
Coauthorship Patterns and Trends in the Sciences (1980–1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies

WOLFGANG GLÄNZEL

ABSTRACT

THE PRESENT STUDY AIMS AT describing both the common and the distinguishing features of coauthorship trends and patterns in selected science fields. The relation between coauthorship schemes and other bibliometric features, such as publication activity and citation impact are analyzed. I show that, while copublication activity has grown considerably, the extent of coauthorship and its relation with productivity and citation impact largely varies among fields. Besides universally valid tendencies, subject specific features can be found.

INTRODUCTION

Authorship is a primary bibliometric descriptor of a scientific publication. Its trends and patterns characterize the social and even the cognitive structure of research fields. The most characteristic tendency of recent times is intensifying scientific collaboration. Collaboration in research is reflected by the corresponding coauthorship of published results, and can thus be analyzed with the help of bibliometric methods.

Kretschmer has conducted profound analyses of coauthorship patterns as a function of the authors' productivity (e.g., Kretschmer, 1994). She concluded that, in invisible colleges, coauthorship between scientists with the same number of publications is more frequent than between authors of different publication activity and that the opposite is valid in institutionalized communities. On the other hand, the reverse question, whether higher "cooperativity" of authors exhibits a greater publication activity, has lit-

tle been dealt with so far. The relation between collaboration and productivity was first studied by Beaver & Rosen (1979). The authors analyzed scientific papers of the French elite in the early eighteenth century, and concluded that collaboration is associated with higher productivity. In a recent paper, Braun, Glänzel, & Schubert (2001) have analyzed the relation between cooperativity and productivity in different author categories in the field of neurosciences. In the following study, I extend some of these results to broader science fields.

Bibliometric meso and macro studies concerned with the analysis of copublication patterns at the institutional (e.g., Hicks, Ishizuka, Keen, & Sweet, 1994; Hicks & Katz, 1997), and the national level (Gómez, Fernández, & Méndez, 1995; *REIST-2*, 1997; Glänzel, 2001) have shown a growing copublication activity. This applies to both scientific collaboration between industry and universities and research cooperation at the domestic, national, and supra-national level. These studies have also proved that international collaboration is—at least on the average—associated with a higher citation impact.

Besides economic and political factors, intra-scientific factors (e.g., Luukkonen, Persson, & Silvertsen, 1992), especially changing communication patterns and increasing mobility of scientists, are influencing collaboration. These factors motivate cooperation in “less expensive” areas, such as pure mathematics, and theoretical research in social sciences, too. The growing share of copublications in theoretical fields could be substantiated in the named literature.

The question arises whether one can observe the same tendencies also at the lowest level of aggregation, that is, at the level of individual publications and of authors. In the light of the above considerations, the following three questions will be answered:

- Does the development of coauthorship at the micro level, that is, at the level of individual papers, follow the trend of intensifying collaboration found at the meso (institutional) and macro (i.e., national and supra-national) level, particularly in the context of international research collaboration?
- Has the cooperativity any influence upon the authors' productivity?
- Do multiauthored papers exhibit a greater citation impact than publications with single authors?

These issues have to be addressed and answered at each level of aggregation separately since the results by Gómez, Fernández, & Méndez (1995) and Katz (2000) have shown that different types of collaboration may exhibit contradictory effects. For instance, while some types of collaboration exhibit Matthew effect, others exhibit the inverse effect (see Katz, 2000). Therefore, conclusions made for a higher level of aggregation cannot be simply assigned to a lower one and vice versa. Consequently, the results of

the following analysis should not be generalized as being valid for all types of scientific collaboration.

DATA SOURCES

All papers recorded in the annual volumes of the Science Citation Index (SCI) of the Institute for Scientific Information (ISI) as *article*, *letter*, *note*, or *review* were taken into consideration. For instance, documents of the type *corrections*, *editorial material*, *bibliographical items*, *meeting abstracts*, *book reviews*, *news items*, etc. have been omitted. The latter types are from the bibliometric viewpoint not considered conveyers of relevant scientific information related to original research results, and are thus not regarded as citable items. All (co)authors indicated in the corresponding search field have been taken into account. Author names were taken as recorded into the database, no corrections have been made for spelling variants or for adjustment of homonyms.

