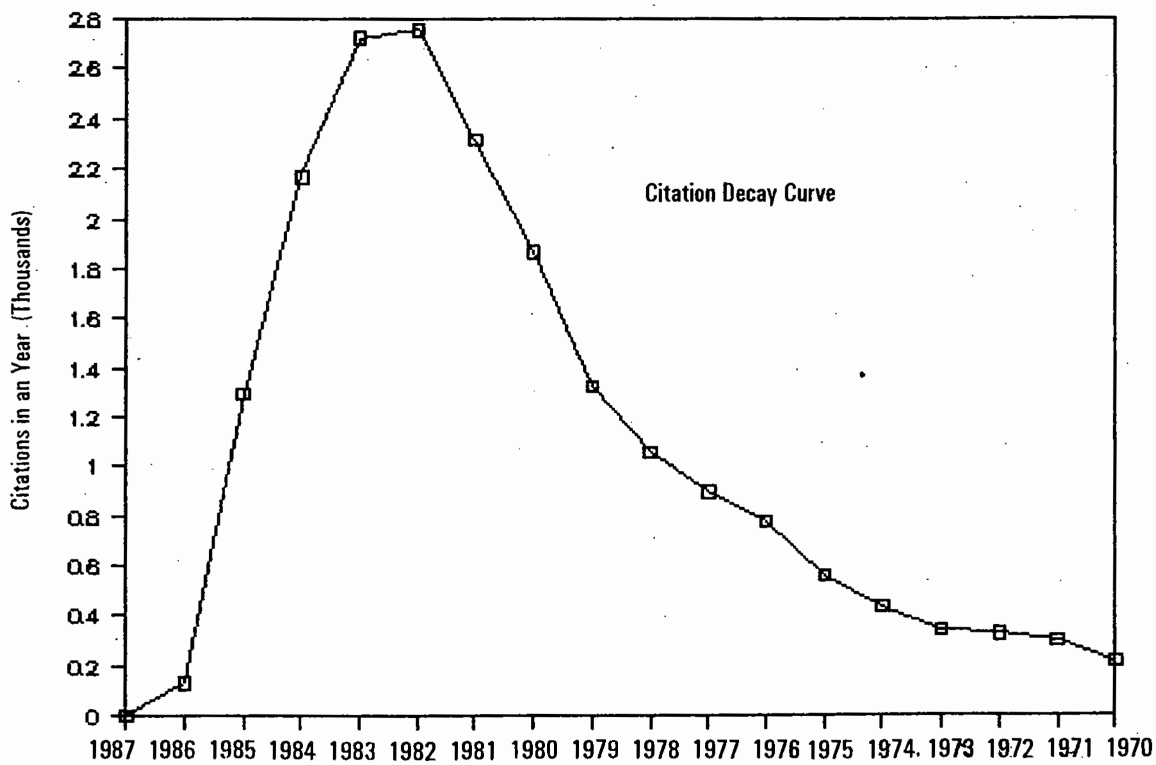


Fig. 12 : Year—Wise Distribution of Citations in Immunology

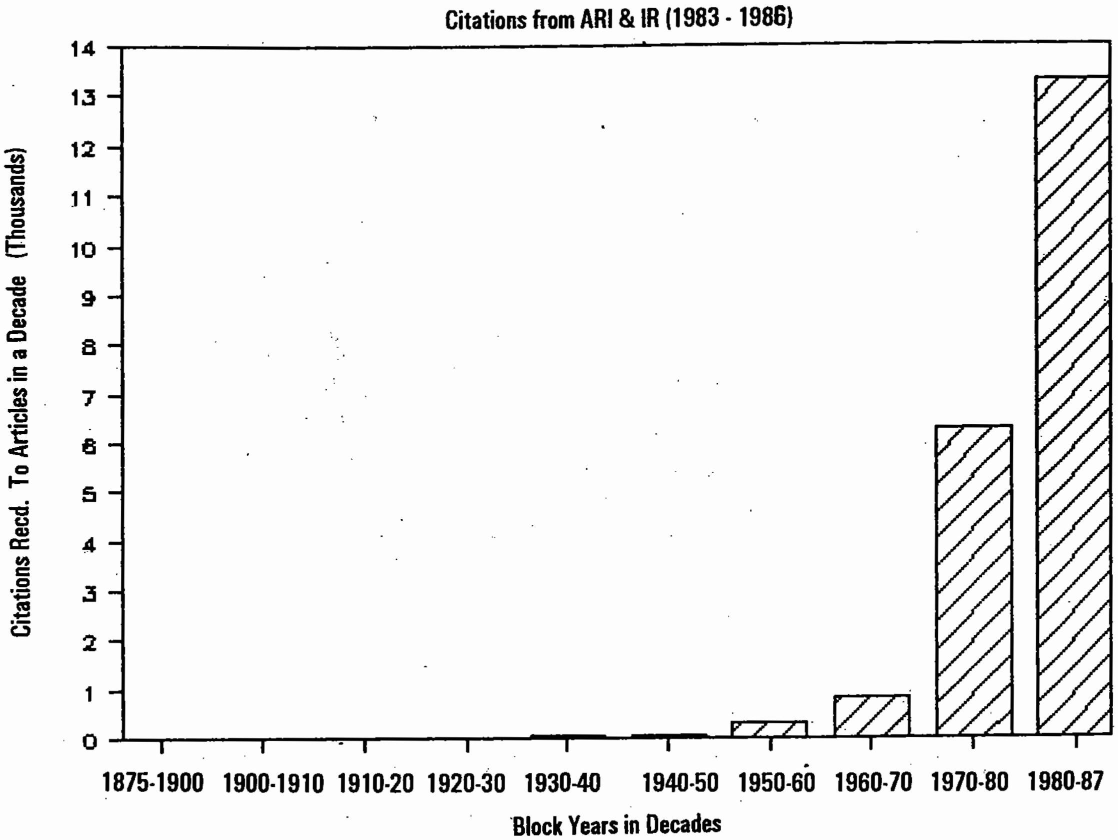


Fig. 13 : Year—Wise Growth in Rate of Citations

recent ones. The year-wise distribution of citations, done only on combined citation base consisting of 20833 citations to research articles obtained from the two reviewing journals, vary from a maximum of 2757 (13.23% in 1982) followed by 2721 (13.06%) and 2317 (11.12%) in 1983 and 1981 respectively to minimum of one citation to a number of research papers published during 19th century or during first three decades of present century. The oldest cited article was published in 1874 and the citations to research articles published from 1874 to 1949 was between 1 and 12. A gradual and steady increase in citations to the articles published from 1950 onwards is quite evident from the Table. Figure 11, 12 and 13 depict year-wise distribution of citations in immunology on the basis of combined citations from the two reviewing journals.

4. Testifying Bradford's Law of Scattering of Literature

Bradford's[1] empirical law of concentration for articles in the scientific periodicals proved to be a landmark event in the field of bibliometrics. Bradford's law states that the articles on a given subject concentrate heavily in a relatively small core of highly productive journals.

This law provides a very convenient base for estimating the size of a subject literature and a means of estimating how many journals must be checked to obtain a specified degree of completeness. A general form of this law says that the cumulative number of papers on a given subject (R (n)) will be related to the n journals in which they appear :

$$R(n) = R(1) + K. \log (n)$$

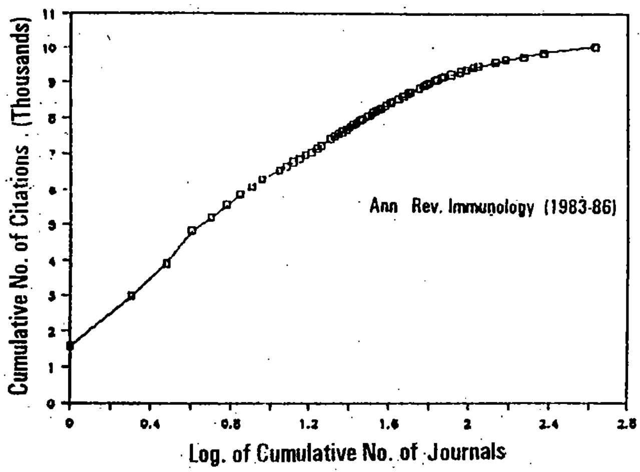


Fig. 14 : Bradford's Bibliograph

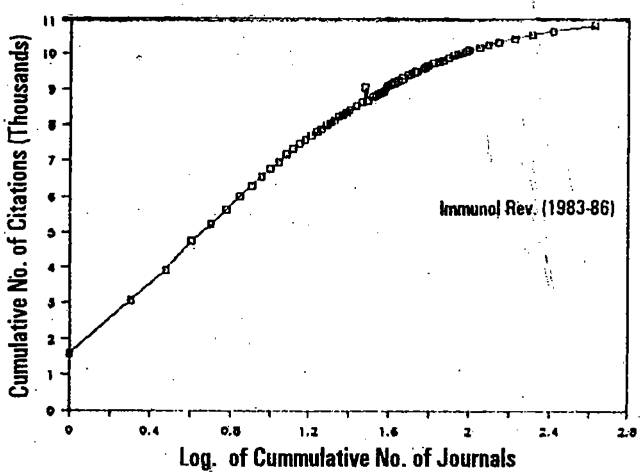


Fig. 15 : Bradford's Bibliograph

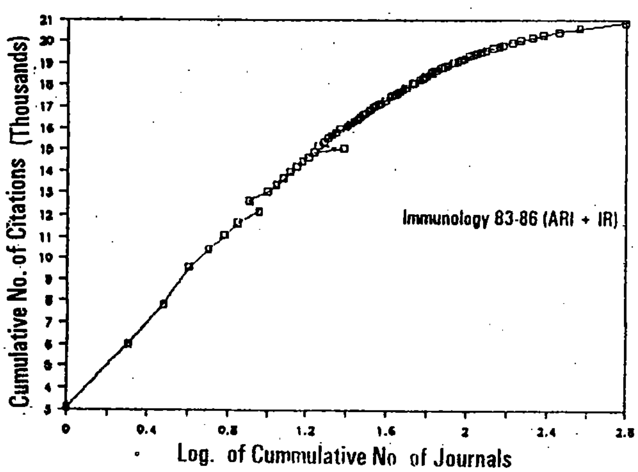


Fig. 16 : Bradford's Bibliograph

Journals were arranged in decreasing order of frequency of citations received from the Ann. Rev. Immunol., the Immunol. Rev. and the combined citation from both the reviewing journals respectively. 425 and 423 journals referred to in the Ann. Rev. of Immunol. (vol. 1 - 4) and the Immunol. Rev. (vol. 69 - 94) yielded 10,024 and 10809 citations respectively while 619 journals referred too in the combined citation base from the two reviewing journals yielded 20833 citations. Bradford's law of scattering is testified by plotting a graph by taking cumulative number of citations (R (n) on the ordinate against the logarithm of cumulative number of journals log(n) on the abscissa. It is observed that the resulting bibliograph in all the three cases, initially starts with the rising as in exponential nature and then follows into a linear curve indicating the observation of Bradford's law of scattering. The upward curving bottom of the bibliograph represents the small unclear zone of the most relevant journals. The upper end of the curve represent the peripheral zone where relevant references are widely scattered among a great number of journals. Experimental curve as plotted here in figure 14, 15 and 16 for the three sets of data confirm to the theoretical curve of Bradford's law.

It should be noted that the empirical data in figure 14, 15 and 16 begin to droop away from a straight line. This droop is characteristics of Bradford's plots and is called the **Groos Droop.** Groos[8] after whom the droop is named first noticed that the bibliograph eventually form a curve which droops below the linear prediction of the Bradford law. It can be interpreted either as a measure of incompleteness of the search if a complete bibliography on a subject is being testified, or as related to the infiniteness of the population of journals and papers. Since the empirical curve droops below the straight line, the empirical data indicate that fewer journals actually have one or more articles in immunology.

Comparative Study of Core Journals in Biomedical Sciences

Comparison between the core journals in immunology and in other disciplines of biomedical sciences is made to find out correlation amongst these disciplines. Sengupta[13] has evolved core lists of journals in various disciplines of biomedical sciences using Annual Reviews published in the respective disciplines. Table 6 shows citations collected from first 35 core journals in the field of immunology in comparison with that of biochemistry, microbiology, physiology, pharmacology and medicine evolved by Sengupta. The comparion reveals that 35 core journals in the respective discipline of biomedical sciences accounts for

Table 6 : Citation Received by First 35 Core Journals in Various Disciplines of Biomedical Sciences

Period covered	Discipline	Cumulative No. of Journals	Cumulative No. of Citations	No. of Citations in 35 Journals	% of Citations in 35 core Journals
1968-70	Biochemistry	533	16,276	13,221	81.25
1968-70	Microbiology	624	8,951	6,323	70.62
1968-70	Physiology	633	9,814	6,297	64.16
1968-70	Pharmacology	781	9,596	5,297	55.17
1965-69	Medicine	975	14,201	7,510	52.87
1983-86	Immunology (ARI)	425	10,024	8,272	82.52
1983-86	Immunology (IR)	423	10,809	8,878	81.97
1983-86	Immunology (ARI + IR)	619	20,833	17,024	81.72

Table 7 : Six Core Journals Common to Various Disciplines of Biomedical Sciences, showing Rank and % of the Citations Received by a Journal in a Discipline.

1. denotes core rank of a journal, and
2. denotes % of citation received in various disciplines

S.No.	Common core journals	Immunol. (ARI)		Immunol. (IR)		Immunol. (ARI + IR)		Med.		Biochem.		Microbiol.		Physiol.		Pharmacol.	
		1	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*
1.	Nature	3	9.36	3	7.86	3	8.59	7	2.18	6	4.52	4	5.05	5	3.01	2	4.26
2.	Proc. Nat. Acad. Sc.	4	8.92	4	7.72	4	8.31	33	0.66	2	8.44	2	6.92	28	0.66	8	0.85
3.	Science	7	2.89	12	1.83	8	2.34	15	1.48	9	3.28	10	2.42	8	2.48	5	2.55
4.	J. Biol. Chem.	9	2.24	20	0.79	11	1.49	28	0.91	1	13.48	6	3.94	21	0.88	10	1.75
5.	Fed. Proc.	21	0.79	35	0.41	25	0.59	11	1.52	12	1.98	30	0.57	11	2.09	9	2.21
6.	Ann. NY Acad. Sc.	30	0.45	26	0.53	29	0.49	17	1.34	28	0.44	25	0.74	23	0.80	8	2.26

81.72 (ARI + IR), 81.25, 70.62, 64.16, 55.17 and 52.87% of total citations respectively. Concentration of literature in the field of immunology in first 35 core journals (82.522%, 81.135% and 81.717%) in the core journals evolved from the citations collected from the Ann. Rev. Immunol. and the Immunol. Rev. and the combined citations from both the reviewing journals respectively and biochemistry (81.25%) is evident. This reveals that the proliferation of literature in these two disciplines is much higher in comparison to the other disciplines of biomedical sciences.

Further all the six disciplines of biomedical sciences share six journals given in Table 7 with core journals of immunology.

All the six journals with exception to J. Biol. Chem. belong to general sciences category. Further, four of the six journals are amongst first ten journals of the ranked lists in atleast five disciplines of medical sciences including immunology. Besides these six journals, there are ten more journals featuring amongst first thirty five core journals in more than three disciplines of biomedical sciences. That means there are sixteen journals out of thirty five core journals that are common to various well established disciplines of biomedical sciences. The significant overlap in the coverage of these journals is in confirmation to the Garfield's law of Concentration[5] which states that the tail of the literature of one discipline consists of a large part of the cores of the literature of the other disciplines.

6. Conlusion

Majority of signficant contribution to the world of science are known to be reported in a relatively small number of high impact journals in a given speciality. The citations, as indicators of scientific "quality" or "importance" can be used for measuring the impact of a scientist, or an article or an organization or a publishing journal on the scientific and technical development. For decades, librarians have used citation counts to determine the adequacy of a collection of periodicals, the optimum size of backfiles, binding and retention schedules and for cost-benefit analysis in the management of subscription budgets.

The citations collected from the reviewing literature have a definite edge ever the citations collected from other sources. The near-perfect correlation between the ranks of core periodicals evolved from the two reviewing journals and the similarity amongst various bibliometric parameters serve as validation to the selection of reviewing journals as source of citation in preference to the primary journals.

Ranking of periodicals based on the quantitative consideration as judged by the frequency of citation counts is a simple technique that can be used as a guidance to the acquisition policy of periodicals. Ranked list of periodicals evolved from the citation base consisting of 20833 references has 103 periodical titles accounting for 19306 (92.97%) citations.

The J. Exp. Med. (Rank #2) and J. Immunol. (Rank #1) are the two journals exclusively devoted to the publication of research articles in immunology. The two journals between themselves contribute close to 30% of the citations in all the three ranked lists. The Nature and the Proc. Nat. Acad. Sci. are the other two journals that rank 3rd and 4th respectively and are common to all the three lists. The four journals amongst themselves contribute close to 45% of citations. English is lingua-franca of immunologists. Nearly 96% of the literature in immunology is published in English and remaining 4 % in journals that are multilingual.

The citation data collected for this study obey the Bradford's law of scattering implying that the law is not only applicable on the complete literature of a subject but also on selected population of citations. Half-life of the journals in immunology was determined to be less than six years. This fact may be considered while buying or maintaining the backfiles of journals.

