Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 60, No. 3 (2004) 421–432

Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies

OLLE PERSSON,[a] WOLFGANG GLÄNZEL,[b,c] RICKARD DANELL[a]

[a] *Inforsk, Department of Sociology, Umeå University, Umeå (Sweden)*
[b] *Katholieke Universiteit Leuven, Steunpunt O&O Statistieken, Leuven (Belgium)*
[c] *Hungarian Academy of Sciences, Institute for Research Organisation, Budapest (Hungary)*

Several research studies and reports on national and European science and technology indicators have recently presented figures reflecting intensifying scientific collaboration and increasing citation impact in practically all science areas and at all levels of aggregation. The main objective of this paper is twofold, namely first to analyse if the number or weight of actors in scientific communication has increased, if patterns of documented scientific communication and collaboration have changed in the last two decades and if these tendencies have inflationary features. The second question is concerned with the role of scientific collaboration in this context. In particular, the question will be answered to what extent co-authorship and publication activity, on one hand, and co-authorship and citation impact, on the other hand, do interact.

The answers found to these questions have strong implication for the application of bibliometric indicators in research evaluation, moreover, the construction of indicators applied to trend analyses and studies based on medium-term or long-term observations have to be reconsidered to guarantee the validity of conclusions drawn from bibliometric results.

Introduction

Recent studies have shown that both the number of co-authored scientific publication and citation impact has steadily increased in all subject fields during the last two decades. An intensifying collaboration has been reported for all fields and at practically all level of aggregations, for instance, by GLÄNZEL & DE LANGE (1997) and GLÄNZEL (2001) for the macro level, by GÓMEZ et al. (1995) for the meso level, and by DING et al. (1999) and GLÄNZEL (2002) for the micro level. Furthermore, growing citation rates per paper could be observed for all fields of science in the last two decades (see, *REIST*-2, 1997 and GLÄNZEL, 2001) and journals are celebrating their rising *impact factors*. The number of publications (articles, letters, notes and reviews) has increased between 1980 and 1998 by roughly *one third* whereas the number of citations received by them has grown by *three quarters* in the same period.

Received January 30, 2004

Address for correspondence:
OLLE PERSSON
Inforsk, Department of Sociology, Umeå University
SE-901 87 Umeå, Sweden
E-mail: olle.persson@soc.umu.se

0138–9130/2004/US $ 20.00

Since more citations received is the same as more citations given, then the disproportionate growth of citations must be related to the reference behaviour of papers and their authors. The most obvious explanation is that the reference lists have become longer. However, if there is an increased preference for source items the effect could be the same without the reference lists becoming longer. To what extent these two explanations hold is not known. It could either be one of them, or a combined effect.

Then, another set of explanations are needed to understand changes in reference behaviour. For example, we would expect reference lists to become longer simply because there are more publications to cite as we move forward in time. However, that would imply a proportionate growth of citations to publications, and it doesn't explain why citations grow faster than publications. An additional explanation might be that authors cite more because it has become easier to search and access relevant literature, and eventually also to the advantage of the literature covered by the citation databases. Otherwise, if growth of citations and references or that of publication activity and authors is not proportional, then a changing structure of the underlying communication networks must be assumed. For instance, if productivity (papers per author) is growing faster than the number of publications then an intensifying scientific collaboration and an increasing density of co-publication networks is the only possible explanation.

In this study we will focus on collaboration. It is reasonable to assume that adding authors to papers also means adding references to papers, since authors can be assumed to have different exposure and access to literature. We can also study if the reference behaviour is dependent upon the internationality of the papers, assuming that the collaboration among authors from different countries contributes even more to the number of papers cited.

The change of reference behaviour is not only a phenomena that needs explanation. We should also look at its consequences in terms of bibliometric indicators. The already observed growth of authors per paper is an inflationary trend that have implications for the interpretation and measurement of productivity and impact.