Subject classification of publications was based on the field assignment of journals (in which the publications in question appeared) according to the major fields of science representing the life sciences, the natural sciences, and mathematics. In particular, the fields of Biomedical Research (BRE), Chemistry (CHE), and Mathematics (MAT) have been selected. The definition of these subject areas is in keeping with the subject scheme used in the 2nd edition of the *European Report on Science and Technology Indicators (REIST-2, 1997)*. The field Biomedical Research includes the following subfields: (1) Pharmacology and Pharmacy, (2) Pathology, (3) Research Medicine, and (4) Immunology. The subject area Chemistry comprises: (1) Inorganic Chemistry and Engineering, (2) Analytical Chemistry, (3) Physical Chemistry, and (4) Organic Chemistry. The field of Mathematics is not subdivided into any particular subfield.

The study is based on papers published in the years 1980, 1986, 1992, 1996, and 1998. Citation counts have been determined in a three-year period on the basis of an item-by-item procedure using special identification keys. In particular, citations were counted in the year of publication and the two subsequent years, that is, in the period 1996–1998 for papers published in 1996. The applicability of the three-year citation window scheme has been proved in several recent methodological studies (e.g., Glänzel & Schoepflin, 1995 and *REIST-2, 1997*).

METHODS AND RESULTS

Theoretical Implications

In a current study by Glänzel & de Lange (2002), the distributions of the number of partner countries over internationally coauthored papers is being analyzed for individual countries in the fields of Biomedical Research, Chemistry, and Mathematics. To date, the analysis has resulted in a

modification of the model assumed in the authors' earlier paper (de Lange & Glänzel, 1997; Glänzel & de Lange, 1997). Originally, a geometric distribution was assumed. This model described extremely skewed distributions with monotonously decreasing probabilities of the number of partners involved. This situation was typical for earlier decades. However, the shapes of the empirical frequency distributions of various countries have changed—they have become less skewed in the 1990s. For some countries, the peak of the distribution is even around the cooperativity value of one or two partner countries. In their study, Glänzel and de Lange have searched for an approximate solution for a suitable distribution within the extended urn model, considering, among others, the geometric, the binomial, the negative binomial, the Poisson, and the Waring distribution. A characterization theorem for discrete probability distributions substantiates that the empirical distributions under study can be found in the "neighbourhood" of the Poisson distribution. One of the basic features of this distribution is that it may take the maximum probability at any value.

From the formally logical point-of-view, increasing international collaboration and increasing multinationality are not automatically tantamount to growing copublication activity of individual authors, since increasing international collaboration might theoretically be caused by a mere replacement of domestic cooperation by international collaboration. However, it is known that coauthorship has increased at all levels of aggregation and, of course, the growth took place at the micro level to a greater extent than at the national/supranational level. Therefore, the application of the above approximate Poisson model seems to be justified to the frequency distribution of coauthors over papers. Consequently, any considerable change of copublication activity of individual authors has to be reflected by the corresponding change of the shape of the empirical cooperativity distribution. In the following sections, the changing shape of the distribution of coauthors over papers will be analyzed, a theoretical explanation for possible observed changes over time, however, will not be given.

Results

In order to answer the first question concerning the trend in coauthorship patterns of individual papers, the distribution of coauthors over publications have been determined for the following four years: 1980, 1986, 1992, and 1998. The mean cooperativity (M), that is, the average number of authors contributing to one paper, is used as an indicator of collaborativity at the micro level. The indicator values for the three selected fields, BRE, CHE, and MAT are presented in Table 1. There is a sharp increase by 48 percent in Biomedical Research. In Chemistry cooperativity increased by 24 percent, and in Mathematics the growth still amounted to 17 percent. This is interesting because cooperativity in the selected lifescience field is traditionally higher than in chemistry or mathematics, where single author-

Table 1. The Development of Coauthors Patterns in Selected Fields (1980–1998) as Reflected by the Mean Cooperativity (M).