References

1. Bradford, S. C. Sources of information on specific subjects. Engineering, 137, 85 - 86, 1934.
2. Bradford, S.C. Documentation. London, Crosby Lockwood & Son Ltd., 1948.
3. Burton, R.E. and R.W. Kebler,. The half-life of some scientific and technical literature. Amer. Docum., 11 (1), 18 - 20, 1960
4. Garfield, E. Journal citations studies. III. Journal of Experimental Medicine compared with Journal of Immunology, or how much of a clinician is the immunologists. (*In* : The Essay of an Information Scientists. Vol. 1. California, ISI, 1973) p. 326-329.
5. Garfield, E. The mystery of the transposed journal lists : Wherein Bradford's law of scattering is generalized according to Garfield's law of concentration. (*In* : Essays of an Information Scientists, Vol. 1. Philadelphia, ISI, 1977,) p. 222-223.
6. Garfield, E. The Science Citations Index : As quality information filter. In : Coping with the biomedical literature explosion; A qualitative approach. New York, Rockefeller Foundation, 1978, p. 76.
7. Garfield, E. Ninety-one citation classics from the Journal of Experimental Medicine. Current Contents, (28), 3-13, 1987.
8. Groos, O.V. Bradford's law and the Keenan-Atherton data. Amer. Docum., 18, 46, 1967.
9. Journal Citation Report 1988. Philadelphia, ISI, 1988.
10. Line, M.B. and A. Sandison, Practical interpretation of citation and library use studies. Coll. Res. Libr., 36(5), 393-396, 1975.
11. Pao, M.L. A quality filtering system for medical literature. J. Med. Educ., 50, 353-359, 1975.
12. Sengupta, I.N. A new bibliometric technique for quantitative evaluation of core journals of basic biomedical sciences. (*In* : The Library and Information Science Horizon (ed. B. Guha). New Delhi, Allied, 1986,) p. 319-368.
13. Ibid.

APPENDIX : TABLE 1-3

Table 1 : Ranked List of Periodicals in Immunology in the Decreasing Order of Citation (Annual Revieiws of Immunology 1983—1986)

Rank No. (1)	Name of Journal (2)	Year of Incept. (3)	Country (4)	Citat. Recd. (5)	% of Col. 5 (6)	Cum. Citat. (7)	% of Col. 7 (8)	1988 Impact Factor@ (9)
1	J. Immunol.	1916	USA	1582	15.78	1582	15.782	6.91
2	J. Exp. Med.	1896	USA	1393	13.90	2975	29.679	11.83
3	Nature	1869	UK	938	9.36	3913	39.036	15.76
4	Proc. Natl. Acad. Sci.	1915	USA	894	8.92	4807	47.955	10.03
5	Cell	1974	USA	382	3.81	5189	51.766	23.91
6	Europ. J. Immunol.	1970	FRG	359	3.58	5548	55.347	4.34
7	Science	1883	USA	290	2.89	5838	58.240	16.46
8	Immunol. Rev.*	1969	DENMARK	232	2.31	6070	60.555	9.67
9	J. Biol. Chem.	1905	USA	224	2.23	6294	62.789	6.49
10	J. Clin. Invest.	1924	USA	133	1.33	6427	64.116	7.59
11	Immunology	1958	UK	133	1.33	6560	65.443	2.66
12	Mol. Immunol. !!	1964	USA	133	1.33	6693	66.770	1.95
13	Cell. Immunol.	1970	USA	123	1.23	6816	67.997	1.81
14	N. England J. Med.	1812	USA	107	1.07	6923	69.064	21.15
15	Immunogenetics	1974	USA	98	0.98	7021	70.042	3.01
16	Biochemistry	1964	USA	95	0.95	7116	70.990	4.01
17	Transplantation	1963	USA	92	0.92	7208	71.907	2.98
18	Lancet	1823	UK	86	0.86	7294	72.765	14.48
19	Ann. Rev. Immunol.	1983	USA	81	0.81	7375	73.573	25.41
20	Adv. Immunol.	1961	USA	81	0.81	7456	74.381	16.40
21	Fed. Proc.	1942	USA	79	0.79	7535	75.170	0.33
22	Scand. J. Immunol.	1972	UK	71	0.71	7606	75.878	2.19
23	Clin. Exp. Immunol.	1966	UK	70	0.70	7676	76.576	2.28
24	Infect. Immun.	1970	USA	67	0.67	7743	77.245	3.21
25	EMBO J.	1982	UK	66	0.66	7809	77.903	10.94
26	Diabetes	1952	USA	62	0.62	7871	78.522	4.42
27	Nucleic Acids Res.	1974	UK	53	0.53	7924	79.050	4.30
28	Biochem. J.	1906	UK	47	0.47	7971	79.519	3.94
29	Cold Sprng. Harb. Symp. Quant. Biol.	1933	USA	46	0.46	8017	79.978	2.25
30	Arthr. Rheum.	1958	USA	45	0.45	8062	80.427	4.81
31	Ann. N. Y Acad. Sci.	1823	USA	45	0.45	8107	80.876	0.75
32	J. Cell Biol. $	1955	USA	43	0.43	8150	81.305	9.75
33	Biochem. Biophys. Res. Commun.	1959	USA	42	0.42	8192	81.724	3.17
34	Transplant. Proc.	1969	USA	41	0.41	8233	82.133	1.10
35	Biochim. Biophys. Acta	1947	HOLLAND	39	0.39	8272	82.522	2.28
36	J. Mol. Biol.	1959	UK	38	0.38	8310	82.901	6.56
37	Cancer Res.	1941	USA	37	0.37	8347	83.270	4.30
38	Blood	1946	USA	36	0.36	8383	83.629	6.85
39	Clin. Immunol. Immunopathol.	1972	USA	36	0.36	8419	83.988	1.97
40	Proc. Soc. Exp. Biol. Med.	1903	USA	33	0.33	8452	84.318	1.33
41	Ann. Rev. Biochem.	1931	USA	31	0.31	8483	84.627	48.31
42	Diabetelogia	1965	FRG	31	0.31	8514	84.936	3.81

Table 1 contd.

Rank No, (1)	Name of journal (2)	Year of incept. (3)	Country (4)	Citat. recd. (5)	% of Col. 5 (6)	Cum. citat. (7)	% of Col. 7 (8)	1988 Impact factor@ (9)
43	J.Natl. Cancer Inst.	1940	USA	31	0.31	8545	85.245	2.71
44	Immunol. Today	1980	UK	29	0.29	8574	85.535	10.65
45	Int. J. Cancer	1966	USA	29	0.29	8603	85.824	3.16
46	Am. J. Pathol. %	1924	USA	27	0.27	8630	86.093	4.36
47	Ann. Intern. Med.	1922	USA	26	0.26	8656	86.353	8.47
48	Am. Rev. Respiratory Dis. !	1917	USA	26	0.26	8682	86.612	4.48
49	Int. Arch. Allergy Appl. Immunol.	1950	SWITZERL	25	0.25	8707	86.862	1.31
50	Clin. Res.	1953	USA	23	0.23	8730	87.091	0.12
51	Mol. Cell. Biol.	1981	USA	22	0.22	8752	87.310	7.73
52	Prog. Allergy	1966	SWITZERL	22	0.22	8774	87.530	3.75
53	J. Allergy Clin. Immunol.	1929	USA	22	0.22	8796	87.749	2.95
54	Proc. Royal Soc. (London) B. Biol. Sc.	1905	UK	22	0.22	8818	87.969	2.23
55	Ann. Inst. Pasteur/Immunol. %%	1887	FRANCE	22	0.22	8840	88.188	1.38
56	J. Investigative Dermatol.	1938	USA	21	0.21	8861	88.398	3.47
57	Springer Sem. Immunopathol.	1978	FRG	21	0.21	8882	88.607	1.68
58	Nature-New Biol. **	1869	UK	21	0.21	8903	88.817	
59	Neurology	1951	USA	20	0.20	8923	89.016	2.97
60	Australian J. Exp. Biol. Med. Sci.	1924	AUSTRALI	20	0.20	8943	89.216	0.90
61	J. Infectious Dis.	1904	USA	19	0.19	8962	89.405	4.91
62	J. Immunol. Methods	1971	USA	19	0.19	8981	89.595	2.17
63	FEBS Letters	1968	HOLLAND	18	0.18	8999	89.775	3.57
64	Immunobiology	1909	FRG	18	0.18	9017	89.954	2.71
65	Prostaglandlins	1972	USA	18	0.18	9035	90.134	2.12
66	Arch. Dermatol.	1920	USA	18	0.18	9053	90.313	2.00

@ From Journals Citations Report
* Formerly Transplantation Review
! Formerly Am. Rev. Tuber. Pulmon. Dis.
$ Formerly J.Biochem. Biophys. Cytol.
!! Formerly Immunochemistry
** Ceased Publication
% Formerly J. Med. Res.
%% Renamed as Research in Immunology

Table 2 : Ranked List of Periodicals in Immunology in the Decreasing Order or Citation (Immunological Reviews 1983—1986)

Rank No. (1)	Name of Journal (2)	Year of Incept. (3)	Country (4)	Citat. Recd. (5)	% of Col. 5 (6)	Cum. Citat. (7)	% of Col. 7 (8)	1988 Impact Factor@ (9)
1	J. Exp. Med.	1896	USA	1592	14.728	1592	14.728	11.83
2	J. Immunol.	1916	USA	1473	13.628	3065	28.356	6.91
3	Nature	1869	UK	851	7.873	3916	36.229	15.76
4	Proc. Natl. Acad. Sci.	1915	USA	837	7.744	4753	43.973	10.03
5	Europ. J. Immunol.	1970	GERMANY	469	4.339	5222	48.312	4.34
6	Transplantation	1963	USA	425	3.932	5647	52.244	2.98
7	Immunol. Rev.*	1969	DENMARK	361	3.340	6008	55.583	9.67
8	Cell	1974	USA	288	2.664	6296	58.248	23.91
9	Cell. Immunol.	1970	USA	263	2.433	6559	60.681	1.81
10	Transplant. Proc.	1969	USA	210	1.943	6769	62.624	1.10
11	Immunogenetics	1974	USA	199	1.841	6968	64.465	3.01
12	Science	1883	USA	198	1.832	7166	66.297	16.46
13	Immunology	1958	UK	157	1.452	7323	67.749	2.66
14	Lancet	1823	UK	141	1.304	7464	69.054	14.48
15	New Engl. J. Med.	1812	USA	129	1.193	7593	70.247	21.15
16	Blood	1946	USA	102	0.944	7695	71.191	6.85
17	Clin. Exp. Immunol.	1966	UK	97	0.897	7792	72.088	2.28
18	Scand. J. Immunol.	1972	UK	95	0.879	7887	72.967	2.19
19	Tissue Antigens	1971	DENMARK	91	0.842	7978	73.809	1.52
20	J. Biol. Chem.	1905	USA	86	0.796	8064	74.604	6.49
21	Immunol. Today	1980	UK	84	0.777	8148	75.382	10.65
22	J. EMBO	1982	UK	81	0.749	8229	76.131	10.94
23	Ann. Inst. Pasteur (Immunol.) !!	1887	FRANCE	69	0.638	8298	76.769	1.38
24	Adv. Immunol.	1961	USA	65	0.601	8363	77.371	16.40
25	Human Immunol.	1980	USA	58	0.537	8421	77.907	3.22
26	J. Clin. Invest.	1924	USA	57	0.527	8478	78.435	7.59
27	Ann. NY. Acad. Sci. (USA)	1823	USA	57	0.527	8535	78.962	0.75
28	Ann. Rev. Immunol.	1983	USA	54	0.500	8589	79.462	25.41
29	Nucleic Acids Res.	1974	UK	54	0.500	8643	79.961	4.30
30	Mol. Immunol. !	1964	USA	51	0.472	8694	80.433	1.97
31	Clin. Immunol. Immunopathol.	1972	USA	49	0.453	8743	80.886	1.97
32	Arthr. Rheum.	1958	USA	48	0.444	8791	81.330	4.81
33	Ann. Rheum. Dis.	1939	UK	48	0.444	8839	81.774	2.01
34	Proc. Royal Society (UK) B. Biol. Sc.	1905	UK	46	0.426	8885	82.200	2.23
35	Fed. Proc.	1942	USA	44	0.407	8929	82.607	0.33
36	Prog. Allergy	1966	SWITZER	43	0.398	8972	83.005	3.75
37	American J. Pathol. %	1924	USA	42	0.389	9014	83.393	4.36
38	J. Reprod. Immunol.	1981	UK	40	0.370	9054	83.764	1.67
39	Exp. Hematol.	1973	USA	36	0.333	9090	84.097	2.71
40	J. Nat. Cancer Inst.	1940	USA	34	0.315	9124	84.411	2.71
41	J. Immunol. Meth.	1971	USA	31	0.287	9155	84.698	2.17

Table 3 contd.

Rank No. (1)	Name of Journal (2)	Year of Incept. (3)	Country (4)	Citat. Recd. (5)	% of Col. 5 (6)	Cum. Citat. (7)	% of Col. 7 (8)	1988 Impact Factor@ (9)
42	Human Immunology	1980	USA	59	17489	0.283	83.949	3.22
43	Proc. Soc. Exp. Biol. Med.	1903	USA	57	17546	0.274	84.222	1.33
44	Ann. Intern. Med.	1922	USA	54	17646	0.259	84.702	8.47
45	J. Cell Biol. $	1955	USA	53	17699	0.254	84.957	9.75
46	Ann. Rheum. Dis.	1939	UK	52	17751	0.250	85.206	2.01
47	Ann. Rev. Biochem.	1931	USA	50	17801	0.240	85.446	48.31
48	J. Immunol. Methods	1971	USA	50	17851	0.240	85.686	2.17
49	Cancer Res.	1941	USA	48	17899	0.230	85.917	4.30
50	Immunobiology	1909	FRG	47	17946	0.226	86.142	2.71
51	J. Investigative Dermatol.	1938	USA	46	17992	0.221	86.363	3.47
52	Biochem. Biophys. Res. Commun	1959	USA	46	18038	0.221	86.584	3.17
53	Biochem. Biophys. Acta	1947	NE	46	17592	0.221	84.443	2.28
54	J. Mol. Biol.	1959	UK	45	18083	0.216	86.800	6.56
55	Int. J. Cancer	1966	USA	44	18127	0.211	87.011	3.16
56	Exp. Hematol.	1973	USA	40	18238	0.192	87.544	2.71
57	Springer Sem. Immunopathol.	1978	USA	40	18278	0.192	87.736	1.68
58	J. Reprod. Immunol.	1980	UK	40	18198	0.192	87.352	1.67
59	Int. Arch. Allergy Appl. Immunol.	1950	SWITZER	39	18356	0.187	88.110	1.31
60	Nature-New Biol.~	1869	UK	39	18317	0.187	87.923	
61	Mol. Cell. Biol.	1981	USA	38	18394	0.182	88.293	7.73
62	Diabetologia	1965	USA	37	18431	0.178	88.470	3.81
63	Australian J. Exp. Biol. Med. Sci.	1924	AUSTRAL	35	18466	0.168	88.638	0.90
64	J. Infectious Dis.	1904	USA	33	18557	0.158	89.075	4.91
65	Br. Med. J.	1832	UK	33	18524	0.158	88.917	3.14
66	J. Mol. Cell Immunol.	1986	USA	33	18590	0.158	89.233	2.89
67	Acta Pathol. Microbiol. Scand.	1929	DENMARK	31	18158	0.149	87.160	1.67
68	J. Rheumatol.	1974	CANADA	29	18619	0.139	89.373	1.47
69	American J. Med.	1946	USA	28	18647	0.134	89.507	2.73
70	Phil. Trans. Royal Soc. B. Biol Sci.	1832	UK	28	18675	0.134	89.641	
71	Lab. Invest.	1952	USA	27	18702	0.130	89.771	5.14
72	Am. Rev. Respiratory Dis. !	1917	USA	26	18780	0.125	90.145	4.48
73	J. Allergy Clin. Immunol. !!	1929	USA	26	18754	0.125	90.021	2.95
74	Cell Tissue Kinetics	1968	UK	26	18728	0.125	89.896	0.75
75	FEBS Letters	1968	NE	25	18805	0.120	90.265	3.57
76	Clin. Res.	1953	USA	25	18491	0.120	88758	0.12
77	Am. J. Obstet. Gynecol.	1920	USA	24	18829	0.115	90.381	1.93
78	Br. J. Exp. Pathol.	1920	UK	24	18853	0.115	90.496	0.95
79	J. Clin. Endocrinol. Metab.	1941	USA	23	18876	0.110	90.606	4.09
80	Arch. Dermatol.	1920	USA	23	18899	0.110	90717	1.99
81	J. Cell Biochem. %	1972	USA	22	18921	0.106	90.822	4.47
82	J. Am. Med. Assoc. (JAMA)	1848	USA	21	19005	0.101	91.225	5.28
83	Am. J. Hum. Genet.	1948	USA	21	18963	0.101	91.024	4.79
84	Neurology	1951	USA	21	18984	0.101	91.125	2.97
85	J. Reticuloend. Soc.	1964	USA	21	18942	0.101	90.923	
86	J. Am. Chem. Soc.	1879	USA	20	19025	0.096	91.321	4.57
87	J. Immunogenetics	1974	UK	20	19045	0.096	91.417	1.12
88	Prostaglandlins	1972	USA	19	19064	0.091	91.509	2.12

Table 3 contd.