Furthermore, at the level of countries an increased level of collaboration could lead to diminishing differences among the countries in citation impact – and thus in a sort of convergence. Such a tendency could be even stronger since internationally co-authored papers are generally more cited than domestic. The answers to these questions are expected to have strong implication for the application of bibliometric indicators in research evaluation. Moreover, the construction of bibliometric indicators applied to trend analyses and studies based on medium-term or long-term observations might have to be reconsidered.

Data sources and data processing

The study is based on the analysis of all papers indexed in the 1980-2000 annual volumes of the *Science Citation Index* (SCI) of the Institute for Scientific Information (ISI – Thomson Scientific, Philadelphia, PA, USA). Citations have been determined for three-year citation windows on the basis of an item-by-item procedure using special identification-keys. For the citation analysis, only papers of the type articles, letters, notes and reviews have been taken into account. In order to avoid biases by a possibly growing number of letters or reviews, the analysis of reference literature has been restricted to articles and notes. Citations to these publications have been determined on an item-by-item basis for three-year citation windows beginning with the publication year.

Methods and results

In order to approach the study of inflationary bibliometric values from the viewpoint of scientific collaboration systematically, we will address and answer several research questions in the context of possible measurable influence of collaboration on bibliometric indicators. First, we have to analyse if such disproportional trends expressing inflationary tendencies as described above can really be observed.

What are the measurable symptoms of the inflationary effects?

Figure 1 illustrates the disproportionate growth of references, citations and authors in relation to publications for all fields combined. The year 1980 presents the level 100%. The citing-cited time window is the same for references and citations, e.g., for the year 1980, references are the sum of references to 1980 from publications in 1980, 1981 and 1982. Since citations are counted from publications indexed in the SCI database whereas references cover all cited literature, the number of citations is a fraction of those references going to the source items of 1980.

While papers have grown by 36 percent between 1980 and 1998, authors have grown by 64 percent, citations received by 76 percent and references made by 93 percent. These results are in line with observations based on an earlier study as reported in the introduction, namely, that the number of publications increased between 1980 and 1998 by roughly *one third* whereas the number of citations received by them has grown by *three quarters* in the same period. Expressing the trends presented in Figure 1 in verbal terms, we can say that production per author is decreasing while citation rate per paper is increasing. There is also an increase in citations per authors, although not linear, which means a higher pay-off for less effort. Since the growth of references is stronger than the growth of citations it means that there is not a tendency towards an increased preference for citing source items. Thus, the growth of reference lists appears to be the only direct cause of growing citations.

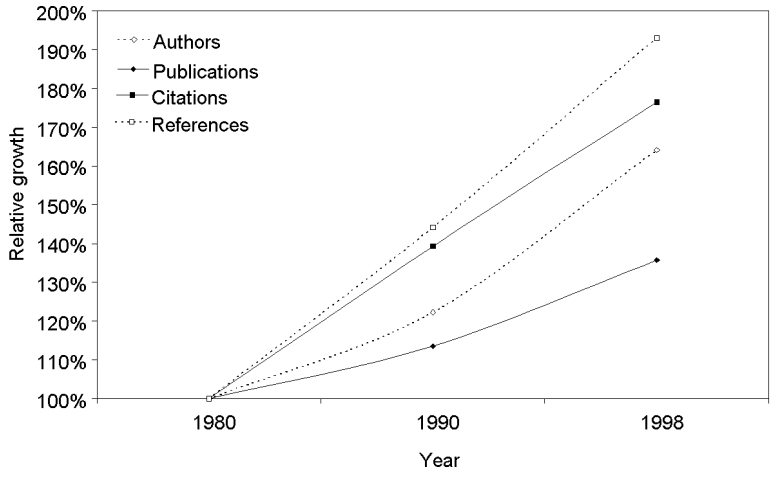

Figure 1. Relative growth of publications, authors, references and citations (all fields combined)

Collaboration appears to be a potential explanation to the disproportionate growth of authors, references and citations. Of course, the fact that the number of authors grows faster than the number of papers is due to increased collaboration. However, the question arises of if collaboration also increases the length of reference lists.