Subject Field	1980		1986		1992		1998	
	Papers	M	Papers	M	Papers	M	Papers	M
Biomedical Research	64501	3.47	74630	3.96	86544	4.57	98795	5.13
Chemistry	66576	3.07	69703	3.27	80083	3.50	94600	3.82
Mathematics	14385	2.22	11892	2.30	13362	2.36	18729	2.59

ship was always typical of the field. Field-specific characteristics of coauthorship patterns have therefore deepened.

Since bibliometric distributions are discrete rather than continuous and are often skew, the interpretation of mean values requires the application of additional statistical tools besides the use of mean values. In order to visualise field-specific changes in coauthorship patterns, the frequency distributions of coauthors over papers is presented in Figure 1. The tails of the distributions proved to be long, and have therefore been cumulated.

The share of papers with a low number of coauthors in Biomedical Research shrunk steadily between 1980 and 1998. Thus, the share of papers with one or two authors halved (from 16 percent [27 percent] in 1980 to 7 percent [13 percent] in 1998), and the share of papers with three authors decreased from 24 percent in 1980 to 16 percent in 1998. The share of papers with four coauthors did not change during the eighteen years of observation. The share of papers with five or more authors considerably increased, so that multiauthored papers became predominant and characteristic for the field.

There is a similar, yet not quite as pronounced, trend in Chemistry. While a chemistry paper published in 1980 was most likely to have two coauthors (33 percent), the local maximum moved to three authors with a share of 25 percent in 1998. It is worth mentioning that one quarter of all papers published in 1998 had at least five authors.

The intensifying collaboration and the associated increase of the share of multiauthored papers in Chemistry and in Biomedical Research does not really surprise. The trend towards coauthorship in Mathematics is, however, somewhat striking. In 1980, about two thirds of all papers were single authored and only 6 percent of all journal publications had more than two coauthors. Eighteen years later, in 1998, most papers are still single authored, but the share of papers with one and two authors almost coincides. About 25 percent of all mathematical publications have at least three authors. Although the distribution remains very skew in this field, a considerable increase in individual copublication activity can be observed in the last two decades.

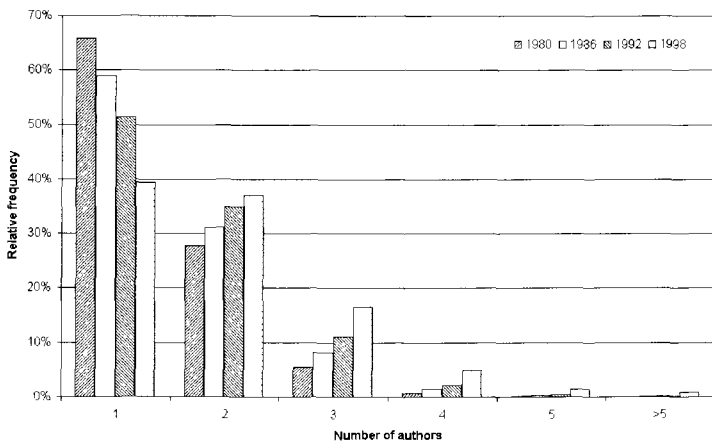
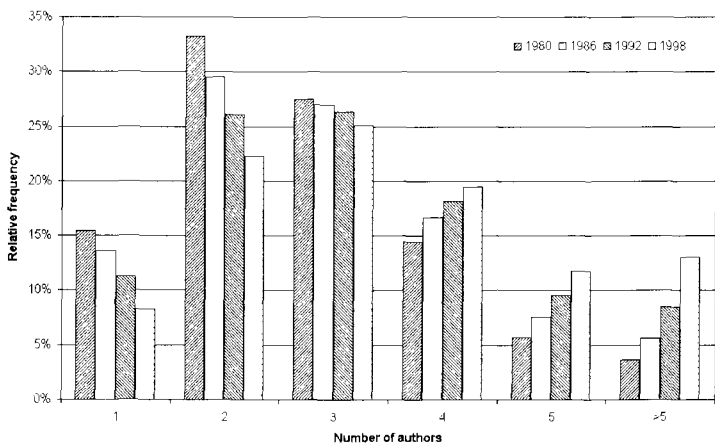
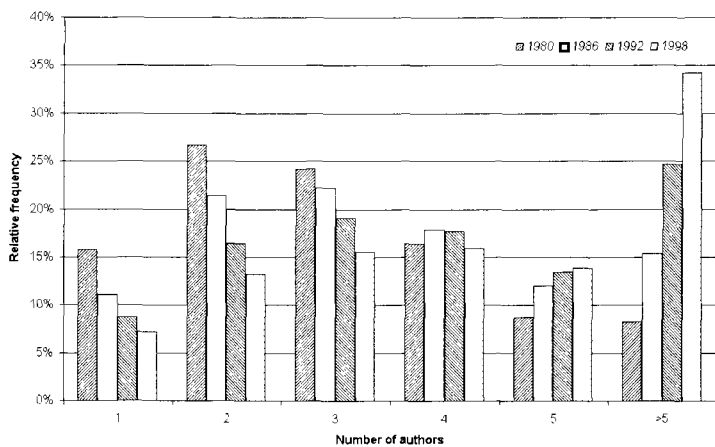