Rank No. (1)	Name of Journal (2)	Year of Incept. (3)	Country (4)	Citat. Recd. (5)	% of Col. 5 (6)	Cum. Citat. (7)	% of Col. 7 (8)	1988 Impact Factor@ (9)
89	Thymus	1979	NE	19	19083	0.091	91.600	0.86
90	C.R. Acad. Sci.	1935	FRANCE	19	19102	0.091	91.691	0.37
91	Ann. Rev. Genetics	1967	UK	17	19136	0.082	91.854	15.12
92	J. Physiol.	1878	UK	17	19119	0.082	91.773	3.91
93	Rev. Infect. Dis.	1979	USA	17	19153	0.082	91.936	2.74
94	Europ. J. Biochem.	1967	USA	16	19169	0.077	92.013	3.15
95	J. Cell Physiol.	1932	USA	16	19185	0.077	92.089	3.15
96	Surgery	1937	USA	16	19217	0.077	92.243	1.73
97	Am. J. Reprod. Immunol.	1980	USA	16	19201	0.077	92.166	1.25
98	Ann. Rev. Microbiol.	1947	USA	15	19277	0.072	92.531	8.44
99	Adv. Cancer Res.	1953	USA	15	19262	0.072	92.459	6.54
100	Gastroenterology	1943	USA	15	19232	0.072	92.315	6.13
101	Hoppe-Seyler's Z. Physiol. Chem.**	1877	FRG	15	19307	0.072	92.675	1.93
102	Br. J. Dermatol.	1886	UK	15	19292	0.072	92.603	1.78
103	Br. J. Obstet. Gynaecol.	1902	UK	15	19247	0.072	92.387	1.63

@ From Journals Citations Report
* Formerly Transplantation Review
! Formerly Am. Rev. Tuber. Pulmon. Dis.
$ Formerly J.Biochem. Biophys. Cytol.
!! J. Allergy
% J. Supramol. Struc. & Cell Biol.
** Renamed as Biol. Chem. : Hoppe-Seyler's
%% Formerly Immunochemistry
~ Ceased Publication
~~ Formerly J. Med. Res.
++ Renamed as Research in Immunology

Mapping Knowledge : The Use of Co-Word Analysis Techniques for Mapping a Sociology Data File of Four Publishing Countries (France, Germany, UK and USA)

Xavier Polanco and Luc Grivel
Centre National de la Recherche Scientifique (CNRS),
Institut de l' Information Scientifique et Technique (INIST),
2 allée du parc de Brabois, 54514 Vandoeuvre-lès-Nancy Cedex, France

The first goal of this study is to map knowledge embedded in bibliographical references by means of co-word analysis. This paper discusses a scientometric approach in the field of the social sciences covered by the FRANCIS database. We apply Bradford's law to build the datafiles that will be the input of the automatic clustering and mapping process. The results are presented in two parts : (1) structuring information and identifying the emerging research subjects (cluster analysis), (2) representing these subjects graphically in a two-dimensional space (network analysis). The discussion about knowledge and information spaces provides a perspective : the production of cognitive maps of any developing knowledge field stored in a database at any time.

1. Introduction

We group bibliometrics as well as scientometrics under informetrics. What is Informetrics for? In our field of performance, informetrics operates the following functions : analyzing, assessing and mapping scientific and technical information (STI). The analysis is aimed at answering strategic needs and serving scientific and techinical monitoring purpose. The end product is "information on information". There are two kinds of STI assessment : a metrical assessment of information flows (articles, journals, reports, patents), and a qualitative assessment of the information processed (relevance). Mapping (or graphical representation) consists in presenting STI as maps on which to position both information contents and research actors.

Moreover, informetrics is for us a research programme in the context of an information industry. The Institut de l' Information Scientifique et Technique (INIST) is an integrated information centre, created by the French Centre National de la Recherche Scientifique (CNRS) for worldwide promotion of French and European research. Its mission is to collect and process the results of research and to make them immediately accessible.

Scientometric analysis has mostly been applied in the natural and life sciences. A small number of studies have used scientometric tools to analyse the research developments in the social sciences. Whereas scientometric tools have proved their usefulness as monitors of research developments in the natural and life sciences, evidence on this point is lacking almost completely for the humanities and social sciences disciplines. This paper is an attempt to apply a scientometric approach in the field of the social sciences, and to evaluate its potential usefulness.

The first goal of the study is to map knowledge or "subject maps" as Price said (1986, p. 269). According to Small and Garfield (1988, p. 46) : "The notion that science can be mapped was first clearly stated by D. Price during the 1960's". In order to map knowledge, we use co-word analysis (Callon, Law, Rip 1986). We have implemented (SDOC programmes) the co-word analysis in order to classify and visualize the STI. It is based on the keywords assigned to scientific documents. As a general definition, we shall take a co-word map of scientific information to be the representation of the topology of relationships between distinct subject areas or research themes, which are embedded in the database from which the data has been extracted.

In this paper, we are going to describe the application of our informetric chain (based upon the analysis and processing of word associations in a database) to the social sciences information, in the specific field of sociology. For this purpose, we use the FRANCIS database produced by INIST in France. FRANCIS is a unique set of 20 multidisciplinary bibliographic data bases covering the core of the world literature in Humanities, Social sciences and Economics. Then, we shall limit our analysis to sociology information just as it is stored in a particular database. We will focus our attention on the results of the treatment of the four sets of bibliographic data each corresponding to one of the following publishing countries : France, Germany, United Kingdom and United States of America. We emphasize that this four-country comparison does not represent a complete survey of the state of the art.

citation and co-citations analysis in the scientometric community. The idea to use keywords to describe the distribution of units of information in a scientometrics sense, is due to the Centre de sociologie de l'Innovation de l'Ecole de Mines de Paris. The first article in a journal describing this method was published in 1983 (Callon and alii, 1983; 1986).

According to indexing documentary tradition, a keyword is an indicator of the subject content of a document. We are ready to admit that the problem here is the quality of the indexing process. This problem is known as the "indexer effect" (see Healey and alii, 1986, p. 245; see also Polanco, 1993). It is important to note what the main characteristics of the method are. As we know, the first quantitative pattern of keywords is the frequency. Bibliometricians know that the frequency distribution of words is governed by Zipf's law. The second pattern is the keywords co-occurrence. The idea of co-occurrence is central. Co-words are, as its name indicates, a relationship indicator (as well as citation and co-citation); this cannot be expressed by an isolated word, as for instance the keywords of frequency one. The third level is the formation of clusters. A cluster is a group of associated keywords indexing a number of

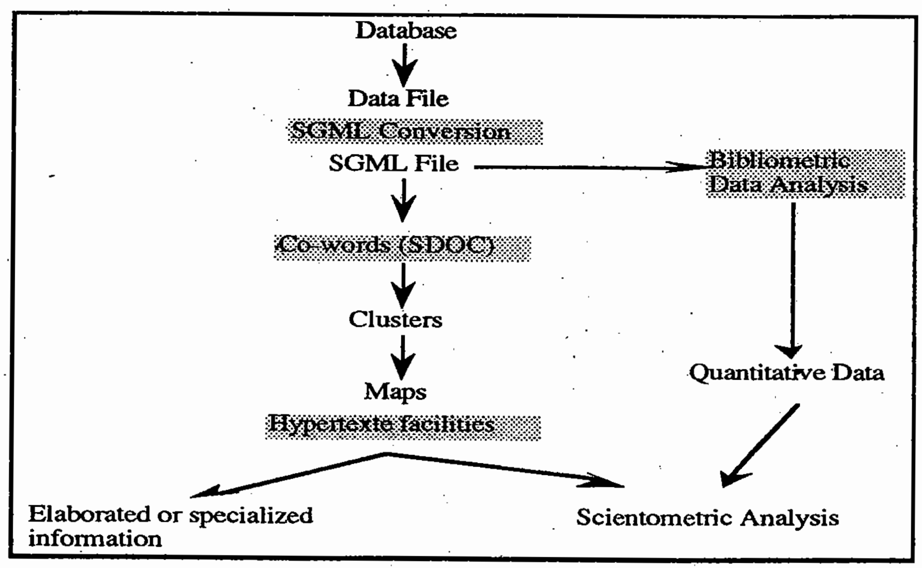


Figure 1 : The informetric chain operates at a number of levels

2. Method

2.1. Co-words analysis

Co-words analysis is an alternative tradition to a more well known and wide-spread tradition of

papers which are the units of information about the subject indicated by the cluster. In this sense, each cluster is an indicator of knowledge founded on frequency and co-occurrence statistical values. The last procedure is to map knowledge. Usually scientific knowledge is in the form of collections of bibliographic data. A representation is a set of conventions about how to describe information. Thinking seriously about scientific information of any sort requires thinking seriously about what representation is best suited to the domain from which the bibliographic information comes from. Indeed, the main purpose of the paper is to demonstrate the central importance of finding good representations capable of bearing good descriptions.

2.2. SDOC programmes

SDOC is an original implementation of co-word analysis from the point of view of computer information technology. The programmes are implemented under the Unix operating system and written in C language, according to a toolkit philosophy based on modularity and data flow communication rather than a "press-button software package". Thus, the modules are organized in a toolbox (Ducloy and alli, 1991).

We use SGML (Standard Generalized Markup Language) to describe bibliographical references whatever their source; and SGML standard is used as pivot format and as specification language for coding intermediate data (see Figure 1).

Co-words analysis has been broken down into the following steps, each corresponding to an independent module communicating with others by file : (1) Building keywords index from a set of bibliographic references ; (2) Computing cooccurrences of keywords and measuring the similarity of the keywords; (3) Cutting up the keywords associations network into clusters; (4) Classifying references into clusters; (5) Building scatter diagrams or maps. In the process, SGML is the specification language of data transmitted from one step to another. The modules are parameterized by the input and the output filename and its specific parameters. Statistics tools and visualization tools are available to assess the results. Finally, the hypertext technique provides the analysts with an interactive working tool (Grivel and Lamirel, 1993).

The clustering algorithm which groups the associated keywords into clusters is an adaptation of the single link clustering algorithm. After the clustering process, the documents are assigned to the clusters.

In order to permit an easier consultation and interpretation of the clustering and mapping results, SDOC converts the "knowledge network" represented by the clusters into hypertext nodes and links. The goal of this conversion is to allow the user to visualize very quickly the knowledge organization of a topic, the key figures, the organisations.

3. Data & Bibliometric Analysis

We use the bibliometric analysis in order to build the datafile that will be the input of the automatic clustering and mapping process (SDOC programmes). This bibliometric analysis is based on some bibliometric indicators, as for instance, the document type, the date of publication, and the country of publication.

3.1. Construction of the data file

A funnel-shaped step by step process is applied on the basis of the selected bibliometric indicators as criteria of construction of the final "target datafile". The main steps of this funnel-shaped process are : (1) the extraction of a datafile from the FRANCIS database; (2) from this source datafile a first raw datafile is constructed by means of the document type bibliometric indicator application; (3) a second datafile is extracted from the former by the application of the country of publication as criteria of selection (4) and finally, the input datafile to SDOC programmes results from the application of Bradford's law.

The first step in an informetric analysis is to extract the "target literature" from an information retrieval database (Turner and alii, 1988). In a present case, we simply start from the literature collected and processed from 1989 to 1991 by the FRANCIS database on sociology.

The size of this datafile is of 13.942 records; there are different types of documents : journal articles, books, reports, proceedings and Ph. D. This is our starting raw datafile from which we shall build a second datafile. The decisions here was to focus on the journal articles.

There are predominantly journal articles in our source datafile, they represent 84% of the sociological information stored in the FRANCIS database from 1989 to 1991. There are 720 journals from which 11 661 articles originate. The other sources are books, reports, proceedings, and theses, at 16% of the raw datafile. The reports and theses essentially concern French sociology.

The date of publication of this literature corresponds mainly to the last years of the 1989's sociology. (13.735 records, 98%, between 1986 and 1991).

The authors' institutional affiliation does not appear in the FRANCIS references, so we have used the publishing country of journals for the definition of our "target" literature. As we can see in Table 1., the journals of the four publishing countries selected and the articles issued from these sources represent 70% of the total. The remaining 30% is distributed over 44 publishing countries in the world. This is a long-tailed distribution, and as we know, this type of distribution appears to be characteristic of bibliometrics. We decided to focus our analysis on this literature of the four publishing countries. In this set, France is over-represented, Germany and United Kingdom are approximately equal, and the United States are well represented.

Table 1.

Publishing country	Number of journals	%	Number of references	%
France	270	37.55	3245	27.83
Germany	30	5.42	860	7.38
United Kindgom	49	6.82	1310	11.23
United States	143	19.89	2787	23.90
Total	501	69.68	8202	70.33
All publishing Countries : 48	719	100.00	11661	100.00

Considering the excessive impact of two journals in the set of 501 journals selected, and their specialized nature, we decided to treat them separately and to remove them from the "target literature". They are *Economie et Statistique* (France), and *Journal of Marriage and the Family* (Etats - Unis). The clustering of the 249 articles of *Journal of Marriage and the Family* (Etats-Unis) provided 21 clusters and we obtained 11 clusters from the 138 articles of the journal *Economie et Statistique* (France). This case underlines that, if the number of references is statistically significant, one can proceed to a content analysis of these references using SDOC, as the one we discuss in section 4 of this paper. We shall not present here the results obtained from these two journals.

So our "target literature" becomes at last : France 269 journals as a source of 3.107 references; United-States 142 journals as a source of 2.538 references. Germany and United-Kingdom remains unchanged.

3.2. Application of the Bradford Law

To select the "target literature", we applied "Bradford's law of rank distribution". The four-country sociological journals were ranked by decreasing productivity of articles. Then for each country, we defined 4 groups (see Table 2) composed of the most productive journals so that their cumulative number of references reach respectively 25%, 50%, 75% and 100% of the corpus.

Table 2.

Publishing country	Number of references	Number of journals with % of references			
		25%	50%	75%	100%
France	3107	12	31	68	269
Germany	860	3	6	15	39
United Kingdom	1310	3	8	15	49
United States	2538	7	17	32	142

We defined as "nuclear zone" the journals which produce 50% of the references. We focused on the four-country nuclear zone (Table 3), in order to map the most important publications, of course according to FRANCIS database coverage on sociology.

Table 3.

Publishing Country	(S) Journals	%	(R) References	%
France	31	11	1568	50
Germany	7	15	462	53
United Kingdom	8	16	676	52
United States	17	12	1287	51

So, the "target" bibliographical data, that we used for the mapping process, is not only a set of sources (S) and a set of references (R), but also the application of a function expressing the source-reference relationships; it is the Bradford's ranking analysis. From the point of view of the date of publication, the nuclear zone is a sociological literature published during 1987 -1990. France is over-represented comparatively to other publishing countries. It may be an expression of the wish of exhaustivity to cover national literature. A certain eclecticism is expressed by a two-level literature : one is more strictly scientific or academic, the other one corresponds more to an enlightenment literature. The catagory of enlightenment publications includes popularizing articles and reviews in magazines. We take the distinction between 'scholarly' and 'enlightenment' publications from Nederhof and alii, (1989, p. 427 - 428). This is not the case for the other countries where the journals selection appears much stricter. These facts only express a policy of coverage of journals. We cannot use these data to compare countries' productivity. The inequality existing in the productivity of the four countries is not a problem for the goals we have fixed in our introduction.

4. Results and Commentary

The obtained results are presented in two parts. The first one is dedicated to the presentation of the lists of clusters and the second one to the mapping of the clusters on scatter diagrams. It corresponds to two phases of the method. In the first phase, it is a question of structuring information and identifying the emerging research subjects (cluster analysis). The second phase is the graphic representation of these subjects in a two-dimensional space (network analysis).

Figure 2 allows us to distinguish two other phases concerning the information processing, (1) a first machine-based phase, the SDOC application, and (2) the phase where there is the action of an expert or knowledgeable person. Our information processing is based on cluster and network analysis techniques, in consequence the expert's goal is to study the themes and networks. In this second phase, hypertext represents an analytical tool which allows navigation through the information space of clusters and networks.

4.1. Cluster analysis

Cluster analysis is, as we know, the generic name for a wide variety of procedures that can be used to create a classification. The procedure empirically forms clusters or groups of key words. The clustering method is a multivariate statistical procedure that starts with a bibliographical data set containing information about a subject and attempts to reorganize the bibliographical information into relatively homogeneous groups. As we have already noted in section 2, the coword clustering method (implemented by SDOC programmes) is designed to create groups or clusters of associated keywords (co-words) as a means to indicate some numbers of research themes. In this particular application on sociology data file, we have applied the Equivalence Index. If we call C_{ij} the cooccurrence number of two keywords i and j, C_i and C_j their occurrence numbers, the Equivalence Index (E_{ij}) is given by the following equation :

$$E_{ij} = C_{ij}^2 / (C_i \times C_j)$$

The clustering algorithm which groups the associated keywords into clusters is an adaptation of the single link clustering algorithm. All the elements which are to be initially clustered constitute a large flat association network, i.e. a system of relationships where the keywords are related to each other. The separation of the association network into clusters is done according to a readability criteria : the cluster size (minimum and maximum number of components) and the number of associations in the cluster. If a pair of terms belongs to the same cluster, the association between the terms is an internal association. If they belong to two different clusters, the algorithm tries to aggregate the clusters into one by merging them. The merger is authorized if the size of the resulting new cluster respects the "readability

SDOC	Analytical Action	Object Study
Automatic Classification ⟶ Clusters	Cluster Analysis	Research Subjects or Themes
Graphic Representation on two-dimensional space (y,x) ⟶ Maps	Network Analysis	Global & Local Networks

Figure 2 : Human-machine information processing

criteria" If not, the association is considered as an external association. In this application, the parameters for each datafile were : minimal size of the clusters = 4 keywords; maximal size of the clusters = 10 keywords; maximal number of external associations = 10; maximal total number of associations = 20.

