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

Scientometrics,
Vol. 61, No. 3 (2004) 405–426

Visibility of collaboration on the Web

HILDRUN KRETSCHMER,[a,b] ISIDRO F. AGUILLO[c]

[a] *NIWI, The Royal Netherlands Academy of Arts and Sciences, Amsterdam (The Netherlands)*
[b] *COLLNET, Hohen Neuendorf (Germany)*
[c] *CINDOC, Madrid (Spain)*

The emerging influence of new information and communication technologies (ICT) on collaboration in science and technology has to be considered. In particular, the question of the extent to which collaboration in science and in technology is visible on the Web needs examining. Thus the purpose of this study is to examine whether broadly similar results would occur if solely Web data was used rather than all available bibliometric co-authorship data.

For this purpose a new approach of Web visibility indicators of collaboration is examined. The ensemble of COLLNET members is used to compare co-authorship patterns in traditional bibliometric databases and the network visible on the Web. One of the general empirical results is a high percentage (78%) of all bibliographic multi- authored publications become visible through search of engines in the Web. One of the special studies has shown Web visibility of collaboration is dependent on the type of bibliographic multi-authored papers.

The social network analysis (SNA) is applied to comparisons between bibliographic and Web collaboration networks. Structure formation processes in bibliographic and Web networks are studied. The research question posed is to which extent collaboration structures visible in the Web change their shape in the same way as bibliographic collaboration networks over time. A number of special types of changes in bibliographic and Web structures are explained.

Introduction

With the importance of collaboration in research and technology growing worldwide, it has become necessary to examine the processes involved in order to become aware of the implications for the future organization of research as well as those for science and technology policy.This has led to an increase in the number of scientific studies of this topic internationally. (GLÄNZEL, 2002; BORGMAN & FURNER, 2002).

The outstanding works of BEAVER (1978), PRICE (1963) and others on the topic of collaboration in science have, over a number of years, encouraged a number of scientists working in the field of quantitative scientific research to concentrate their research in this field. This has led both to an increase in the number of relevant

Received May 4, 2004

Address for correspondence:
HILDRUN KRETSCHMER
NIWI, The Royal Netherlands Academy of Arts and Sciences
P. O. Box 95110, 1090 HC Amsterdam, The Netherlands
E-mail: Hildrun.Kretschmer@niwi.knaw.nl

0138–9130/2004/US $ 20.00

publications concerning this topic in international magazines, and to an increase in the number of lectures in international conferences (BASU, 2001; BRAUN et. al., 2001; DAVIS, 2001; HAVEMANN, 2001; WAGNER-DÖBLER, 2001; KUNDRA & TOMOV, 2001).

The emergence of the Internet and the Web have led to changes in the process of scholarly publishing and communication, in the way that scientists and scholars search for and find information about patterns of national and international collaboration (HERRING, 2002; INGWERSEN, 1998; KLING & MCKIM, 2000). The influence of these new information and communication technologies (ICT) on collaboration in science and technology has also to be considered in light of the work on the topic of collaboration patterns, especially the question of the extent to which collaboration in science and in technology is visible on the Web.

Therefore in the year 2000 the time had come to create a global interdisciplinary research network, COLLNET, on the topic 'Collaboration in Science and in Technology' made up of 64 members from 20 countries of all continents. The members intended to co-operate on both theoretical and applied aspects on the topic 'Collaboration in Science and in Technology' (KRETSCHMER et al., 2001). The focus of this group is to examine the phenomena of collaboration in science, its effect on productivity, innovation and quality, and the benefits and outcomes accruing to individuals, institutions and nations of collaborative work and co-authorship in science as well as collaboration in e-science (More details of COLLNET, see Web site: www.collnet.de).

The EU has recently financed a new project (WISER) to further investigate the potential of creating new indicators of the Web for use in science and technology policy making. The study of collaboration in e-science is one focus of this project including the question of the extent to which collaboration structures visible in the Web follow similar rules as collaboration networks measured by traditional bibliometric data. About the half of the EU-project members are COLLNET members, too.