Figure 1. Frequency Distribution of Coauthors Over Papers in Biomedical Research (top), Chemistry (center), and Mathematics (bottom).

After having found an answer to the first question, namely, that copublication activity at the micro level follows the trend of intensifying scientific collaboration observed at the meso and macro level, we can consider the interrelationship between cooperativity and the authors' productivity as formulated in the second question. Figure 2 shows the average publication activity vs. mean cooperativity plot of the authors in Biomedical Research, Chemistry, and Mathematics for papers indexed in the 1996 volume of the SCI. For authors in Biomedical Research there is a peak of productivity around the cooperativity value of six coauthors. In Chemistry, this peak of productivity can be found around the mean cooperativity of three to four. Finally, in Mathematics, mean publication activity takes its maximum value in the case of one to two coauthors. Otherwise, no unambiguous "effect" on publication activity can be found for the number of authors involved. Collaboration is thus not associated with higher productivity at the level of individual authors. In Mathematics, productivity is even slightly decreasing with growing copublication activity. Here, authors who are—on the average—publishing alone or with only one coauthor are the most productive ones. Although "team work" exhibits higher productivity than single authorship in the two other fields, beyond a field-characteristic level, productivity distinctly decreases with growing cooperativity.

The third question addressed in the introduction is concerned with the citation impact attracted by multiauthored papers. To answer this question, all *article*, *letters*, *notes*, and *reviews* indexed in the 1996 volume of the SCI and assigned to the three selected subject areas have been processed. Cita-

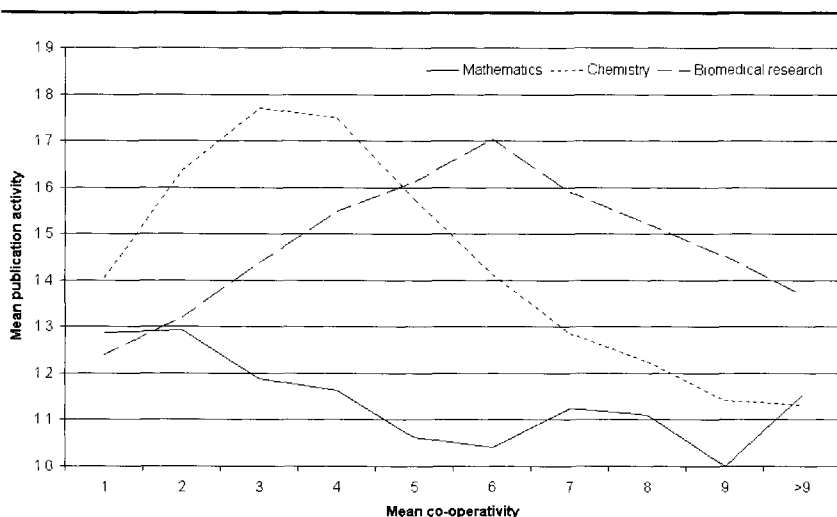


Figure 2. Plot of Average Productivity vs. Average Cooperativity in 1996.