After the clustering process, the documents are associated to the clusters. A document is related to a clauster if, within its indexing terms, there is at least one pair of terms which can constitute either an internal association or an external assciation. We associate a list of authors, and a list of document sources to each cluster, as this information is available in the studied datafile.

Number of lines	Definition of the statistical parameters
[1]	Minimal cooccurrence of keywords (cooccurrence threshold)
[2]	Initial number of documents
[3]	Number of documents with at least a couple of keywords whose cooccurrence value is superior or equal to the cooccurrence threshold
[4]	Number of clusters
[5]	Number of documents in the clusters
[6]	Number of documents appearing only in one cluster

Table 4.

	France	Germany	U K	U SA
[1]	4	2	3	4
[2]	1568	462	676	1287
[3]	1119	392	498	938
[4]	28	24	17	20
[5]	944	324	434	756
[6]	493	156	233	422

These are the main global indicators which allow us to adjust the clustering process by measuring the loss of information in function of the cooccurrence threshold and then the ratio number of references in the clusters / initial number of references. Table 4 provides only the data corresponding to our final choice for that application. We have tried to find a good compromise between the number of clusters for each data file and the loss of information due to both the selected cooccurrence threshold and the clustering parameters.

The statistical variables which characterize each cluster are the following :

Number of columns	Definition of the statistical parameters
[1]	Cluster's saturation threshold
[2]	Density, the mean of the internal associations
[3]	Centrality, the mean of the external associations
[4]	Number of keywords defining the subject
[5]	Number of internal associations (between the keywords defining the subject)
[6]	Number of external associations with other subjects (or clusters)
[7]	Number of citations of a subject by other subjects
[8]	Subject's bibliographic information (number of references)
[9]	Specific subject's bibliographic information

We indicate for each cluster the quantitative value of these parameters. The values of the first three cloumns [1], [2], [3] in the tables below are obtained by the Equivalence index; those of the columns [4], [5], [6] are the size parameters of clusters which results from parameters fixed a priori for building clusters. The values of the last two columns [8] [9] concern documents classification by clusters. These are the indicators which allow us to characterize the clusters.

In the tables 5 to 8 in the appendix, each cluster is a row and each staistical parameter a column. Then we can choose a parameter, and rank the clusters according to their quantitative values in the selected column. Here, the clusters have been sorted by [2] *density value,* the mean of the internal associations which characterizes the strength of the links between the words making up the cluster (intra-cluster associations). The stronger these associations are, the more the subject corresponding to the cluster constitutes an integrated unit of information (or knowledge) *Centrality* [3] measures, for a given cluster, the intensity of its external associations with other clusters (inter-clusters associations). The more of these associations there are, and the stronger there are, the more this cluster designates a subject that is considered important in the knowledge network. The word *citation* [7] is used to indicate the fact that one cluster has been cited in the external associations of another cluster. When one cluster, by its external associations, refers to another cluster, the latter has been cited by the former as a related item of information. The bibliographic informtion represented by a given cluster is measured and characterized by the parameters [8] and [9]. The column [9] is also an indicator of the bibliographic independence of a cluster in relation to other clusters.

The name of a cluster is only a label. The heuristic use to label the clusters is to choose the keyword which appears the most frequently in the associations. The name of a cluster suggested automatically may sometimes be more a mask than a source of information. The program should allow an expert to change the name in this case, as, for instance, for the *Relations* cluster in the four lists of clusters. But taking into account that this cluster is related to a significant number of bibliographic references, SDOC programmes permit us to come back to this number of references, to isolate it in a datafile and to process this datafile in order to again obtain a classification of the information masked by the label, a visualisation of information. We call this action the *"russian doll" procedure.*

One can also see the use of the word *region* in the singular and plural forms. This demonstrates a certain indexing policy and indicates for us the need to adopt methods to normalize the indexing vocabulary in input in order to correct these undesirable effects.

These tables of clusters enable us to know something about the problems studied and their relative importance in the datafile. We can then analyse in more detail each element, that is to say (1) the keywords which form one cluster, (2) the internal and external associations with other clusters, (3) the sources, (4) the authors and (5) the titles of articles belonging to clusters. The conversion of all this data into hypertext hugely facilitates these operations. It increases the analysis and assessment task performance of this information, previously structured by the automatic clustering process.

We can also compare the research subjects in each case; for instance to compare the European publishing research in sociology, to compare it as a whole with the United States, from the point of view "study subjects" (similarities, differences) and areas of research as for instance social, economical or political areas. We can also focus on a subject in the four countries (transverse analysis), for instance, technological innovations or social deviances (see maps below).

Another possibility is to use the co-word clustering process as an instrument for bibliographic retrieval. Retrieval systems are designed to enable a user to query a database of documents or document surrogates. In this sense, we have a co-word based retrieval system, where the user can navigate through clusters in different subject areas of research and immediately identify their authors, journals, titles of papers.

Looking at the scatter diagrams is the next step of the co-word analysis. The scatter diagram for any set of keyword-clusters shows what we call a "knowledge space "(Meincke and Atherton, 1976), or "information space" (Brookes, 1980). In this space, clusters are the indicators of items of knowledge and their positions are indicators of the *density* (Y axis) and *centrality* (X axis) of this item of knowledge. Such diagrams are included in the next section of this paper. Each scatter diagram is a representaion of a set of clusters using the values of the columns [2] and [3] of the tables 5 to 8

4.2. Representing Knowledge in Scatter Diagrams

From a perspective of analysis, the first stage of description was the cluster analysis, and now the second step is the network analysis. Relations are principally the subject of network analysis. A network is a type of relation linking a defined set of clusters (unit of information). The clusters can be defined as micro-networks or graphs and the maps as macro-networks. They are the building blocks of our network analysis.

We propose a two dimensional device for visualizing the organization of objective knowledge diffused by bibliographic data (information). We develop a representation of information items. The chief output in a spatial representation, consisting of a configuration of subjects (or clusters), as on a map. Each subject in the configuration corrsponds to one item of information. This configuration reflects the "hidden structure" in the data, and often makes the data much easier to comprehend.

Before going into details about the description, a remark must be made about the sense of the scatter diagrams in our procedure. We use them as a way to produce a knowledge representation. "A representation has been defined to be a set of conventions for describing things. Experience has shown that designing a good representation is often the key to turn hard problems into simplest ones, and it is therefore reasonable to work hard on establishing what symbols a representaion is to use and how those symbols are to be arranged to produce descriptions of particular things" (Winston, 1977, p. 179).

On the other hand, as Poppers says (1979, p. 108 - 109) there are two different senses of knowledge, the first is "knowledge in the subjective sense, consisting of a state of mind", and the second is "knowledge in an objective sense, consisting of problems, theories, and arguments as such. Knowledge in this objective sense is totally independent of anybody's claim to know; it is also independent of anybody's belief, or disposition to assent; or to assert, or to act. Knowledge in the objective sense is knowledge without knower; it is knowledge without a knowing subject". Knowledge is taken by us in an objective sense, consisting of journal literature, the medium through which natural or social scientists report their own original work and in which they evaluate work done by others.

Two main categories of problems arise from the study of scientific knowledge. One deals with the act of producing knowledge; the other is concerned with the very structures of knowledge produced by scientific activity. (see Popper, 1979, p. 112-113). We are concerned in our study by this second category of problems. Co-word analysis is a way of mapping the structure of scientific knowledge expressed by authors in their publications.

What do maps actually represent? On the one hand, they represent a set of clusters which designate specific centres of interest or themes or subject areas. On the other hand, they represent a network structure. It is a two-dimensional space. The Y axis called "density indicator" is defined by the strength of the internal word associations. It is thought to indicate internal coherence of the subject area. The X axis called "centrality indicator" is defined by the strength of the external associations. It indicates the role of a subject area in structuring a field or research.

When Derek de Solla Price said that the pattern of bibliographic references indicates the nature of the research front, he was clearly thinking of the citation analysis (Price, 1965). The citation of one paper by another in its footnotes or bibliography was the basis of his idea tha science can be mapped. The co-word analysis is another tradition in mapping science. We emphasize that co-word maps are representations of knowledge structures network.

The figure 3 shows that with two theoretically important attributes, *density* and *centrality*, we have four possible combinations (see Callon et alii, 1991, p. 165 - 167).

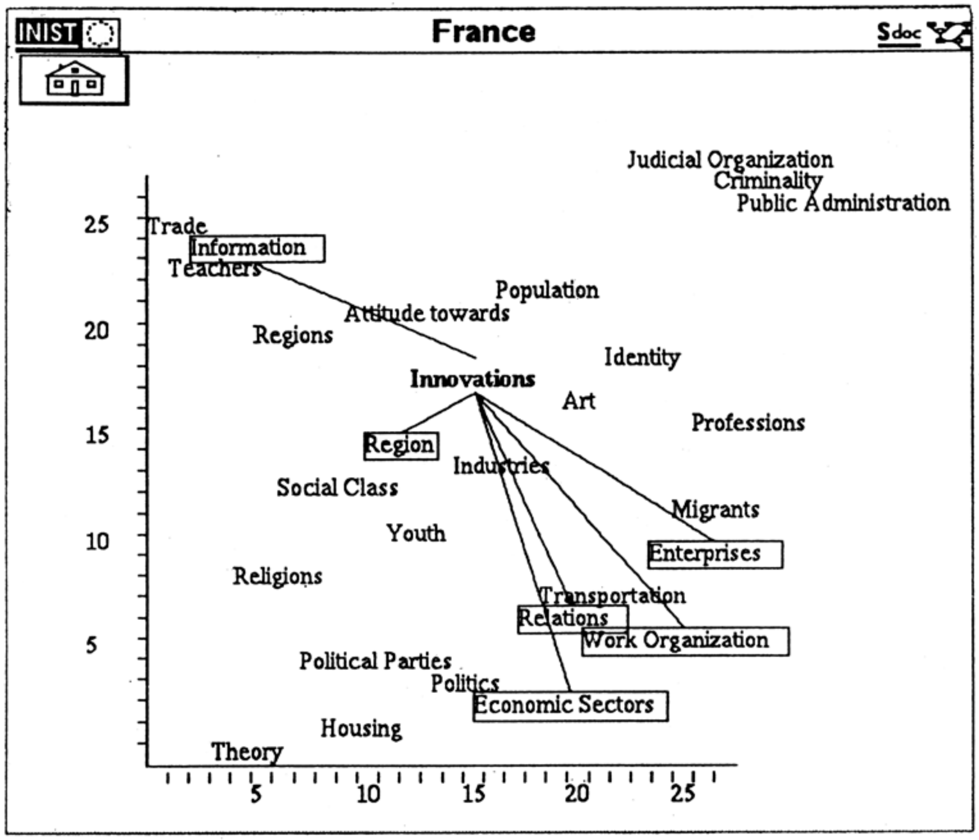
Y		
	High density and Low centrality	High density and centrality
	Low density and centrality	Low density and High centrality
	X	

Figure 3: Scatter Diagram and categorized classes of clusters as indicators of research subjects identified from bibliographic data by computer programmes.

In examining a scatter diagram, the first thing to look for is the clusters distribution in these four zones of the diagram. The clusters are scattered according to the mean value of the internal associations (along the Y - axis), and of the external associations (along the X - axis). The information provided by the diagrams concerns the relative importance of themes or subjects (clusters) according to these two attributes : *density* and *centrality*. This relative importance of clusters is set up from the network of internal associations of each cluster (position along the Y - axis), and external associations between the clusters (position along the X - axis). The first value (along the Y - axis) defines categories of subjects more or less coherent and integrated as units of information. The second value (along the X - axis) defines more or less isolated or linked clusters, this is the notion of *centralness* of a theme in the knowledge space.

Our scatter diagrams are not metric spaces; the fact that two or three clusters are close to one another does not mean that they are closely linked to each other. On the other hand, we arrange the clusters by rank on the Y and X - axis. The number of ranks is equal to the number of clusters. So, the maps can be interpretated as rows on the Y - axis and columns on the X - axis.

Now we are going to show how the maps can be used to help the analysis of research themes such as



Map 1

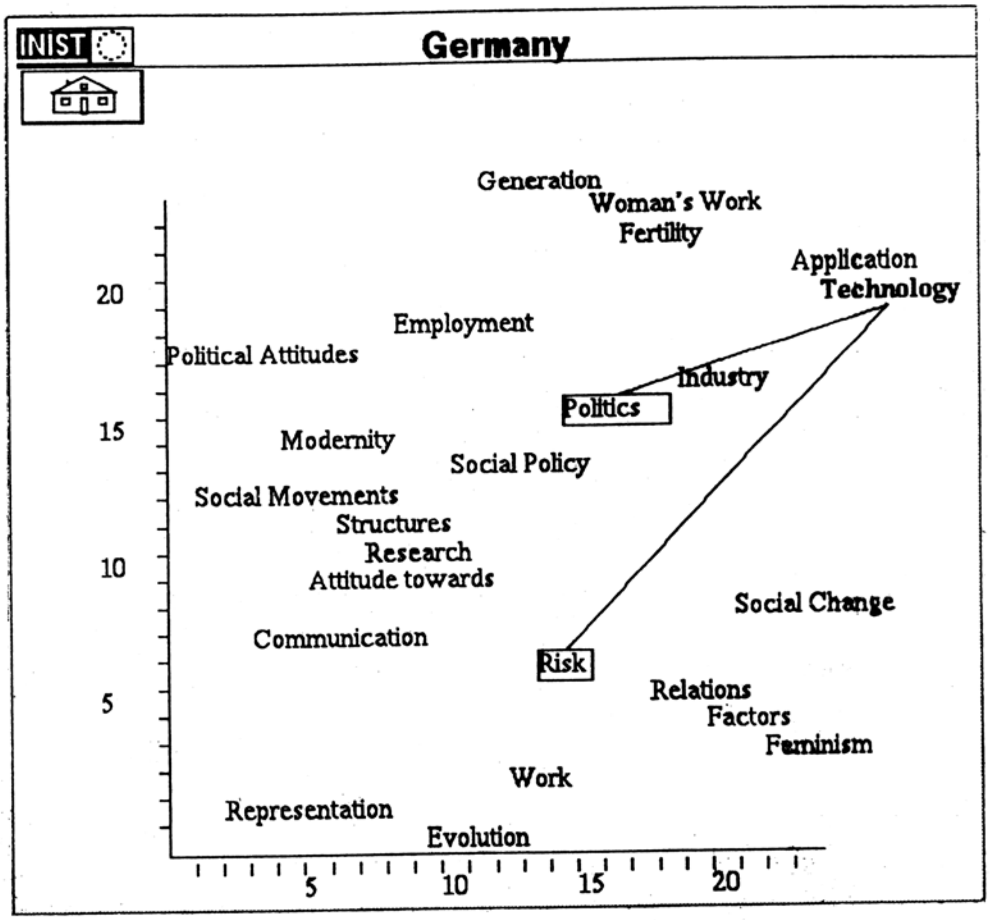
those linked with other themes forming a network in this way. Certainly, the analysis may descend to the level of the authors, sources and articles each time.

The map 1 shows a set of three clusters together in a position of high density and high centrality. They are the themes about *Public Administration, Criminality* and *Judicial Organization.* This centrality is specially explained because they are closely connected, but at the same time each one represents an integrated internal unit of information on this subject (or high density). In reality, they represent an information area that is the result of the weight of certain specialized journals in security, criminology and laws in the sources of the data file. This area is open to *Politics* and *Professions* by means of the external associations of the *Judicial Organization* cluster.

The map 4, which gives a representation of the sociological literature published in journals edited in the United States, also highlights a dimension of social deviance. This is again the same phenomenon, that is to say the important weight of the sources of information specialized in these subjects.

In the case of the European maps (maps 1, 2 and 3), our choice has been to show here the maps that highlight the associations of clusters as *Technology* or *Innovation* or *Technological Innovation* with the other clusters. This is in order to show how a cluster, that we consider as a graph, or in other words a micro-network, can be situated inside a larger network, macro-network or context. We can compare the position of the subject *Technology & Innovation or Technological Innovation* in the three contexts.