Therefore, in a pilot study, the co-authorship network of all of the COLLNET members from bibliometric data has been compared with the co-authorship network from webometric data. New webometric indicators are defined to measure the visibility of collaboration in the Web. COLLNET was selected for testing these new webometric indicators because of our personal familiarity with the COLLNET members, which gives rise to the possibility to make personal requests during testing. In addition, background information and explanations of special changes in both bibliometric and webometric network structures over a longer time period are possible.

Social network analysis (SNA) was used for the analysis of both the collaboration network measured by traditional bibliometric data and Web collaboration network.

The research question posed is to which extent collaboration structures visible in the Web follow the same rules as collaboration networks measured by traditional bibliometric data. Thus, the purpose of the study is to examine if we would get broadly similar results when just using Web data than all data.

Data

The last COLLNET data are from June 2003.

Sample set

The bibliographies and Web data of the 64 COLLNET members were examined, under them:

- 26 female and 38 male scientists,
- 30 members from the European Union (EU) and 34 from non-European Union countries (N).

From the 34 members from the non-European Union countries (N) we have :

- 3 from Australia,
- 7 from America (4 of them from North America),
- 19 from Asia,
- 4 from Eastern Europe,
- 1 from South Africa.

Bibliometric data

As usual, the bibliometric method for the study of collaboration is the investigation of co-authorships. Collaboration between countries, collaboration between institutions, or collaboration between individual scientists is examined in the literature (GLÄNZEL, 2002). In the present paper collaboration between COLLNET members is studied.

Beyond co-authored articles registered in SCI or other data banks, the range of entire collaboration between scientists is also reflected in all other publications, such as jointly authored books, manuscripts, etc.

Assuming that the reflection of collaboration in the Web is not limited to articles in SCI or other data bases, a request was made to all the 64 COLLNET members for their complete bibliographies, independently of the type of the publications and independently from the date of appearance of these publications.

As, for example, the range of collaboration between two scientists is much broader when writing a common book than when writing an article, this fact should become visible also in the Web.

From these bibliographies all publications were selected that appeared in co-authorship between at least two COLLNET members. Thus, it concerns 223 bibliographic multi-authored publications. From this, the respective number of common publications between two members was determined as the basis for the analysis of the co-authorship network.

The co-authorship network developed according to this method covers the entire lifetime collaboration between the COLLNET members.

Webometric data

Two different kinds of data collection for the study of collaboration in e-science are presented. On the one hand Web hyperlinks between homepages of scientists are collected and on the other new Web visibility indicators of collaboration are the basis for the data collection.

Homepages of COLLNET members and Web hyperlinks

TERVEEN & HILL (1998) report on an empirical investigation into emergent collaboration: "Links between web sites can be seen as evidence of a type of emergent collaboration among web site authors". The authors have used SNA for analysis of the link structures.

It was intended to use the same method for the analysis of hyperlink structures between the homepages of the COLLNET members.

17 COLLNET members had placed homepages on the Internet. However there were not any links between the homepages!

New Web visibility indicators of collaboration

According to VAUGHAN & SHAW (2003) Web citations refer to Web text citations or mentions of published papers on the Web. These authors searched for citations to each article on the Web, using the Google search engine. The search strategy was to enter the article title in quotation marks (i.e. phrase search in Google).

Among others there are different categories of citing items, for example the citation of a publication in the on-line version of an article or lists of bibliographies for the students or publication lists in own homepage, etc.

Vaughan and Shaw's method of searching article quotations in the Web (Web citations) was used successfully with the additional use of the Alltheweb search engine (www.alltheweb.com), albeit in a slightly modified form, to measure the visibility of the collaboration in the Web with the following definitions of new indicators:

- The *Web visibility rate of a multi-authored publication won by bibliographic data (WVP)* is measured as a frequency of the different Web sites on which this bibliographic publication is mentioned after entering the full title of the co-authored publication into Google or Alltheweb.

- The *Web visibility rate of a pair of collaborators (WVC)* is equal to the sum of Web visibility rates, WVP, of all of their co-authored publications.