Collaboration and reference behaviour

Figure 2 shows a positive relationship between the number of co-authors and the length of reference lists for all fields combined. For papers published in 1980, we can see that adding an author makes the reference lists grow by half a paper. This also holds for year 2000 even if the effect of adding authors is somewhat stronger. For both periods, the relationship is almost perfectly linear, at least when we set the upper level of co-authors to twelve. From the mathematical-statistical viewpoint this is justified since papers with at most 12 co-authors represent more than 99% of all papers published in the two years, 1980 and 2000. The fluctuations beyond this level are to a large extent conditioned by the small sample size. On the other hand, the extremely high number of co-authors in several fields, for instance, in high-energy physics, where the number of co-authors often exceeds 200, clearly obeys other rules than "standard collaboration" in science.

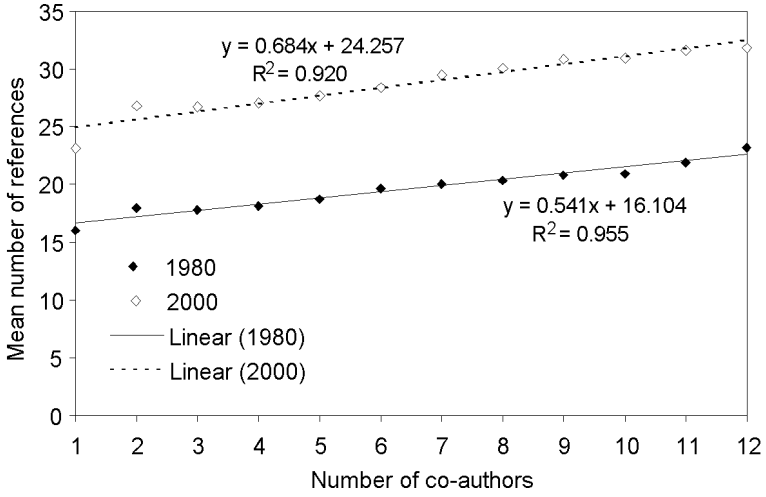

Figure 2. Relationship between number of co-authors and length of reference lists (all fields combined)

If we look at the intercepts, the difference between the periods is about eight references, which means that the increase in collaboration is only a partial explanation to the growth of references.

Another question concerning the effects of collaboration is if international collaboration adds more references compared to domestic papers, since authors from different countries can be assumed to have different information exposure and habits. Figure 3 decomposes the trends in Figure 2 into national and international papers. Internationally co-authored papers have approximately two references more than domestic papers. The parallel regression lines implies that each new author adds about half a reference for domestic as well as international papers and for both periods.

So far, we have found that more authors means more references, too, especially if we combine authors from more than one country. The next question is: What kind of citations do they add? We can expect that authors have a preference for their own papers. This is supported by findings of VAN RAAN (1998) who shows that the share of self-citations is higher in co-authored papers and especially for international papers. However, the question here is if the self-citation rate of collaborative papers is increasing over time and thus a cause of the inflation in citations. An analysis of self-citations 1991-1999 basing on 3-year citation windows, shows that the share of self-citation even decreases slightly: 1991: 29,1%, 1995: 28.3% and 1999: 26.7%. The total number of self-citations, however, increases to a lesser extent than non-self citations. For the publication year 1992, have found the following approximation on the basis of a linear regression model for the number of co-authors (≤ 12, representing 99.5% of all

papers) and the expected (self-)citation rate under the condition that the paper had the given number of co-authors. $y = 0.19 \cdot x + 0.33$ for self-citations and $y = 0.58 \cdot x + 0.36$ for foreign citations. The correlation is sufficiently high $r \sim 0.99$. To conclude, the self-citing tendency of collaboration does not explain the growth of citations per paper. Moreover, collaboration has a stronger effect on foreign citations, i.e., on non-self citations, than on self-citations, which, in other words, means that self-citing is more than compensated for by the citing through foreign papers.