In the case of France, this subject appears at an average position along the two axes, and is linked (new communication technologies) with *Information,* a subject whose position on the map indicates that it is cohesive but without centralness. At the same time, the other associations refer to significative clusters from the centrality point of view, but lowly integrated as unit of information. All these clusters constitute an economic area of sociologial research. (There we also have the example of "indexer effect" : a same concept



Map 2

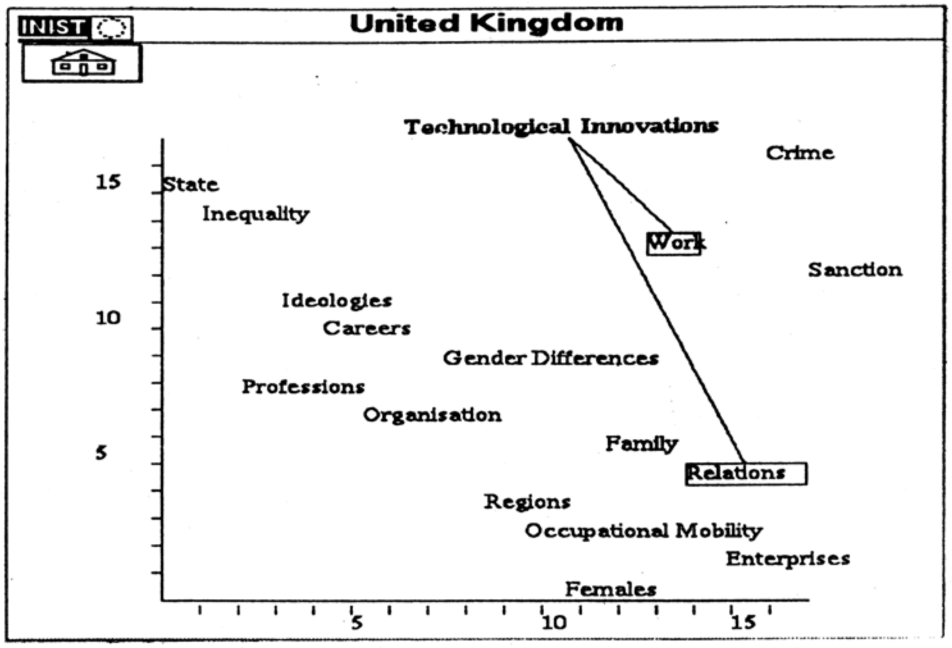
is expressed in the singular and the plural form, *Region* and *Regions,* whenever they constitute a single and same concept).

On the map 2 concerning the German journals of sociology, the cluster *Technology (Innovation or Technological innovation)* is plot at a high value along the two centrality and density axes. By means of its internal associations, it exhibits three sectors (1) computerisation, (2) enterprise and industrial enterprise, and (3) human genetic engineering; the external associations refer to clusters *Politics* and *Risk* (more specifically the nuclear risk). In the cluster *Politics,* we have a junction concerning "mass media" and "public opinion".

On the France map, the technological innovation theme is linked to economic development and work organisation changes (also visible on the United Kingdom map). Whereas on the map 2 (Germany), this theme is associated to the risks and social impacts of the computerisation and the genetic technologies applied to human reproduction.

Now, if we look at the map 3 United Kingdom map, the *Technological Innovation (or Innovation)* is a high density and high centrality cluster, associated with *Work* and *Relations.* Again, we find the ambiguous word *Relations* as a descriptor and then as a label of a cluster. But the *"Russian doll procedure"* is handy to visualize what is hidden under this subject because of the number of records aggregated in this cluster (128 records). *Work* is a cluster in which we find sociological studies on skill and deskilling problems because of the technological changes, and the *Work* cluster is associated by its external associations to the theme *gender differences.* This is the context in which the social studies of technological innovation are situated in our information space.

As in the case of the literature published in journals edited in France and the United-States, the United Kingdom map shows that the subjects *Sanction* and *Crime* stand out. This is an indicator of the relative importance of the sociological research ded-



Map 3

icated to social deviance problems.

The United States map is a representation of the important weight of the specialized publcations in social deviance and anomy. The network is a graphical representation of the information essentially published by the journals *Criminology, Crime and Delinquency* followed by *Social Forces* and *Social Problems.*

The information on technological innovation is not visualized on the map, this informationis inside the *Regions* cluster, because the studies concern the agriculture, and their source is the *Rural Sociology* journal. On the contrary, the sociological studies on technological innovation published by European journals appear in an industrial context of change and innovation published by European journals appear in an industrial context of change and innovation (France, Germany, United Kingdom), also in the context of communication technologies (France, German) and human genetic engineering (German).

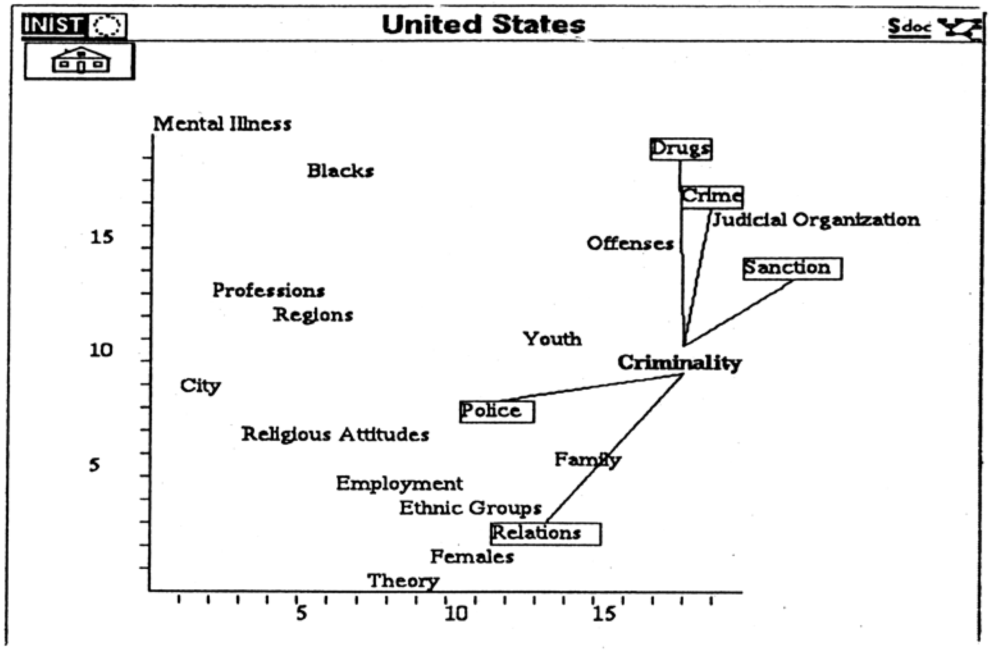
This brief overview of four cases of network analysis underlines the important emergence of the structural properties of the information. In addition, we can note the problems induced by the extreme sensibility of the co-word analysis to the quality of indexing. Beforehand an important effort of normalization is needed. A second problem is always to consider explicitly the type of journals which are at the origin of the information thatwe analyse and represent, especially when it is a question of representing the results of a given field of research.

Finally, it is interesting to underline that maps allow a user to visualize the knowledge structure of the document data file. The idea is to present information within a cognitive structure so that the experts can assess its validity. On the other hand, as Brookers says (1981, p. 10) : "As a map grows it will reach a stage at which it could be used as a database"

5. Conclusion

We would like to stress two main purpose concerning our approach. The first is to map knowledge structures, and the second is to watch science activity by means of its bibloiographic output as items of information.

Mapping knowledge structures : this discussion about knowledge and information spaces provides a perspective, the production of cognitive maps of any



Map 4

developing knowledge field stored in the database at any time. Furthermore, SDOC programmes rely on the hypertext paradigm to represent the thematic maps, and allow the user to navigate through a hyperspace composed of clusters, relationships between clusters, documents related to these clusters, and so on. Such a hypertext map would become of strategic interest to those with competence in the field.

Watching science activity : the coword maps visualize the structure of relationships between subjects of research and the way in which this network evolves with time. Thus, this method may be useful to identify subject research areas, and to investigate the distribution of publications, institutions, countries, in these areas of research. The goal is to indicate « who is doing what, where and when » (4W) with respect to the topics and centres of interest identified on the maps.

6. Epilogue

Today, the informetric techniques and the databases may be considered, in our opinion, as the contemporary instruments for representing and visualizing the state of scientific knowledge (natural and social sciences), the way Galileo turned the telescope on the heavens and set up the modern scientific revolution at the beginning of the Seventeenth Century.

Furthermore, we think that Price's instrumentality theory of innovation (see Price, 1984) can be applied to the informetric techniques field which offer new instrumentalities in order to produce a more empirical approach vis-a-vis traditional epistemology, taken to be the theory of scientific knowledge. As we know, Price coined the term instrumentality in order to indicate methods and techniques from which spring a scientific change or a new technology.

Table 5. France

No.	Name	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	Judicial Organization	0.211	0.326	0.044	4	3	2	2	17	6
2	Criminality	0.070	0.279	0.052	7	15	5	1	21	14
3	Public Administration	0.099	0.241	0.058	10	13	1	5	36	21
4	Trade	0.208	0.236	0.005	4	3	2	0	20	6
5	Information	0.070	0.230	0.010	5	5	5	1	26	8
6	Teachers	0.052	0.183	0.007	6	6	2	0	27	18
7	Population	0.073	0.162	0.029	9	8	4	2	40	14
8	Attitude towards	0.032	0.159	0.019	8	7	10	5	69	18
9	Regions	0.020	0.158	0.015	6	5	10	5	65	6
10	Identity	0.021	0.154	0.042	10	13	8	5	41	9
11	Innovations	0.039	0.153	0.023	7	9	10	4	57	8
12	Art	0.127	0.151	0.038	4	3	1	1	15	9
13	Professios	0.100	0.135	0.051	10	13	4	9	59	25
14	Region	0.031	0.134	0.020	8	9	8	7	68	16
15	Industries	0.052	0.130	0.025	5	4	7	2	28	7
16	Social Class	0.052	0.129	0.017	4	3	2	1	18	6
17	Migrants	0.060	0.124	0.049	10	13	4	12	89	32
18	Youth	0.022	0.123	0.020	8	8	10	5	53	15
19	Enterprises	0.045	0.117	0.044	10	11	8	17	84	24
20	Religions	0.080	0.117	0.113	6	5	8	2	44	14
21	Transportation	0.059	0.115	0.034	10	16	1	8	64	32
22	Relations	0.036	0.109	0.030	10	10	6	28	176	31
23	Work Organization	0.076	0.107	0.042	8	10	8	11	73	12
24	Political Parties	0.030	0.105	0.018	10	11	7	8	110	42
25	Politics	0.030	0.105	0.024	10	11	8	10	101	33
26	Economic Sectors	0.018	0.094	0.025	10	10	10	12	88	11
27	Housing	0.044	0.093	0.019	10	9	6	2	61	28
28	Theory	0.027	0.051	0.012	4	3	10	2	56	28

Table 6. United Kingdom

No.	Name	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	Technological Innovation	0.100	0.284	0.020	5	6	6	2	16	4
2	Crime	0.067	0.224	0.052	9	13	6	4	29	15
3	State	0.032	0.205	0.010	4	3	5	0	25	14
4	Inequality	0.083	0.204	0.011	6	6	5	0	27	13
5	Work	0.032	0.184	0.029	8	10	9	10	65	16
6	Sanction	0.114	0.183	0.065	10	15	2	6	31	17
7	Ideologies	0.040	0.168	0.015	4	3	3	1	19	7
8	Careers	0.039	0.163	0.019	6	6	10	6	27	6
9	Gender Differences	0.018	0.138	0.022	8	7	10	7	59	17
10	Professions	0.041	0.135	0.012	6	6	4	0	27	11
11	Organisation	0.017	0.106	0.019	7	7	8	3	40	16
12.	Family	0.021	0.102	0.028	9	8	8	4	44	11
13	Relations	0.042	0.088	0.032	10	10	7	35	128	44
14	Regions	0.044	0.084	0.024	4	3	5	3	19	6
15	Occupational Mobility	0.036	0.082	0.026	9	10	9	7	66	18
16	Enterprises	0.048	0.068	0.040	10	15	4	5	49	13
17	Females	0.040	0.057	0.028	6	6	9	17	63	5

Table 7. Germany

No.	Name	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	Generation	0.071	0.298	0.026	9	9	5	3	18	6
2	Woman's Work	0.044	0.270	0.038	10	12	8	4	18	6
3	Fertility	0.071	0.270	0.039	8	8	10	8	23	6
4	Application	0.108	0.247	0.072	9	14	5	24	52	11
5	Technology	0.095	0.243	0.082	7	15	2	11	21	3
6	Employment	0.045	0.229	0.023	9	8	10	4	26	5
7	Political Attitudes	0.222	0.222	0.005	4	3	1	0	6	4
8	Industry	0.100	0.208	0.043	7	7	10	4	18	3
9	Politics	0.038	0.201	0.038	10	12	8	6	26	3
10	Modernity	0.044	0.198	0.017	10	10	9	0	29	16
11	Social Policy	0.029	0.195	0.024	10	16	4	1	26	17
12	Social Movements	0.041	0.189	0.006	4	4	4	0	12	5
13	Structures	0.042	0.188	0.021	7	6	10	3	21	5
14	Research	0.050	0.169	0.023	5	4	10	5	17	3
15	Attitude towards	0.034	0.160	0.017	7	7	10	4	22	9
16	Social Change	0.045	0.158	0.048	10	10	8	12	42	10
17	Communication	0.057	0.157	0.015	4	5	3	0	10	7
18	Risk	0.036	0.145	0.034	9	11	9	4	28	8
19	Relations	0.033	0.143	0.040	8	9	10	48	71	8
20	Factors	0.073	0.142	0.048	10	10	9	8	22	4
21	Feminism	0.087	0.130	0.058	6	8	6	10	19	8
22	Work	0.054	0.125	0.031	7	6	9	9	23	2
23	Representation	0.063	0.122	0.007	6	5	5	0	16	6
24	Evolution	0.020	0.102	0.023	9	9	10	7	28	1

Table 8. United States

No.	Name	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	Mental Illness	0.200	0.263	0.000	5	4	0	0	14	8
2	Drugs	0.062	0.250	0.045	8	12	8	5	47	16
3	Blacks	0.045	0.225	0.015	6	7	7	4	39	9
4	Crime	0.111	0.215	0.054	7	8	8	11	84	16
5	Judicial Organization	0.061	0.203	0.060	10	15	5	11	85	29
6	Offenses	0.019	0.184	0.036	9	9	9	5	45	13
7	Sanction	0.134	0.168	0.090	9	14	3	17	53	8
8	Professions	0.024	0.150	0.010	6	6	4	1	30	10
9	Regions	0.028	0.146	0.013	8	8	9	3	71	33
10	Youth	0.045	0.136	0.032	5	4	10	5	50	10
11	Criminality	0.055	0.132	0.040	7	10	9	13	74	13
12	City	0.020	0.120	0.009	9	8	6	1	49	20
13	Police	0.047	0.116	0.027	5	5	10	4	55	12
14	Religious Attitudes	0.035	0.111	0.012	5	4	4	0	27	17
15	Family	0.049	0.098	0.035	7	6	9	9	67	16
16	Employment	0.036	0.077	0.015	7	7	10	2	65	24
17	Ethnic Groups	0.017	0.059	0.023	9	9	10	7	76	26
18	Relations	0.045	0.058	0.029	9	10	5	28	160	35
19	Females	0.030	0.049	0.024	10	13	5	8	107	39
20	Theory	0.027	0.043	0.019	10	9	4	1	104	68

References

1. Braam, R.R. H.F. Moed, A.F.J. van Raan, "Comparison and Combination of Co-Citation and Co-Word Clustering", in *Select Proceeding of the First International Workshop Science and Technology Indicators,* Leiden, 14 - 16 November 1988, p. 307 - 337.
2. Brookes, B.C. "Information Space", *The Canadian Journal of Information Science,* vol. 5, 1980, p. 199 - 211.
3. Brookes, B.C. "The Foundations of Information Science. Part IV : Information Science : The Changing Paradigm", *Journal of Information Science,* vol. 3, 1981, p. 3 - 12.
4. Callon, M. J-P. Courtial, W.A. Turner, S. Bauin, "From translations to problematic networks : An introduction to co-words analysis", *Social Science Information,* vol. 22, n° 2, 1983, p. 191 - 235.
5. Callon, M. J. Law and A. Rip (eds), *Mapping the Dynamics of Science and Technology.* London, Macmillan Press, 1986.
6. Callon, M. J-P. Courtial, F. Laville, "Co-word analysis as a tool for describing the network of interactions beteween basic and technological research : The case of polymer chemistry", *Scientometrics,* vol. 22, n° 1, 1991, p. 155 - 205.
7. Ducloy, J. P. Charpentier, C. Francois, L. Grivel, "Une boîte à outils pour le traitement de l'Information Scientifique et Technique", *Génie logiciel,* n° 25, 1991, p. 80 - 90.
8. Grivel L. et J-Ch. Lamirel, "An analysis tool for scientometric studies integrated in an hypermedia environment", in *Proceedings of ICO93* 4th International Conference on Cognitive and Computer Sciences for Organizations, Montreal, (Quebec) Canada, 4-7 mai 1993, p. 146 - 154.
9. Healey, P. H. Rothman, P. Hoch, "An Experiment in Science Mapping for Research Planning", *Research Policy,* vol. 15, 1986, p. 233 - 251.
10. Meincke P. and P. Atherton, "Knowledge Space : A Conveptual Basis for the Organization of Knowledge", *Journal of the American Society for Information science,* vol 27, 1976, p. 18 - 24.
11. Michelet, B. L'analyse ds associations. Paris : Thèse de doctorat, 1998.
12. Nederhof, A.J. R. A. Zwaan, R.E. de Bruin, P.J. Dekker, "Assessing the Usefulness of Bibliometric Indicator for the Humanities and the Social and Behavioural Sciences : A Comparative Study", *Scientometrics,* vol. 15, n° 5-6, 1989, p. 423 - 433.
13. Polanco, X. "Analyse stratégique de l'information scientifique et technique. Construction de clusters de mots-clés", *Sciences de la société,* n° 28, 1993, p. 111 - 126.
14. Popper, K.P. *Objective Knowledge.* Oxford : The Clarendon Press, 1979.
15. Price, D. de S. "Network of Scientific Papers", *Science,* vol. 149, n° 3683, 1965, p. 510 - 515.
16. Price, D. de S. "The Citation Cycle", p. 269 in *Little Science, Big Science ... and Beyond.* New York, Columbia University Press, 1986.
17. Price, D. de S. "The Science- Technology Relationship, the Craft of Experimental Science, and Policy for the improvement of High TEchnology Innovation", *Research Policy,* vol. 13, 1984, p. 3 - 20.
18. Small H. and E. Garfield, "The Geography of Science : Disciplinary and National Mappings ", in *Science Citation Index 1988,* Philadelphia : Institut for Scientific Information, p. 46 - 58.
19. Turner, W. G. Charton, F. Laville, B. Michelet, "Packinging Information for Peer review : New Co-word Analysis Techniques", in A.F.J. van Raan (ed), *Handbook of Quantitative Studies of Science and Technology.* Amsterdam : Elsevier Science Publisher, 1988, p. 291 - 323.
20. Winston, P. H. *Artificial Intelligence.* London : Addison Wesley Publishing Co., 1977.