tions have been counted for the period 1996–1998. Unlike in the Journal Citation Reports, journal *impact factors* have here been calculated for one source year (1996) and a three-year citation window (1996–1998). The plot of the average coauthorship of journals vs. journal impact factor for the three fields is presented in Figure 3.

All plots reflect almost uncorrelated patterns. The application of the F -test shows that the two variables can practically be considered independent in all selected fields. The corresponding statistics are presented in Table 2. $F_1 = 1$ for all three samples. It has to be mentioned that there is a slight decline for Chemistry and a certain increase for Mathematics. In case of Biomedical Research, the correlation coefficient is zero. According to the F -test, the two variables are independent at any reasonable confidence level in Biomedical Research. The critical value for degrees of freedom at a confidence level of 99.5 percent is 7.88; that is, the F -values for Chemistry and Mathematics are below this threshold.

In verbal terms, high-impact journals tend to publish chemistry papers with a somewhat lower number of coauthors on the average. The reverse statement applies to mathematics. However, there is no pronounced relation between the journal impact factor and the average cooperativity of papers published in the journal under study, and the hypothesis that the two variables are independent can be accepted at the above confidence level.

Now the question will be answered whether multiauthored papers exhibit a greater citation impact than publications with single authors. First, I will analyze the share of cited papers as a function of the number of coauthors. Both number and share of cited papers with k coauthors are presented in Table 3.

The well-known fact that biomedical research attracts, on the average, higher citation rates than chemistry, and that chemistry literature itself is, on the other hand, more frequently cited than mathematics, is reflected by the share of cited papers. Within each subject area, a clear dependence of the citedness variable on the number of coauthors can be observed. In particular, the share of cited papers grows with the increasing number of coauthors. Roughly speaking, about three quarters of all papers with at least four coauthors each are cited in the three-year period beginning with the year of publication.

Figure 4 presents the mean citation rate of papers as a function of cooperativity. In all three fields, there is a pronounced tendency of growing citation impact if the number of coauthors increases. The drop at the “high-end” of cooperativity in the mathematical sample can be explained in terms of statistical reliability. Only twenty-five papers, that is, 0.15 percent of all mathematical papers under study, have more than eight coauthors each. The decrease might therefore be considered statistically not significant. The field average of citation impact is reached at a cooperativity of fifty-six in Biomedical Research, at thirty-four in Chemistry, and at two in Mathemat-