Collaboration and performance measurement

In this section, we analyse performance measurement in the context of increasing scientific collaboration. First, it is possible interaction with productivity is studied; the second analysis is concerned with its influence on citation impact.

Collaboration and publication activity. Increased collaboration means growing fractional contributions to the papers by the authors. If this is not considered when comparing authors we will get an illusory productivity growth. Figure 4 presents the percentage share of author producing 1, 2, … papers a year for all fields combined. The "average productivity" has roughly increased by one half paper, from about 2.5 papers in 1980 to 3 papers in 2000. When we look at the distribution of papers over authors, we find that the mean productivity has increased, which is also reflected in the fact that the share of low-productivity authors has decreased, that of medium and high productivity has increased.

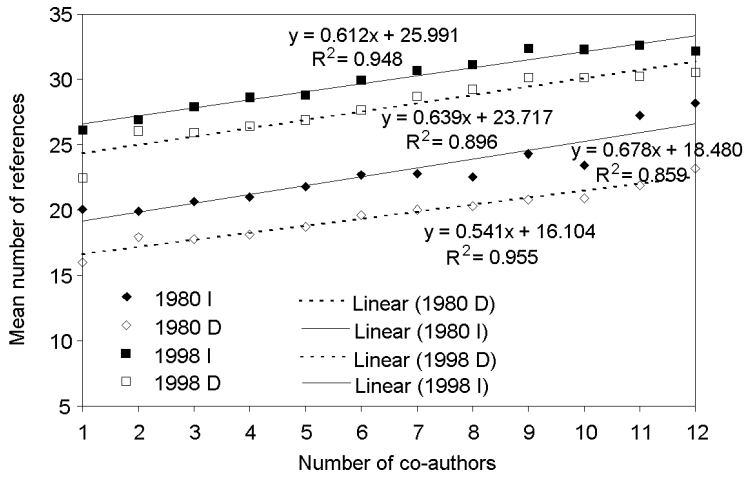

Figure 3. Relationship between number of co-authors and length of reference lists controlling for international collaboration (all fields combined)

Figure 5 shows that the shift towards multiple co-authorship is even more pronounced than shift to medium and high productivity. This supports the conclusion that the growth of papers per author is really an effect of collaboration rather than of becoming more productive. However, it is conceivable that the citation impact resulting from collaboration is superior. This will be discussed in a later section.

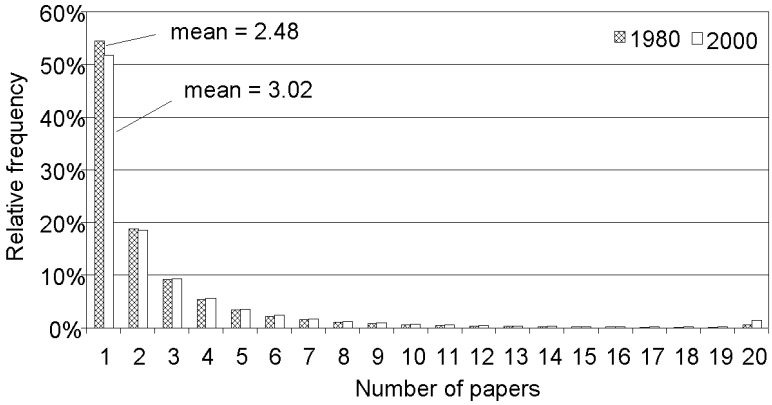

Figure 4. Change of publication activity of individual authors in time (all fields combined)