NATIONAL INSTITUTE OF SCIENCE, TECHNOLOGY & DEVELOPMENT STUDIES

Dr. K. S. Krishnan Road, New Delhi 110012 (INDIA)

Phones: 5714064, 5743227; **Telex**: 031-77182;
Telefax: 011-5754640; **E-mail**: nistads@sirnetd.ernet.in

NISTADS is a research Institute of the ***Council of Scientific and Industrial Research.*** It specialises in analysing issues of science and technology and their interface with society within a broad spectrum of economic, social, historical and cultural perspectives -- variously termed as science policy studies, science and technology studies, or science of science.

NISTADS provides consultancy services and undertakes sponsored studies in the above areas for national agencies, central and state government departments, industries, etc.; and international organisations like World Bank, UNESCO, WAITRO, IDRC, American Institute of Physics.

MAJOR RESEARCH AREAS

The research projects undretaken at NISTADS are essentially multi-disciplinary and carried out by teams drawn from the Institute's scientific staff (about 55) having qualifications and experience in diverse disciplines of natural sciences, engineering, mathematics, statistics, and disciplines of social sciences, management, etc. Research programmes, and likewise the Institute's expertise, are built around the following major themes, under each of which a large number of studies have been completed and published:

- **(a) Technology assessment and forecasting;**
- **(b) Resource planning and utilisation for regional development;**
- **(c) Application of S&T in rural areas;**
- **(d) Sociology of science;**
- **(e) Women and science and technology;**
- **(f) History and philosophy of science;**
- **(g) Science and technology indicators;**
- **(h) Research and training on R&D management;**
- **(i) Information systems and data-bases for S&T;**
- **(j) Science Communication;**

SERVICES OFFERED BY NISTADS

S&T Policy Studies & Analyses on : (i) Planning, organization and management of science; (ii) Social relations of science; (iii) Geographic Information Systems for planning at regional level; (iv) Technology assessment and forecasting; (v) Technology development, acquisition and transfer, (vi) Bibliometric studies (to determine research trends); (vii) Issues related to women in S&T.

Design/Setting up of Information Systems and Databases on : (i) S&T indicators; (ii) R&D project management; (iii) Management of natural resources for regional development; (iv) Bibliographic and non-bibliographic databases; (v) Software development : Fonts for Indian languages, multilingual graphic-user interface %wr databases and text.

Survey Studies on: (i) Impact assessment; (ii) Techno-economic surveys for rural industries;
(iii) Opinion surveys on S&T related issues.

Information Supply on: (i) S&T archival records; (ii) Literature on science policy; (iii) S&T Indicators for India and other developing countries; (iv) Research material for making films on S&T related public issues and on Indian scientists.

Training provided to scholars and government functionaries from India and other developing countries in areas of expertise of the Institute.

NISTADS' PUBLICATIONS

Books, Published Research Papers, Reports, Proceedings, Working Papers, 'List of Publications (1979-94)', Annual and subject lists can be supplied on request.

Current Literature on Science of Science: The monthly journal includes abstracts of papers(from about 250 international journals), digests of reports, book reviews and statistics in areas related to science policy studies. **Annual subscription:** Rs.150 (in India); £33 or $62 in developing countries; and £48 or $90 elsewhere.

Recent NISTADS' Publications in Scientometrics and Informetrics

1. Jain, Ashok; Pruthi, S. **'Indian S&T Indicators'.** (pages 275), Under publication.
2. Nagpaul, P.S. **'Scientometric Profile of Academic Science in India'.** (pages 335) Price Rs.335.00, US$67.00. Sponsored by Dept. of Science and Technology, GOI.
3. Pruthi, S. **'Opinion survey of Scientists on Peer Review System of SERC in India (Ist Report)'.** (pages 21) Price Rs.20.00, US$4.00.
4. Pruthi, S. **'Opinion Survey of Concerned Citizens on Draft - Paper for a New Technology Policy and Issues -- A comparison of R&D and Industry'.**
5. Nagpaul, P.S. **'Structural and Organizational Characteristics of Academic Science'.** (pages 190) Price Rs.190.00, US$38.00. Sponsored by Dept. of Science and Technology, GOI.
6. Nagpaul, P.S. **'Characteristics of Scientific Potential in Academia'.** (pages 500). Price Rs.500.00, US$100.00. Sponsored by Department of Science and Technology, Govt. of India.

The Stability of Social Sciences Citation Index® Journal Citation Reports® Data for Journal Rankings in Three Disciplines.

Thomas E. Nisonger
Indiana University, School of Library and Information Science, Bloomington, Indiana 47405, U.S.A.

The stability of journal rankings, based on *Social Sciences Citation Index Journal Citation Reports (SSCI JCR)* data, are investigated for the years 1980 through 1990 in library and information science, economics, and geography. Three data analysis techniques are used : 1) pairwise correlational analysis; 2) average movement in rank; and 3) overlap among top-ranked journals. Major findings are that : 1) rankings by impact factor are less stable than those based on total citations received; 2) discipline size influences the results; and 3) top-ranked journals exhibit more stability than middle or lower-ranked ones. It is concluded that one should not rely on a single year of *JCR* data for journal assessment purposes.

1. Introduction

Beginning with Gross and Gross's famous ranking of chemistry journals, [13] various types of citation data have been used to evaluate and rank journals in innumerable disciplines. Citation analysis represents one of the two most commonly used journal ranking methods, along with the subjective-based perceptions of experts. Journals have also been ranked by usage, number of articles produced, and diverse measures of cost-effectiveness, although these approaches are less frequently employed. An annotated bibliography of the last decade's major journal ranking studies is included in Nisonger. [28]

Journals rankings are of practical value to scholars for manuscript submission decisions and promotion-and-tenure dossier evaluation purposes as well as to librarians who must make journal subscription and cancellation decisions. A list of a discipline's top-ranked journals can be especially valuable for serials collection management in libraries because these journals are often considered the core.

Journal rankings can be used from either a macro or a micro perspective. A macro-use examines the relative rankings of the journals in a particular discipline, while a micro-use would focus on the status of a specific journal. Broadus (pro) [4,5] and Line (con) [23] have debated the utility of *JCR* journal rankings to libraries.

The *Social Science Citation Index Journal Citation Reports* tabulates citation data conerning approximately 1,400 titles, based on the citations these journals received in roughly *6,500 titles covered by the SSCI, the Science Citation Index, and the Arts & Humanities Citation Index. [40] The JCR* offers separate rankings by five criteria : 1) total citations for all years; 2) impact factor; 3) immediacy index; 4) source items published in the current year; and 5) citations in the current year to issues of the journal published during the two proceding years, i.e., 1992 citations to 1991 and 1990. A new set of data is provided by the *JCR* each year.

Literature Review

Social Sciences Citation Index or Science Citation Index Journal Citation Reports data have been used to rank or rate journals in numerous subjects, including physics by Inhaber, [14] acoustics by Cawkell, [7] public administration by Colson, [9] social science psychology by Feingold, [11] reading research by Summers, [35] economics by Liebowitz and Palmer, [20] sociology and political science by

Christenson and Sigelman, [8] political science by Nisonger, [29] psychology by Buffardi and Nichols [6] plus Rushton and Roediger, [30] and criminology and criminal justice by Stack. [34] Garfield has provided citation rankings in countless subject areas. [41] Furthermore, *JCR* citation data has been incorporated in a number of journal selection models for libraries, including those by Kraft and Polacsek [17] plus Bennion and Karschamroon. [2]

Previously published perception-based journal rankings or ratings in the disciplines under study should be briefly noted, including Malouin and Outreville [24] in economics, two rankings in geography by Lee and Evans [18, 19] as well as Kohl and Davis, [16] Blake, [3] Tjoumas, [36] and Tjoumas and Blake [37] in library and information science. Kim [15] used *JCR* data to further analyze the findings of Kohl and Davis. Altuna Esteibar and Lancaster [1] offered alternate rankings of library and information science journals utilizing citations in course reading lists, doctoral dissertations, and faculty publications. Rankings based on *JCR* data have been compared with usage-based rankings by Scales [31] and Line [22] and, for physics journals, with usage and productivity rankings by Singleton. [33]

Most citation-oriented journal rankings have been based on impact factor, which, although subject to controversy, is generally acknowledged to be the most valid citation measure for journal evaluation purposes. See, for example, Buffardi and Nichols, [6] Dometrius, [10] and Vinkler [38] to mention only a few. However, numerous studies have used total citations received rather than impact factor, especially prior to the *JCR*. The use of impact factor for purposes other than journal ranking is beyond this paper's scope.

Problem Statement and Conceptual Framework

Most of the previously noted studies have utilized only a few years of *JCR* data. Thus, the question naturally arises concerning how consistent or stable the *JCR* journal rankings are from year to year, and to what extent these rankings would vary if different years had been used. The question of stability is not only of theoretical interest, but it also has considerable implications concerning use of journal citation rankings for practical purposes. If a journal ranked 4th within its discipline in 1989 and 32th in 1990, one could not confidently rely on either year's ranking as a reliable measure of the journal's true status. Thus, high stability implies that data from one or two years can be used with confidence. Medium or low stability would indicate that data from a broader range of years is necessary.

The stability of *SSCI* citation data has not been systematically investigated, although a number of researchers have previously examined the issue. Buffardi and Nichols found a .91 correlation ($p < .01$) between the 1977 and 1978 *SSCI JCR* impact factors for a set of 96 psychology journals and a .84 correlation ($p < .01$) between the 1975 and 1978 impact factors for 70 psychology journals. [6] A Study of 59 sociology journals by Gordon revealed a .974 correlation between 1977 and 1978 total citations and a .844 correlation between 1977 and 1978 impact factor (both $p < .001$). [12] Sievert's analysis of *Online* and *Online Review* compared both *JCR* immediacy index and impact factor for 1979 through 1986. More significantly, rank within their discipline, as indicated by *JCR*, during these years ranged from 1 in 1981 to 8 in 1982 for *Online* and 1 in 1980 to 26 in 1984 for *Online Review* (which rebounded to 2 in 1986). [32] A few studies have included longitudinal comparisons of impact factor scores without offering explicit rankings. Lindsey exhibited the *JCR* impact factors for 48 sociology journals from 1981 through 1988, [21] while Vlachy plotted the impact factor trends of approximately 250 physics journals from 1974 to 1983. [39]

This study's objective is to analyze the stability of journal rankings, based on *SSCI JCR* citation data, in three disciplines : economics, library and information science, and geography. The primary focus will fall on ranking by impact factor, while ranking by total citations will also be examined. The data will be analyzed from 1980 through 1990, with ranking comparisons made at yearly intervals.

To control for the impact of discipline size (i.e., the number of journals covered in the *SSCI JCR*), economics was chosen as a large discipline, library and information science to represent a medium-sized discipline, and geography as a small subject area. The journals listed under these disciplines in section 8 of the *JCR*, "*SSCI* Journals by Category Ranked by Impact Factor" for the years 1980 through 1990 formed the basis of the analysis. [42] (The number of covered journals ranged from 103 to 130 for economics; 48 to 57 for library and information science; and 21 to 28 for geography.)

Methodology

For practical decision-making purposes by both librarians and scholars a journal's ordinal rank within its discipline is the crucial variable rather than the absolute value of the impact factor or total citations received. Thus, the primary focus in this study will be on the rankings rather than the raw data. It is generally known that citation patterns vary among disciplines and subject areas. Consequently, a journal's raw citation scores are exceedingly difficult to evaluate unless placed in a comparative context with similar journals. While it is true that notable year-to-year shifts in rank could reflect minor changes in the absolute data, a full investigation of this issue is beyond the study's scope. It should be noted, however, that one of the three techniques used in this analysis (Pearson's Product Movement Correlation) is based on underlying data rather than rank.

The methodological literature concerning journal ranking has focused on the benefits and drawbacks of various approaches, e.g., subjective opinion of experts versus citation-based rankings or impact factor versus some other citation measure. Little or nothing has been published on methodology for investigating overall stability in journal rankings. Two general strategies are available. The first, as used by Gordon and Buffardi and Nichols as noted above, may be termed a "year-by-year" approach. The focus is on changes between years in the rankings within a set of journals, typically the journals in a particular discipline. The second strategy, employed by Sievert, Vlachy, and Lindsey, could be called a "journal-by-journal" approach as variation in the ranking of specific journals over a range of years is analyzed.

This paper uses the "year-by-year" approach to analyze the consistency of *SSCI JCR* rankings for economics, geography, and library and information science from 1980 to 1990. Three specific techniques are used : 1) correlational analysis; 2) average movement in the rankings; and 3) overlap among the top ten and top twenty ranked journals. Due to the exceedingly high year to year correlations for rankings based on total citations received (see the "Results" section) and the fact that most recent citation-based journal ranking studies have used impact factor, techniques 2 and 3 have been applied only to impact factor rankings.

Calculation of the average year-to-year movement in rank for each title in the set represents a modification of an approach employed by Mankin and Bastille, who analyzed two different methods for ranking journals by usage at the Masachusetts General Hospital Library. [25] The movement in rank between years was figured for each journal common to the two years. The direction of movement was not considered in the calculation, as it is irrelevant to the issue under analysis. A journal moving from 10 to 13 displays equal stability with one moving from 10 to 7. Ties were prorated by dividing the total of their ranks by the number of tied positions, so that two journals tied for 6th and 7th place were both counted as 6.5. (i.e., 13 / 2) or journals in a three-way tie for 49, 50, or 51 were each ranked as 50 (i.e., 150 / 3). A title change was considered to result in a different journal.

After the annual mean and median movement in rank were calculated, the results were also normalized for size to correct for disparities among the three disciplines and the fact that within a discipline the number of journals covered in *JCR* varied from year to year. [43] The following formula illustrates the correction for the mean 1980 to 1981 movement of economics journals :

$$\frac{\text{1980 / 1981 Mean Movement}}{\text{No. of Journals Covered in 1981}} \times 100$$

$$= \frac{10.6}{103} \times 100 = 10.29$$

The result represents a normalized mean or median movement as a percentage of the discipline's journals covered in *JCR*.

Analysis of overlap among the top journals is modeled upon the approach used by Scales, [31] and Line, [22] both of whom analyzed overlap among the most frequently requested journals at the British Library Lending Division and the most highly cited journals in the *SCI* and *SSCI*. The method is self-evident and requires no explanation.

Tests of statistical significance are not appropriate for this study because *JCR* journal coverage is not based on randomly selected samples. Also, the top ten and top twenty journals (for which stability is analyzed in Table 3) do not represent independent sets. Consequently, interpretation of the results is based primarily on visual observation of the data.

Nieuwenhuysen [26] and Nieuwenhuysen and Rousseau [27] have proposed methods for estimating the random yearly fluctuation in raw impact factor scores. This present study focuses on overall stability in the rankings without addressing the extent to which score fluctuations represent random variation, genuine changes in status, or, possibly, errors or quirks in the *JCR*.

Limitations

A number of general limitations to journal rankings based on *SSCI JCR* citation data must be acknowledged. The data tabulates citations from all 6,500 journals in the ISI database, not just from titles in the same discipline, so a journal's centrality to its discipline is not measured. Also, not all journals are included in the *SSCI JCR*. Citation analysis does not measure a journal's value for functions other than communicating research, such as current awareness or support of teaching. The quality of individual articles in a specific journal title may vary immensely. While it is generally understood that citation data can not be used for comparing journals in different disciplines, some observers feel they can not be used for valid comparisons among different types of journals or different subfields within the same discipline. Moreover, a journal's citation record is only one of many significant assessment variables.