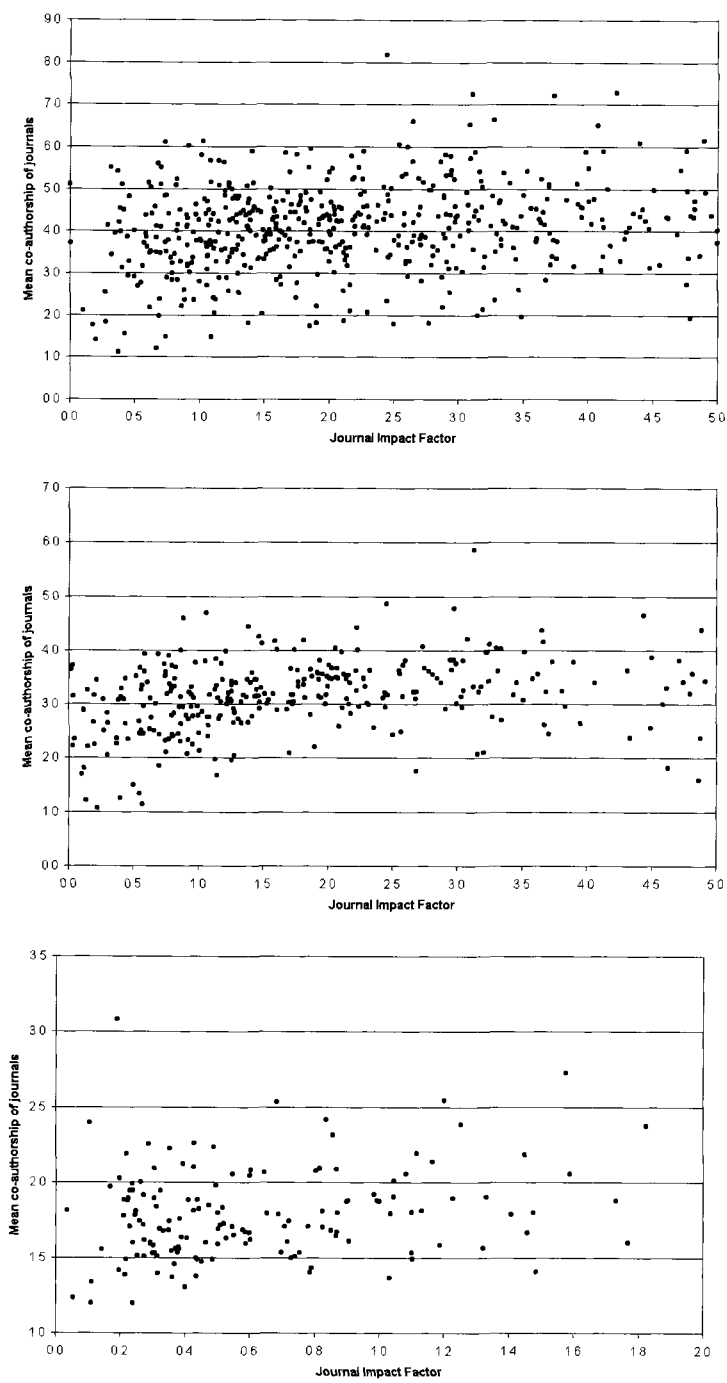


Figure 3. Plot of Average Coauthorship of Journals vs. Journal Impact Factor in 1996. Biomedical Research (top), Chemistry (center), and Mathematics (bottom).

Table 2. Statistics Derived From the Linear Regression Analysis of Average Coauthorship of Journals vs. Journal Impact Factor in 1996.

Statistics	Biomedical Research	Chemistry	Mathematics
r^2	0.000	0.019	0.049
$df(f_2)$	614	348	150
<i>F</i> -statistics	0.01	6.75	7.79

Table 3. Share of Cited Papers as a Function of the Number of Coauthors in 1996.

Number of Coauthors (<i>k</i>)	Number of Papers with <i>k</i> Coauthors			Share of Cited Papers with <i>k</i> Coauthors		
	BRE	CHE	MAT	BRE	CHE	MAT
1	8151	8241	6777	53.1%	47.2%	41.5%
2	12927	20893	6151	68.9%	67.9%	51.3%
3	15201	21884	2406	70.3%	69.9%	56.7%
>3	55928	34066	942	77.5%	73.9%	70.6%

Biomedical Research (BRE), Chemistry (CHE), and Mathematics (MAT).

ics. It is worth mentioning that these thresholds roughly coincide with the local maximum values in the productivity vs. cooperativity plot in Figure 2. There is, however, no causal relation conditioning such coincidence. In all, multiauthored papers exhibit a clearly greater citation impact than publications with single authors in the three selected fields.

In this context, the question of (author) self-citation has to be discussed. The above citation patterns have not been checked for self-citations. Self-citation analysis has been omitted for the following two reasons. As mentioned in the Data Sources section, no corrections have been made for spelling variants of author names or for adjustment of homonyms. This may result in considerable errors in self-citation statistics. Moreover, Figure 2 shows that the mean publication activity does not exceed two papers per year. That is, it can be concluded indirectly that the higher citation rates are not a consequence of possible self-citations alone, and growing citation impact has to be explained mainly with other aspects of scientific communication.