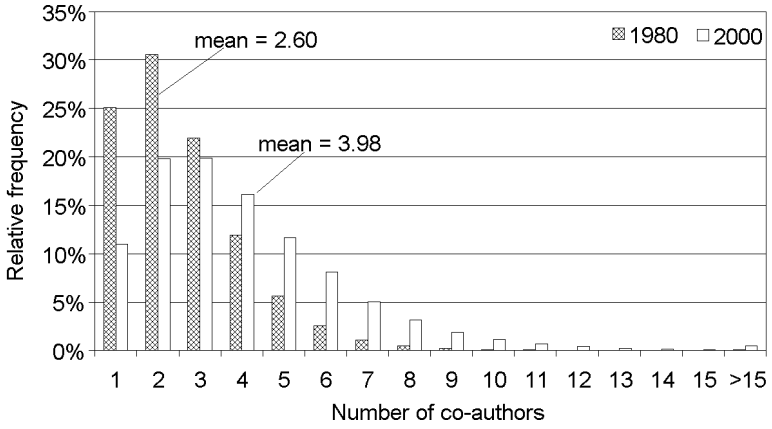

Figure 5. Change of co-publication patterns of individual papers in time (all fields combined)

The obvious way to get around the productivity paradox is to use fractionalised counting when distributing papers over authors. Then the sum of fractions would be the same as the number of papers, and we would get a much better indicator of productivity differences. Since the level of multi-authorship has increased, we can be sure that the fractional counting will yield a lower mean value for year 2000 in Figure 4.

The above-mentioned productivity paradox becomes, however, even more complicated in the light of the suggested fractional counting scheme if we refer to the results by BRAUN et al. (2001) and GLÄNZEL (2002) who have shown that in the fields of neuroscience, biomedical research, chemistry and mathematics, increasing number of co-authors is positively correlated merely to the extent of a field-specific value amounting to about 5-6 co-authors in the life sciences, 3-4 in chemistry and 2 in mathematics. Beyond this level, the correlation between number of co-authors and their productivity turns negative. Although there is no immediate and apparent explanation for this phenomenon, the threshold at which collaboration reaches its maximum effect on productivity roughly correlates with average collaboration in the field. However, the fact that newcomers, transients, and terminators, i.e., scientists who are entering and/or leaving the community of active authors, are less affected as has shown by BRAUN et al. (2001) might be a symptom of "occasional" co-authorship of persons involved in higher-than-average collaboration.

Collaboration and citation impact. Figure 6 presents the change of the distribution of citation over papers from 1980 to 1998 on the basis of a 3-year citation window. The share of uncited papers has considerably decreased (by more than 6 percent points); both, the share of medium and more frequently cited papers have increased. The "average citation rate" has increased by almost one citation. This trend does not change if we exclude non-cited papers; also the (conditional) mean citation rate of *cited papers* increased by almost one citation, particularly, from 4.77 in 1980 to 5.67 in 1998.

From Figure 1 we can conclude that the ratio citations/author also will increase since citations grow faster than authors. This is also reflected when we distribute citations over authors. In a recent study on changing collaboration patterns in the fields biomedical research, chemistry and mathematics, GLÄNZEL (2002) has shown that the number of co-authors has a strictly positive effect on mean citation rate except for mathematics if the number of co-authors is greater than 7.

Another question concerning the effects of collaboration is whether international collaboration attracts more citations compared to domestic papers. Figure 7 decomposes the relationship between number of co-authors and mean citation rate into national and international papers for all fields combined. While in 1980, internationally co-authored papers attract approximately one citation more than domestic papers and the mean citation rates of both types of co-publications show an almost perfect parallel increase,

the situation completely changed in 1998. Domestic and international papers start from almost the same level if a paper has one, two or three co-authors, but international collaboration causes a more powerful increase if more than three authors are involved.