This specific research project is complicated by the number of idiosyncracies in the *SSCI JCR*. For example, 1) the journals covered in *JCR* vary from year to year; 2) the *JCR's* assignment of journals to subject categories is sometimes questionable; and 3) even journals that are covered by the *JCR* are sometimes omitted from the "Journals by Category Ranked by Impact Factor" section. The analysis is further confounded by changes in journal names. Moreover, no standard exists concerning what would constitute an acceptable level of stability.

Results

The year-to-year correlational analysis results are displayed in Table 1. [44] Two conclusions, applying to all three disciplines, emerge : 1) the correlations for rankings based on total citations are extremely high; and 2) the correlations for impact factor based rankings, while also generally high, are less strong than for rankings structured on total citations received. In all 30 data sets (10 years for 3 disciplines) the year-to-year correlations for total citations were higher than the impact factor correlations. Accordingly, the mean 10 year correlation for total citations versus impact factor was .9960 vs. 9072 in economics, .9363 vs. .7621 in library and information science, and .9608 vs. .7063 in geography. The combined 10 year mean for all disciplines was .9644 for total citations, compared to .7919 for impact factor. It is also noteworthy that for each discipline the variation in the correlations among years was greater for impact factor than total citations.

It initially seems a paradox that rankings based on total citations offer much greater stability than impact factor rankings, because the latter is generally considered to be the more valid citation measure. Yet intuitively this phenomenon can easily be understood. Impact factor is based on citations made to a journal's two previous years, [45] whereas the "total citations" figure contains citations made to a journal's entire historical run. Consequently, a change in the number of citations made to either of the two previous years would cause a greater fluctuation in impact factor than in "total citations."

To address briefly a tangential issue, in all three disciplines the correlations between the impact factor and the "total citation" rankings for the same year were much lower than the year to year correlations for either impact factor or total citations. The mean yearly correlation between impact factor and total citation rankings for the eleven years between 1980 and 1990 was .5103 in economics, .6353 in geography, and .5127 in library and information science. These findings are intuitively unsurprising and could be interpreted as reaffirming the need for impact factor as a separate citation measure.

Table 2 summarizes the year-to-year movement for each title in impact factor rankings from 1980 to 1990 for the three disciplines. The mean, mean corrected for size, median, median corrected for size, and standard deviation are displayed.

The mean movement ranged from 10.6 (in 1980/81) to 15.46 in (1983/84) for economics; 6.33 (in 1981/82) to 9.48 (in 1982/83) in library and information science; and 2.75 (in 1987/88) to 5.68 (1989/90) for geography. In all cases but one (1985/86 for geography) the mean exceeds the median, reflecting the fact that a number of extreme cases skewed the distribution to the right.

It is apparent that a discipline's size (operationally defined as the number of journals covered in the *JCR*) does influence the results. The absolute ten-year mean and median movement in rank was highest in the largest discipline, economics, (12.6 and

Table 1 Pairwise Year-to-Year Correlations for *SSCI - JCR®* Journal Rankings in three Disciplines, 1980-1990.*

ECONOMICS

Years	Total Citations	Impact Factor
1980/81	.9856	.9422
1981/82	.9980	.9416
1982/83	.9976	.8088
1983/84	.9967	.8370
1984/85	.9966	.9123
1985/86	.9974	.8671
1986/87	.9978	.9538
1987/88	.9976	.9343
1988/89	.9987	.9330
1989/90	.9942	.9416
10 Year Mean	.9960	.9072

LIBRARY AND INFORMATION SCIENCE

Years	Total Citations	Impact Factor
1980/81	.9296	.9020
1981/82	9448	.7697
1982/83	.8689	.6250
1983/84	.9353	.7148
1984/85	.9369	.7511
1985/86	.9240	.8254
1986/87	.9359	.7259
1987/88	.9596	.7902
1988/89	.9574	.7374
1989/90	.9708	.7793
10 Year Mean	.9363	.7621

GEOGRAPHY

Years	Total Citations	Impact Factor
1980/81	.8786	.6820
1981/82	.9728	.7474
1982/83	.9432	.4524
1983/84	.9469	.6895
1984/85	.9796	.5679
1985/86	.9786	.8308
1986/87	.9768	.7303
1987/88	.9817	.8799
1988/89	.9794	.9023
1989/90	.9704	.5801
10 Year Mean	.9608	.7063

* All correlations are statistically significant ($p < .01$).

Table 2 Average Year-to Year Movement in Impact Factor Journal Rankings, 1980-1990.

ECONOMICS

Years	Mean	Cor. Mean	Median	Cor. Median	Stan. Dev.
1980/81	10.6	10.29	9	8.74	9.65
1981\82	12.73	11.07	10	8.70	10.77
1982/83	11.55	9.55	10	8.26	10.94
1983/84	15.46	12.57	10	8.13	18.04
1984/85	12.83	10.02	8	6.25	14.46
1985/86	14.03	11.05	9	7.09	14.77
1986/87	11.16	9	8	6.45	10.57
1987/88	11.34	9.07	7	5.60	12.58
1988/89	14.4	11.08	10	7.69	15.55
1989/90	11.94	9.18	10	7.69	10.78
Mean	12.6	10.29	9.1	7.46	12.81

LIBRARY AND INFORMATION SCIENCE

Years	Mean	Cor. Mean	Median	Cor. Median	Stan. Dev.
1980/81	6.72	14	6	12.5	5.83
1981\82	6.33	12.17	5	9.62	5.35
1982/83	9.48	16.93	7	12.5	8.37
1983/84	8.66	16.04	5	9.26	8.33
1984/85	9.27	17.17	8	14.81	7.99
1985/86	8.8	15.44	5.5	9.65	8.9
1986/87	8.66	15.19	7.5	13.16	7.1
1987/88	7.23	13.39	6	11.11	5.89
1988/89	9.22	17.07	6.25	11.57	8.27
1989/90	8.3	15.09	4.5	8.18	7.96
Mean	8.27	15.25	6.08	11.24	7.4

GEOGRAPHY

Years	Mean	Cor. Mean	Median	Cor.Median	Stan. Dev.
1980/81	3.2	15.24	2	9.52	3.97
1981\82	3.42	17.1	2	10	3.32
1982/83	4.21	18.3	3	13.04	4.18
1983/84	3.86	16.8	3	13.04	2.93
1984/85	3.39	13.56	2	8	3.14
1985/86	4	14.81	4	14.81	2.74
1986/87	4	16	3	12	3.28
1987/88	2.75	11	1	4	2.91
1988/89	3.6	14.4	3	12	3.48
1989/90	5.68	20.29	4	14.29	4.98
Mean	3.81	15.75	2.7	11.07	3.49

Table 3 : Mean Year-to-Year Movement of Top Ranked Journals*

ECONOMICS

Years	Top 10	Top 20	Total
1980/81	1.9	4	10.6
1981\82	1.9	4.06	12.73
1982/83	1.13	3.21	11.55
1983/84	1.1	4.45	15.46
1984/85	3.4	5.37	12.83
1985/86	1.3	7.3	14.03
1986/87	3.5	4.1	11.16
1987/88	1.7	4.47	11.34
1988/89	1.11	4.18	14.4
1989/90	1.25	4.94	11.94
Mean	1.83	4.61	12.6

LIBRARY AND INFORMATION SCIENCE

Years	Top 10	Top 20	Total
1980/81	2.22	4.44	6.72
1981\82	4.38	4.44	6.33
1982/83	5.7	8.7	9.48
1983/84	2.9	6.7	8.66
1984/85	8.38	9.76	9.27
1985/86	4.1	8.63	8.8
1986/87	6.6	6.67	8.66
1987/88	8.5	7.6	7.23
1988/89	8.4	9.48	9.22
1989/90	8.6	6.48	8.3
Mean	5.98	7.3	7.4

GEOGRAPHY

Years	Top 10	Top 20	Total
1980/81	2.5	3.21	3.2
1981\82	2.89	3.42	3.42
1982/83	4.4	4.25	4.21
1983/84	3.67	3.47	3.86
1984/85	2.78	3.11	3.39
1985/86	3.38	3.77	4
1986/87	2.9	4.4	4
1987/88	1.8	2.95	2.75
1988/89	1.4	3.3	3.6
1989/90	4.67	5.16	5.68
Mean	3.04	3.7	3.81

* The data has been adjusted for cases in which journals in the top ten or twenty were not included in the previous year's *JCR* ranking.

Table 4 : Year-to-Year Overlap Among Top 10 and Top 20 Journals

ECONOMICS

Year	Top 10 No. Common To Both Years	Top 10 Percentage	Top 20 No. Common To Both Years	Top 20 Percentage
1980/81	7	70	16	80
1981\82	9	90	15	75
1982/83	9	90	14	70
1983/84	9	90	16	80
1984/85	7	70	14	70
1985/86	10	100	18	90
1986/87	9	90	18	90
1987/88	8	80	17	85
1988/89	9	90	14	70
1989/90	8	80	14	70
10 Year Mean	8.5	85	15.6	78

LIBRARY AND INFORMATION SCIENCE

Year	Top 10 No. Common To Both Years	Top 10 Percentage	Top 20 No. Common To Both Years	Top 20 Percentage
1980/81	8	80	12	60
1981\82	7	70	16	80
1982/83	6	60	15	75
1983/84	8	80	12	60
1984/85	6	60	12	60
1985/86	9	90	13	65
1986/87	7	70	16	80
1987/88	7	70	11	55
1988/89	5	50	13	65
1989/90	7	70	16*	80
10 Year Mean	7	70	13.6	68

GEOGRAPHY

Year	Top 10 No. Common To Both Years	Top 10 Percentage	Top 20 No. Common To Both Years	Top 20 Percentage
1980/81	8	80	19	95
1981\82	7	70	18	90
1982/83	8	80	16	80
1983/84	7	70	16	80
1984/85	7	70	17	85
1985/86	8	80	17	85
1986/87	8	80	18	90
1987/88	10	100	16	80
1988/89	10	100	17	85
1989/90	7	70	18	90
10 Years Mean	8	80	17.2	86

* Includes two titles in tie for 20th place in 1990.

9.1 respectively) and lowest in geography, the smallest discipline (3.81 and 2.7, respectively). This result would be expected based on logic and intuition because a large discipline allows more opportunities for a journal's ranking position to fluctuate between years.

However, with a ten-year corrected mean of 10.29 and a corrected median of 7.46, economics displayed the smallest movement relative to number of journals in the discipline, when the results are normalized for size. To what extent this finding reflects the size or other characteristics of the discipline is unclear. The difference between library and information science and geography in the corrected ten-year mean (15.25 versus 15.75) and median (11.24 versus 11.07) are essentially trivial.

This issue of absolute versus relative movement in rank poses a basic philosophical question without an obvious empirically-based answer. Is a movement of 10 places within a set of 100 titles more or less significant than moving 5 positions in a discipline containing 20 journals? The answer may depend on the purpose for which the ranking is being used.

Table 3 analyzes the mean movement of the top ten and top twenty journals in the three disciplines. (Analysis of the top twenty geography journals is obviously less significant because of the smaller number of *JCR*-covered titles in that area.) It is readily apparent that highly ranked journals exhibit greater stability, i.e., they move a smaller number of ranking positions from year to year. In all three disciplines the ten-year mean movement of the top ten and top twenty journals was smaller than for the whole discipline, while the top ten was also more stable than the leading twenty. The difference was particularly pronounced in economics, where the top ten moved an average of 1.83 positions during the decade and the top twenty 4.61, compared to 12.6 for the entire discipline. This may reflect the fact that the leading ten and twenty journals represent a higher percentile in economics than in the other two disciplines, due to economics' larger size. While the finding that high ranking journals exhibit more stability may be expected based on intuition, it has practical implications for the use of *JCR*. Moreover, it would tend to confirm Line's speculation concerning the greater instability of lower-ranked titles. [23]

Table 4 tabulates the year-to-year overlap among the top ten and top twenty titles. In economics, overlap among the top ten ranged from 7 to 10 with a 8.5 mean, and, among the leading twenty, 14 to 18 with a mean of 15.6, while the figures for geography ranged from 7 to 10 with a mean of 8 and 16 to 19 with a 17.2 mean, respectively. It is not immediately clear why the overlap was lower in library and information science where the mean was 7 for the top ten (ranging from 5 to 9) and 13.6 for the leading twenty, ranging from 12 to 16.

In percentage terms the top ten were more stable than the top twenty in economics (85% overlap compared to 78%) and library and information science (70% versus 68%), but not in geography (80% contrasted to 86%). (As noted above, the concept of the "top twenty" is not particularly meaningful in geography because of the small number of its journals included in the analysis.) This finding confirms the greater stability of the top ten versus the top twenty revealed in Table 3, except the tendency is less distinct.

In a practical sense, these findings indicate that anyone using the *JCR* impact factor rankings from a macro perspective to identify a discipline's leading or core journals would find roughly a 20% to 30% change from one year to the next.

Conclusions and Implications

Although precise decision rules have yet to be developed, it seems apparent that one can not reach valid journal assessments for decision-making purposes based on a single year of *SSCI JCR* impact factor data, because the rankings can vary substantially from year to year. This statement applies to both micro-decisions about a specific title's status and macro-decisions that identify a discipline's top-ranked or core journals. When reaching micro-decisions concerning specific titles, the higher the title's ranking the greater confidence one can place in the year-to-year consistency of its status. Thus, special caution (i.e., less reliance on a few years of data) is advised for middle and lower-ranked journals, especially titles that belong to large disciplines.

These findings have varying implications for different uses of *JCR* data. The impact factor ranking s undoubtedly have more reliability for identification of core journals and journal selection decisions by scholars and librarians (where prime attention would be paid to highly ranked journals) than for journal cancellation decisions in libraries, which would tend to focus on titles in the lower rank positions.

Questions for further research include : 1) to what extent do ranking variations represent random fluctuations, significant underlying trends, or errors or quirks in the *JCR;* 2) what decision rules can be developed for effective utilization of *SSCI JCR* data for journal ranking; 3) besides such self-evident factors as the ranking itself and the number of citations received, what variables, e.g., age, type of publisher, country of origin, etc., are associated with high volatility in a journal's ranking by impact factor; 4) how frequently and in what way do scholars and librarians use journal rankings generally and, specifically, citation-based journal ranking or *JCR* data, in their decision-making processes; 5) how stable are the journal half-life data reported each year in *JCR* ; 6) how would comparisons at two and five year intervals differ from one-year interval comparisons; 7) how would analysis of leading journals in terms of percentiles, e.g., the top 25% differ from analysis in terms of absolute numbers, i.e., the top ten and twenty; 8) what insights would be provided by longitudinal tracking of specific titles over the ten year range ? Finally, it would be useful to replicate the study in other disciplines, especially the scientific areas covered in the *Science Citation Index JCR* .

Acknowledgments : The author gratefully acknowledges valuable input from Blaise Cronin, Dean, Stephen P. Harter, Professor, and Judith Serebnick, Associate Professor, at Indiana University's School of Library and Information Science, and Michael E.D. Koenig, Dean, Rosary College, Graduate School of Library & Information Science. Claire Nisonger and Cindy Pierard, my graduate assistant, are thanked for helping gather and tabulate data. Stephen McKinley of the Indiana University Statistical Consulting Service is thanked for valuable statistical advice.

References and Notes

1. Altuna Esteibar, B. & F.W. Lancaster, "Ranking of Journals in Library and Information Science by Research and Teaching Relatedness," *Serials Librarian* 23 (1992) No. 1/2, 1 - 10.
2. Bennion, B.C. & . Karschamroon, Multivariate Regression Models for Estimating Journal Usefulness in Physics, *Journal of Documentation* 40 (1984), NO. 3, 217 - 227.
3. Blake, V.L.P., In the Eyes of the Beholder : Perceptions of Professional Journals by Library / Information Science Educators and District School Library Media Coordinators, *Collection Management,* 14 (1991) No. 3/4, 101 - 148.
4. Broadus, R.N., A Proposed Method for Eliminating Titles from Periodical Subscription Lists, *College & Research Libraries,* 46 (1985) No. 1, 30 - 35.
5. Brodus, R.N., On Citations, Uses, and Informed Guesswork: A Response to Line, *College & Research Libraries,* 46 (1985) No. 1, 38 - 39.
6. Buffardi, L.C. & J.A. Nichols, Citation Impact, Acceptance Rate, and APA Journals, *American Psychologist,* 36 (1981) No. 11, 1453 - 1456.
7. Cawkell, A.E., Evaluating Scientific Journals with *Journal Citation Reports* A Case Study in Acoustics, *Journal of the American Society for Information Science,* 29 (1978) No.1, 41 - 46.
8. Christenson, J.A. & L. Sigelman, Accrediting Knowledge : Journal Stature and Citation Impact in Social Science, *Social Science Quarterly,* 66 (1985) No. 4, 964 - 75.
9. Colson, H., Citation Rankings of Public Administration Journals, *Administration & Society,* 21 (1990) No. 4, 452 - 71.
10. Dometrius, N.C., Subjective and Objective Measures of Journal Stature, *Social Science Quarterly,* 70 (1989) No. 1, 202.
11. Feingold, A., Assessment of Journals in Social Science Psychology, *American Psychologist,* 44 (1989) No. 6, 961-64.
12. Gordon, M. D., Citation Ranking versus Subjective Evaluation in the Determination of Journal Hierarchies in the Social Sciences, *Journal of the American Society for Information Science,* 33 (1982) No. 1 55-57.
13. Gross, P. L. K. & E. M. Gross, College Libraries and Chemical Education, *Science,* 66 (1927) No. 1713, 385-389.
14. Inhaber, H., Is There a Pecking Order in Physics Journals? *Physics Today,* 27 (1974) No. 5, 39-43.
15. Kim, M. T., Ranking of Journals in Library and Information Science : A Comparison of Perceptual and Citation-Based Measures, *College & Research Libraries,* 52 (1991) NO. 1, 24-37.
16. Kohl, D. F. & C. H. Davis, Ratings of Journals by ARL Library Directors and Deans of Library and Information Science Schools, *College & Research Libraries,* 46 (1985) No. 40-47.
17. Kraft, D. H. & R. A. Polacsek, A Journal-Worth Measure for a Journal-Selection Decision Model, *Collection Management,* 2 (1978) No. 2, 129-139.
18. Lee, D. & A. Evans, American Geographers' Rankings of American Geography Journals, *Professional Geographer,* 36 (1984) No. 4, 292-300.
19. Lee, D. & A. Evans, Geographers' Rankings of Foreign Geography and Non-Geography Journals, *Professional Geographer,* 37 (1985) NO. 4, 396-402.
20. Liebowitz, S. J. & J. P. Palmer, Assessing the Relative

Impacts of Economics Journals, *Journals of Economic Literature,* 22 (1984) No. 1 77-88.