CONCLUSIONS AND IMPLICATIONS FOR DATABASE INDEXING AND SEARCH STRATEGIES

In earlier papers concerned with the analysis of international scientific collaboration, the author has found considerable changes in copublication activity and multinationality of publications during a period of ten years (de Lange & Glänzel, 1997; Glänzel & de Lange, 1997; and Glänzel, 2001). Moreover, I observed an increase of citation impact in papers published in

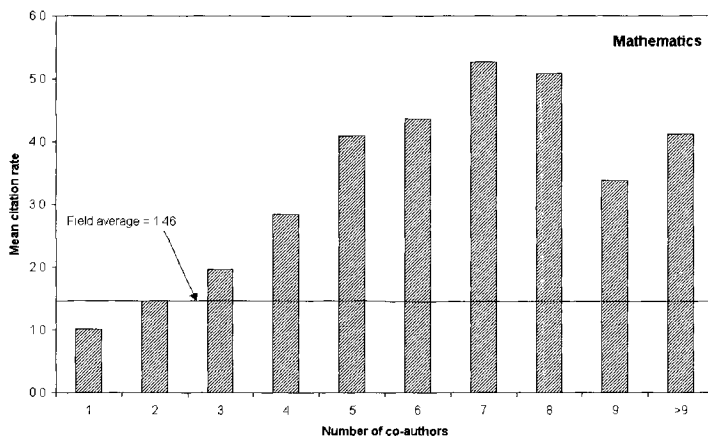
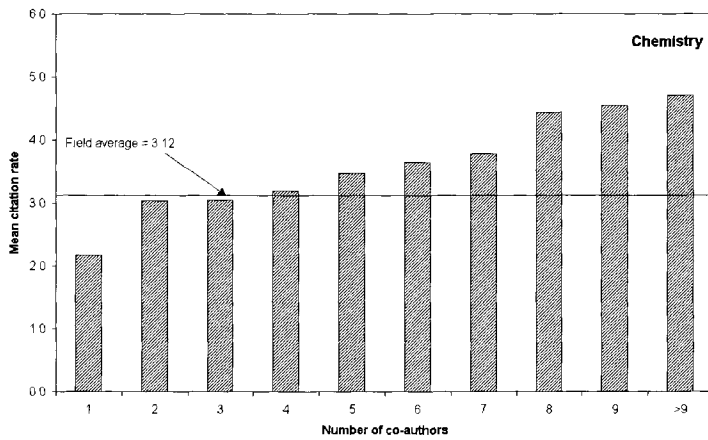
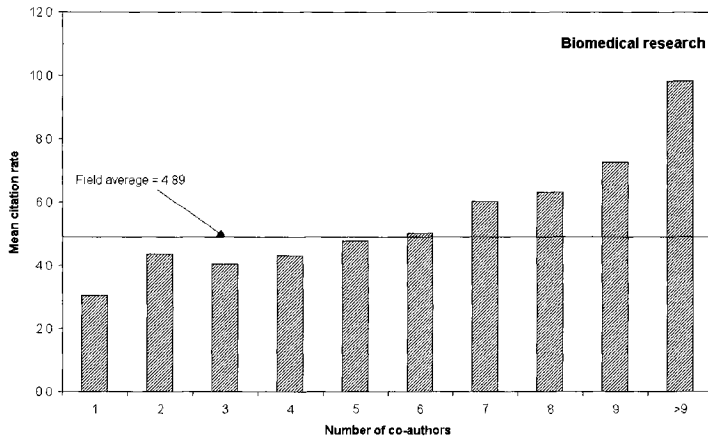


Figure 4. Mean Citation Rate vs. Number of Coauthors in 1996. Biomedical Research (top), Chemistry (center), and Mathematics (bottom).

international cooperation. A similar development could be found at the micro level, although direct parallels must not be drawn because of the different conditions for and different meaning of copublication at the lower level of aggregation.