In this section, we have studied the interaction between scientific collaboration as reflected by co-publication patterns with performance measurement. This is not merely a matter of the underlying "sample size". Figure 7 (similarly to Figure 3 in a previous section) combines the micro with the macro approach; on one hand, collaboration is measured at the micro level since collaboration patterns have been determined for individual authors, on the other hand, citation impact is also compared at the macro level, namely for domestic and international collaboration. This makes the interpretation somewhat more complicated but at the same time more challenging, too. A paper with two co-authors or even a single-authored paper can be an international one, whereas a paper co-authored by, say, eight or ten authors can be results of domestic or even an intra-institutional collaboration. Unlike the correlation between number of co-authors and length of reference lists, internationality has beside the interaction found at the micro level apparently a large influence on citation behaviour.

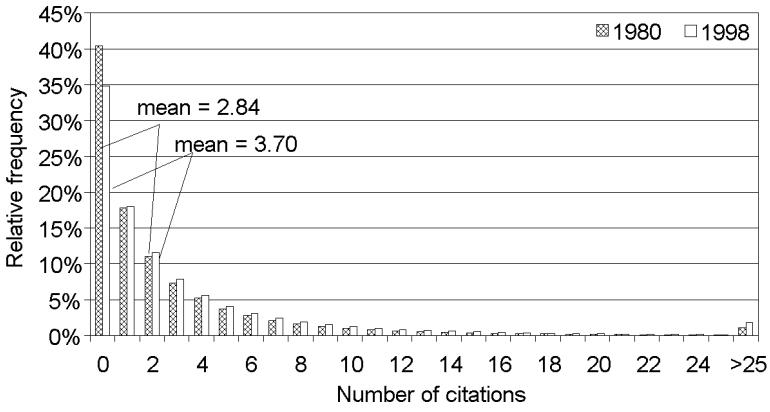

Figure 6. Change of observed citation impact of individual papers (all fields combined)

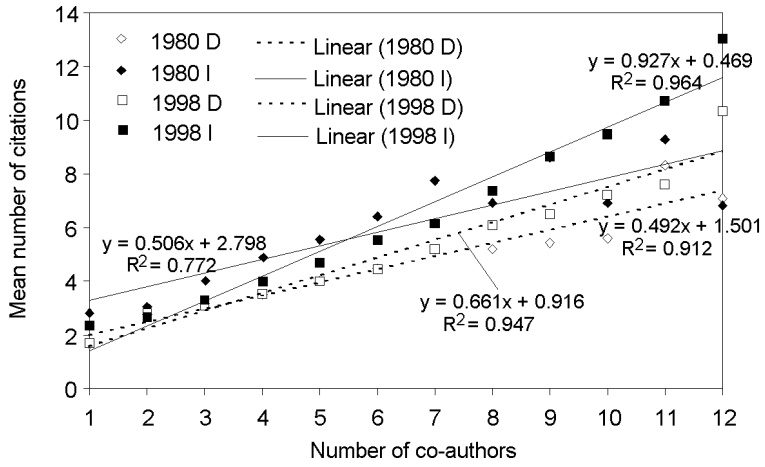

Figure 7. Relationship between number of co-authors and mean citation rate controlling for international collaboration (D – domestic papers, I – international papers, all fields combined)

International collaboration and convergence

The observed increasing international collaboration in almost every field of science is a well known fact, which one would expect leads to less differences in citation rate among the collaborating countries. Table 1 displays the mean observed citation rate (MOCR) for 22 counties ranked by impact in 1980. If we calculate the standard deviation for the first eleven countries it decreases from 0.74 for papers published in 1980 to 0.60 citations for the 1998 papers. The convergence among the most cited countries is apparent. On the other hand, among the eleven less cited nations the variation grows from 0.66 citations to 1.25. This indicates that there is stratification in international collaboration to the advantage of the more cited nations.

If we have a closer look at individual countries Sweden shows a negative trend compared to the other top, mostly European countries, while Switzerland increases strongly. It would be a premature conclusion to say that Sweden is loosing impact because of collaboration and Switzerland not. So there are, of course, other factors besides collaboration that effect citation trends for individual nations.