21. Lindsey, D., Precision in the Manuscript Review Process : Hargens and Herting Revisited; *Scientometrics,* 22, (1991) No. 2, 320.
22. Line, M. B., Changes in Rank LIsts of Serials Over Time : Interlending versus *Citation Data, College & Research Libraries,* 46 (1985) No. 1, 77-79.
23. Line, M. B., Use of Citation Data for Periodicals Control in Libraries : A Response to Broadus, *College & Research Libraries,* 46 (1985) No. 1, 36-37.
24. Malouin, J. L. & J. F. Outreville, The Relative Impact of Economics Journals — A Cross Country Survey and Comparison, *Journal of Economics & Business,* 39 (1987) No. 3, 267-77.
25. Mankin, C. J. & J. D. Bastille, An Analysis of the Differences Between Density-of-Use Ranking and Raw-Use Ranking of Library Journal Use, *Journal of the American Society for Information Science,* 32 (1981) No. 3, 224-228.
26. Nieuwenhuysen, P., Journal Citation Measures : Taking into Account Their Fluctuations from Year to Year, *Journal of Information Science,* 15 91989) No. 3, 175-178.
27. Nieuwenhuysen, P. & R. Rosseau, A Quick and Easy Method to Estimate the Random Effect on Citation Measures, *Scientometrics,* 13 (1988) No. 1/2, 45-52.
28. Nisonger, T. E., *Collection Evaluation in Academic Libraries : A Literature Guide and Annotated Bibliography,* Englewood, CO : Libraries Unlimited, 1992.
29. Nisonger, T. E., A Ranking of Political Science Journals Based on Citation Data, *Serials Review,* 19 (1993) No. 4, 7-14.
30. Rushton, J. P. & H. L. Roediger, III, An Evaluation of 80 Psychology Journals Based on the *Science Citation Index, American Psychologist,* 33 (1978) No.5, 520-523.
31. Scales, P. A., Citation Analyses as Indiacators of the Use of Serials : A Comparison of Ranked Title Lists Produced by Citation Counting and from Use Data, *Journal of Documentation,* 32 (1976) No. 1, 17-25.
32. Sievert, M. C., Ten Years of the Literature of Online Searching : An Analysis of Online and Online Review, *Journal of the American Society for Information Science,* 41 (1990) No. 8, 560-68.
33. Singleton, A., Journal Ranking and Selection : A Review in physics, *Journal of Documentation,* 32 (1976) No. 1, 258-289.
34. Stack, S., Measuring the Relative Impacts of Criminology and Criminal Justice Journals : A Research Note, *Justice Quarterly,* 4 (1987) No. 3, 475-84.
35. Summers, E. G., A Review and Application of Citation Analysis Methodology to Reading Research Journal Literature, *Journal of the American Society for Information Science,* 35 (1984) No. 6, 332-343.
36. Tjoumas, R., Professional Journal Utilization by Public LIbrary Directors, *Serials LIbrarian,* 20 (1991) No. 2/3, 1-16.
37. Tjoumas, R. & V. L. P. Blake, Faculty Perceptions of the Professional Journal Literature : Quo Vadis? *Journal of Education for Library and Information Science,* 33 (1992) No. 3, 173-194.
38. Vinkler, P., Bibliometric Features of Some Scientific Subfields and the Scientometric Consequences Therefrom, *Scientometrics,* 14 (1988) No. 5/6, 453.
39. Vlachy, J., Trends in the Citation Impact of Physics Journals, *Czechoslovak Journal of Physics,* B35 (1985) No. 2, 187-190.

40. The *Journal Citation Reports* was included automatically as the last bound volume in the annual *Social Sciences Citation Index set,* until 1989 when it began being issued on microfiche and had to be purchased separately.
41. Many of these, originally publiished in *Current Contents,* have ben reprinted in E. Garfield, *Essays of an Information Scientist,* 9 vols., Philadelphia : ISI Press, 1977-1988.
42. A few library science journals such as *Library Quarterly and Library Journal,* that are covered by the *SSCI JCR* but were omitted from this section in the early 1980s, were also included in the analysis.
43. The number of journals covered in *JCR* increased throughtout the 1980s in all three disciplines.
44. Pearson Product Movement correlations were calculated with KWIKSTAT, version 1.3.
45. A journals's 1993 ISI impact factor would be calculated by the following formula :

$$\text{Impact Factor} = \frac{\text{1993 Citations to 1992+1991}}{\text{No.of 1992+1991 Citable Items}}$$

*Periodicity in Academic Library Circulation A Spectral Analysis**

William E. McGrath
School of Information and Library Studies
State University of New York at Buffalo
Buffalo, New York 14260, USA

Spectral analysis was used to identify periods in four *circulation* series of 1032 days and four of *pickup* (3073 days) in an academic library. Five observed periods were compared to theoretical periods in a 364-day academic calendar : two were expected, 182 and 7 days; one (122 days) was interpreted as a 112-day semester plus an interim of about 10 days ; two were unexpected (3.5 and 2.3 days). The 182-day (about half an academic year) was strongest in *pickup*. The 7-day (the basic scheduling period) was strong in all series. The 3.5 day was strong in most series. The 2.3 day, found only in *circulation,* is close to every-other-day usage or about 3 cycles per week, and may be due to Monday-Wednesday-Friday class scheduling, or it may be spurious. Only the 182-day period was visible in time charts. No evidence was found for a monthly period, the calendar unit in which library time series are often recorded.

* This paper in full is being published elsewhere

International Society for Scientometrics and Informetrics (ISSI)

From 11 to 15 September 1993, the 'Fourth International Conference on Bibliometrics,Informetrics and Scientometrics' was held in Berlin, Germany. The meeting, which was dedicated to the memory of Derek de Solla Price, the founding father of our field of research, has been a great success. The fact that it was attended by 189 participants from 33 countries shows that this emerging scientific field is gaining importance and worldwide recognition. In order to further stimulate this development, the program committee has decided to found the International Society for Scientometrics and Informetrics (ISSI).

After a period of preparations, the new Society was officially founded on 5 October 1994 in Utrecht, The Netherlands.

ISSI's goals are the advancement of theory, method and explanation of the following areas :

1. Quantitative studies of :
   - scientific, technological and other scholarly and substantive information;
   - the science of science and technology, social sciences, arts and humanities;
   - generation, diffusion and use of information;
   - information systems, including libraries, archives and databases.
2. Mathematical, statistical and computational modelling and analysis of information processes.

In order to achieve them, the organization directs its activities at:

1. Communication and exchange of professional information;
2. Improving standards, theory and practice in all areas of the discipline;
3. Education and training;
4. Enhancing the public perception of the discipline.

In recognition of her expertise and dedication, Dr. Hildrun Kretschmer, organizer of the Berlin conference, has been chosen to be the first president of ISSI.

Secretary-treasurer is Dr. C. le Pair, Technology Foundation (STW),P.O. Box 3021, 3502 GA Utrecht, The Netherlands.

Meanwhile, plans for the other International Conferences took shape.

In the following page you will find an ISSI-membership application form.

ISSI Membership Information

If you are active in the field of Scientometrics or Informetrics, you can become a member of the newly-founded Society. For a membership fee of $ 20.-, you will recieve early information on activities in the field through regularly appearing newsletters.

In view of the fact that many researchers in our field live in developing countries, the ISSI Board has decided on a special, reduced fee of $ 5.— for people from those countries.

JISSI is issued quarterly. Correspondence regarding advertishment and subscription should be made to the publisher. Payments may be sent by Cheque/Bank Draft/Money order.
For cheques outside Delhi but from within India a sum of Rs 15/- should be added. For cheques out side India a sum of U.S. dollar 2.00 or equivalent should be added.
Claims for missing number(s) should be made within four months of the date of issue of such number to the Publisher.

Cheques/Demand Drafts should be drawn in favour of '**Brzark Information Systems (P) Ltd**'.

The order form may please be filled up and sent to :

Brzark Information Systems (P) Ltd,
112, Humayun Pur (Near N.C.C. Office)
Safdarjung Enclave. New Delhi - 110029.

Subscription Rates (All rates include airmail surcharges)

Rates for Canada,USA,Europe,Japan,Singapore,Australia, New Zealand & Middle East:

INSTITUTIONS (including LIBRARIES)		
Annual	....	US $200
For first three years	....	US $400
INDIVIDUALS		
Annual	....	US $100
For first three years	....	US $200
ISSI MEMBERS		
Annual	....	US $75
For first three years	....	US $150

Special Offer for Other (Developing) Countries:

INSTITUTIONS		
Annual	....	US $125
For first three years	....	US $250
INDIVIDUALS		
Annual	....	US $50
For first three years	....	US $100
ISSI MEMBERS		
Annual	....	US $30
For first three years	....	US $60

Rates for Indian Subcontinent:

INSTITUTIONS		
Annual	....	Rs2500
For first three years	....	Rs5000
INDIVIDUALS		
Annual	....	Rs1000
For first three years	....	Rs2000
ISSI MEMBERS		
Annual	....	Rs750
For first three years	....	Rs1500

ISSI Membership Application Form

FAMILY NAME :.. MALE/FEMALE*.......................

FIRST NAME(s) :.. TITLE(s) :...............................
....

INSTITUTION:.. FUNCTION:...............................
....

ADDRESS: ....
....
....
....

PHONE : .... EMAIL : ....

TELEFAX : ....

MEMBERSHIP FEE : $20.00
$ 5.00 (for people from developing countries)

CHECK ENCLOSED/AMOUNT REMITTED TO : *) ABN ALN 2A
ABN-AMRO Bank Utrecht
acc.no. 55.57.54.855.STW

*) please cross out which does not apply

Please return the completed form to : ISSI Administration
c/o Technology Foundation(STW)
PO Box 3021
3502 GA UTRECHT
Netherlands

JISSI BUSINESS REPLY FORM

I/We want to subscribe JISSI for one/three years.

Name : ....

Postal Address : ....

....

....

Postal Code : ....

Country : ....

Telephone : ....

Fax : ....

E Mail : ....

Whether member of ISSI* : Yes / No

I/We enclose the sum of ....

by Demand Draft / Cheque / Money order*

Reference and date of DD/Cheque/MO : ....

Signature

*Please cross out which do not apply

To

BRZARK INFORMATION SYSTEMS (P) LTD,
112, HUMAYUN PUR (NEAR N.C.C. OFFICE)
SAFDARJUNG ENCLAVE. NEW DELHI - 110029.

JISSI : Guidelines to Contributors

JISSI: The international journal of Scientometrics and Informetrics

is a quarterly publication. It provides a forum for dissemination and exchange of ideas and scholarship in the field of Scientometrics and Informetrics.

JISSI is a refereed international journal devoted to the quantitative aspects including mathematical, statistical and computational analyses and modelling of :

- Science of science and technology including technometrics
- Studies on social sciences, arts and humanities
- All informational processes e.g., generation, growth, diffusion, dynamics, use etc. of information.
- Management of information systems and services including libraries, archives, databases, information centres and any other type of document collections

JISSI intends to publish

- research papers of highest standards
- state of the art reviews (historical reviews will not be preferred as such except when filling a gap.)
- short communications
- letters to the editor
- book reviews
- the (irregular) features of reprint of
- classical and very important but apparently obscure papers.
- profiles of important personalities (these should be serious essays discussing the work and not just life sketches)
- glossary of scientometrics and informetrics.
- editorials solicited and unsolicited(guest editorials may be published dealing with emergent issues).

- Manuscripts written in English should be forwarded in triplicate to the honorary executive editor or to any other member of the editorial board. It is preferable that a manuscript be forwarded through an editor who may be interested in the topic of the paper. It would be better if a paper be pre-reviewed and pre-edited before it reaches the Hony Exec Editor.
- The manuscript should be typed or printed, double spaced on A4 size paper with wide (4 cm.) upper, lower and left-hand margins. Manuscripts on diskettes may be considered if composed in Wordstar, Word perfect or Word. A legible hard-copy should accompany the machine readable copy.
- Every manuscript (except for glossary) must have a suitable abstract, a short list of keywords, full address of communication (alongwith E-mail and/orFax numbers whenever possible), address(es) of author(s), necessary and appropriate references.
- The formating of the articles should follow this sequence: title of the paper, name(s) of the author(s), full address (es), abstract, list of keywords, text of the article, acknowledgement, appendice(s), references.
- Tables and diagrams should be on separate sheets and not within the text. But the prefered positions for such matters should be shown in the text.
- Figures, diagrams and other illustrative matters may be sent as camera-ready copies keeping in view the printing area of JISSI. Lettering in illustrative matters should allow for necessary reductions.
- Legends for illustrative matters should be on a separate sheet the legend for table should be above the table and the legend for an illustrative matter should be below it.
- Mathematical symbols and formulas need to be written clearly and legibly. If a preferred symbol cannot be typed it should be written in hand putting the name in words on the margin.

References should be in the following style:

- They should be numbered even if arranged alphabetically. The numbers within the text should be written in box brackets.
- Each reference to journal articles should have the serial number, name(s) of author(s) (surname first) in upper lower, year of pub lication in parenthesis, title of the article, name of the journal (preferably unabbreviated),volume number, issue number in parenthesis, pagination.
- For books - serial no.,names of authors, year of publication in parenthesis, name of the book, publisher, place of publication, pagination.
- For an analytical entry of a part of a book or for an article in an anthology the sequence should be- serial no., name (s) of the author(s), the year of publication in parenthesis, title of the article or the chapter heading,description of the source material starting with the word in and should include the title,author(s)' or editor(s)'name(s), publishers, place of publication within parenthesis, pagination. If the author(s) of the referenced item and the source document be same, repetition may be avoided. In case of proceedings of the meetings, the full description of the meeting should be provided.

Examples :

Journal articles : Rousseau, R. (1994) Similarities between informetrics and econometrics. Scientometrics 30(2-3), 385 - 387.

Books : Garfield, E (1979) Citation indexing : its theory and application in science, technology and humanities. Wiley : New York.

Part of a book : Hertzel, D.H. (1987). Bibliometrics. History of the development of ideas (in Encyclopedia of library and information science. Volume 42,Supplement 7, edited by A. Kent. Marcel Dekker : New York - Basel) 144 - 219.

Style

- Author(s) should use simple sentences and avoid complex or multiple sentences. All contributions should be brief and concise.
- 'I' and 'We' (except in the abstract) rather than the (present) author(s) and the active voice rather than the passive voice should be used.
- Abbreviations are generally to be avoided. When names or phrases for which abbreviations exist occur frequently, the full form along with the abbreviation should be given and the abbreviation can be used subsequently.
- The title should be appropriate, concise and informative.
- Subtitle should be avoided unless necessary.
- Italics (i.e., underline in the manuscript) for common Latin expressions or abbreviations, such as, a priori, et al, ca, cf, e.g., viz. etc should not be used. Italics may be used for special emphasis, for new terms when they are introduced, words in other languages and titles of books and journals. Roman numerals should be avoided. Headings should not be written in capital letters and they are not to be underlined.
- The number of tables should be kept in minimum. Tables should be small and simple.
- Vertical lines in tables are to be avoided.
- Punctuation marks should be avoided as far as possible.

Reprints

- Authors will receive 30 reprints of a published article.

** Camera ready copy of an accepted manuscript will be welcome.Such a copy should follow the printing style of JISSI.For this the Hony Exec Editor may be consulted.

** JISSI has a panel of language editors. Otherwise acceptable articles will be language edited with consent from the author.

** News items can be published in JISSI only as part of ISSI Newsletters.

JISSI : The International Journal of Scientometrics and Informetrics

Volume 1 Number 2 June 1995

CONTENTS

Published by Prabhat Pandey, Brzark Information Systems (P) Ltd. 112, Humayun Pur (Near NCC Office), Safdarjung Enclave, New Delhi -110 029 and printed by him at International Printers, 2, Humayunpur, Safdarjung Enclave, New Delhi -110 029