A theoretical explanation for the considerable change in copublication activity of individual authors is not given. The same applies to the striking trend towards multiauthored publications in biomedical research and chemistry that has been found in the present study. Surprising was the decrease of single-authored papers to a clear minority in mathematics. However, truly multiauthored papers in mathematics, with four authors or more, remain rather the exception than the rule.

The lack of an unambiguous relation between cooperativity and publication activity was somewhat unexpected, although a similar tendency has been shown by Braun, Glänzel, & Schubert (2001) for the field of neurosciences. In particular, a peak of productivity around a field-specific cooperativity value could be found. A question arises as to how much the location of this peak depends on the publication period under study. For longer periods, this local maximum might be taken at somewhat higher cooperativity values; however, these values will remain characteristic for the field.

The theory of a relationship between cooperativity and publication activity was thus not supported by these findings. On the other hand, the theory that multiauthored papers are more likely to be cited, and attract more citations, than single-authored papers was strongly supported and proved to be universal. In particular, the mean citation rate of multiauthored papers in mathematics exceeds the field average by even more than 200 percent. It has, however, to be mentioned that these papers only amount to about 2 percent of all publications in this field. These results are contrasted by the lack of any relation between the impact factor of journals and the mean cooperativity of papers published in them.

From the viewpoint of library and database management, the following implications should be mentioned. Quantitative methods in bibliometrics help to uncover important relations underlying the network of science communication, and to measure their strength. Such relations are established by the thematic linkage that can be measured and described not only with the help of bibliographic coupling and cword and cocitation analysis, but also through the coauthorship or copublication relationship.

In a recent paper, Glänzel & Czerwon (1996) have pointed to classical information retrieval as one possible field of application of bibliographic coupling techniques. In particular, they have shown that these techniques can be used to identify "core documents" representing recent "hot" and other research-front topics. Core documents are thus important nodes in the network of documented science communication. A similar statement holds in the context of scientific collaboration and its citation impact, since citations give a formalized account of the information use and can thus be

taken as a strong indicator of reception. Multiauthored, and first of all internationally coauthored publications, proved to hold key positions within the framework of scientific communication; their citation impact is assumed to exceed standard reception. Apart from the definition of core documents given by Glänzel and Czerwon in the context of bibliographic coupling, other documents, frequently cited and strongly interrelated in terms of theme, can thus serve as core documents in search strategies.

REFERENCES

- Beaver, D. deB., & Rosen, R. (1979). Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship. *Scientometrics*, 1, 133-149.
- Braun, T.; Glänzel, W.; & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, 51, 499-510.
- de Lange, C., & Glänzel, W. (1997). Modelling and measuring multilateral co-authorship in international scientific collaboration. Part I. Development of a new model using a series expansion approach. *Scientometrics*, 40, 593-604.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship. *Scientometrics*, 51, 69-115.
- Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37(2), 195-221.
- 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., & de Lange, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics*, 54 (in press).
- Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature. *Journal of Information Science*, 21(1), 37-53.
- Gómez, I.; Fernández, M. T.; & Méndez, A. (1995). Collaboration patterns of Spanish scientific publications in different research areas and disciplines. In M. E. D. Koenig & A. Bookstein (Eds.), *Proceedings of the Biennial Conference of the International Society for Scientometrics and Informetrics* (pp. 187-196). Medford, NJ: Learned Information.
- Hicks, D.; Ishizuka, T.; Keen, P.; & Sweet, S. (1994). Japanese corporations, scientific research, and globalization. *Research Policy*, 23, 375-384.
- Hicks, D., & Katz, J. S. (1997). *The changing shape of British science*. STEEP special report No. 6, SPRU.
- Katz, J. S. (2000). Scale independent indicators and research assessment. *Science and Public Policy*, 27(1), 23-36.
- Kretschmer, H. (1994). Co-authorship networks of invisible colleges and institutional communities. *Scientometrics*, 30(1), 363-369.
- Luukkonen, T.; Persson, O.; & Silvertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values*, 17(1), 101-126.
- REIST-2. (1997). *The European report on science and technology indicators*. EUR 17639. Brussels: European Commission.