Table 1. Mean observed citation rate (MOCR) for 22 countries in 1980 and 1998 for international papers in all fields combined

Rank in 1980	Country	1980	1998
		MOCR	MOCR
1	Switzerland	6.77	7.15
2	Sweden	6.68	6.16
3	Denmark	6.22	5.88
4	Netherlands	5.67	6.52
5	Finland	5.63	6.53
6	Germany	5.34	5.73
7	USA	5.24	6.70
8	UK	5.08	6.31
9	Italy	5.01	5.91
10	France	4.76	5.52
11	Spain	4.57	5.01
12	Belgium	4.57	5.39
13	Australia	4.39	5.34
14	Canada	4.36	6.53
15	Norway	4.35	5.21
16	Japan	4.23	5.82
17	South Korea	3.82	3.56
18	Mexico	3.81	3.06
19	Greece	3.78	3.59
20	Austria	3.70	5.24
21	Poland	2.85	3.54
22	Turkey	2.47	2.88

Conclusions

The analysis of trends in co-authorship links, publication activity, number of references and citation frequency at the level of individual papers results in the assumption that the density of co-publication and citation networks has increased.

Implementations of the results of the study for research evaluation are two-fold. First, the most important consequence of the sometimes almost inflationary growth of the value of the basic indicators is the need for relative and strictly normalised indicators in bibliometric trend analyses and medium-term or long-term studies.

The mere calculation of averages is consequently no sufficient to compensate the effect as these indicators might show a growth where the development is actually characterised by stagnation or even by decline.

Although basic indicators like absolute publication and citations counts, there shares and averages are widely accepted as useful tools in measuring research performance, their uncritical use can result in incorrect interpretations developments. A proper normalisation of standard measures and the use of relative indicators are, therefore, indispensable for trends analyses or medium- or long-term studies to guarantee the validity of conclusions drawn from bibliometric results.

General tendencies like the observed intensifying (international) collaboration and its interaction with increasing publication activity and citation impact, however, must not be interpreted as universal rules that apply to all actors. Breakdown by subjects and decomposition of bibliometric indicators help to identify those actors whose patterns deviate whether in positive or negative direction from the general national trend.

References

BRAUN, T., GLÄNZEL, W., SCHUBERT (2001), Publication and cooperation patterns of the authors of neuroscience journals, *Scientometrics,* 51 (3) : 499–510.

DING, Y., FOO, S., CHOWDHURY, G. (1999), A bibliometric analysis of collaboration in the field of information retrieval. *The International Information & Library Review,* 30 : 367–376.

GLÄNZEL, W., DE LANGE, C. (1997), Modelling and measuring multilateral co-authorship in international scientific collaboration. Part II. A comparative study on the extent and change of international scientific collaboration links, *Scientometrics*, 40 (3) : 605–626.

GLÄNZEL, W. (2001), National characteristics in international scientific co-authorship, *Scientometrics*, 51 (1) : 69–115.

GLÄNZEL, W. (2002), Co-authorship patterns and trends in the sciences (1980–1998). A bibliometric study with implications for database indexing and search strategies, *Library Trends*, 50 (3) : 461–473.

GÓMEZ, I., FERNÁNDEZ, M. T., MÉNDEZ, A. (1995), Collaboration patterns of Spanish scientific publications in different research areas and disciplines, In: *Proceedings of the Biennial Conference of the International Society for Scientometrics and Informetrics* (Ed. by M. E. D. KOENIG, A. BOOKSTEIN), Learned Inf., Medford, NJ, pp. 187–196.

REIST-2. (1997), *The European Report on Science and Technology Indicators 1997*. EUR 17639. European Commission, Brussels.

VAN RAAN, A. F. J. (1998), The influence of international collaboration on the impact of research results – Some simple mathematical considerations concerning the role of self-citations. *Scientometrics*, 42 (3) : 423–428.