In contrast to the measurement of Web citations by Vaughan and Shaw, who cite all pages of web sites on which an article is mentioned, here only the number of different Web sites on which a multi-authored publication is mentioned is used for measurement of Web visibility of bibliographic multi-authored publications. This decision was made as some Web site authors presented, at the same time, the same list of publications on several pages, only under different criteria of arrangement, for example the publication list in the CV: on a page "chronological" and on another page of the same Web site 'by subject'.

On the other hand, there are some reasons to count all pages of all Web sites as done by Vaughan and Shaw as there are, for example, authors of Web sites who have published two or more different articles on different pages on-line. If the same other publication is cited in these different online articles on the different pages (bibliographic coupling), then all these pages are counted and this other publication receives the appropriate number of Web citations.

By way of an example of the present study for the measurement of the Web visibility of bibliographic multi-authored publications of the COLLNET members, the number of different Web sites was selected as a method after detailed examination of the empirical results because there is a difference between counting Web citations and visibility of collaboration. In further investigations, however, the question of the best suitable method should be revisited, as it deals here with a pilot study with first results in the available paper.

Methods and results

In this paper, general results arising from testing Web hyperlinks and Web visibility indicators as possible suitable data collection strategies for the investigation of collaboration in e-science are presented. In addition, findings of the dependence of Web visibility on the type of the bibliographic multi-authored papers are also presented.

Social network analysis (SNA) is applied to a comparison between bibliographic and Web collaboration networks. Moreover, developmental and structural formation processes in bibliographic and Web networks are studied

General results

Homepages and Web hyperlinks. From the 17 COLLNET members, who have homepages on the Internet, are:

- 7 female (= 27% of the 26 female members) and 10 male (= 26% of the 38 males)
- 12 members from European Union countries (= 40% of the 30 European Union members) and 5 members from N countries (=15 percent of the 34 members from the non- EU countries)

While there appeared to be no difference between the female and male members, a tendency is apparent in favour of the EU when compared to the non-EU countries. A Chi-square test was performed on the EU/N data and the result shows a significant ($p<0.01$) difference between EU and N countries. It would be interesting to perform a similar investigation on a larger sample in the future.

The partitioning of European Union and non-European Union countries took place as Vaughan and Shaw found out that the number of Web citations through EU Web sites is the highest, followed by North America. They refer to a "general pattern relatively lower Web penetration and use beyond Europe and North America". This is in agreement, for example, with a comparison of the Web sites of the universities in the UK with the Web sites of the Indian universities (KRETSCHMER & THELWALL, 2003).

Due to the geographical proximity of the possible COLLNET co-authors in the EU and additionally due to the small number of COLLNET members from North America (only 4 members) only a rough partitioning was made for the investigation, i.e. EU and non EU countries. This division can be done in a more detailed manner in future investigations.

As already explained above, there are no hyperlinks between the homepages, even though collaboration between several COLLNET members exists (223 bibliographic multi-authored publications). From an investigation similar to that of TERVEEN & HILL (1998), it follows that necessary distance must be maintained.

In this connection it is worthy of note that the authors only stated: "The work reported here investigates links between Web sites as a potential ground for emergent collaboration". Thus they neither give an explanation nor empirical support for the validity of this statement.

In contrast to this there are empirical investigations regarding motivations and reasons for the creation of links between Web sites (WILKINSON et al., 2003; THELWALL, 2003). The motivations are various and arise from the fact that Web hyperlinks differ not only from bibliographic citations but also from co-authorships, whereas only a very reduced percentage is seen to be similar to bibliographic citations and a smaller percentage compared to co-authorships.

In a similar direction, concerning co-authorships, the results of an unpublished study of the two authors of the present paper point to the hyperlinks between homepages from approximately 2000 members of the German Society for Psychology (DGPs), in which the very small number of hyperlinks are not in a similar ratio to the high number of bibliographic multi-authored papers.

Web visibility indicators of collaboration. Contrary to the planned investigations of the structures of hyperlinks between homepages, here the slightly modified form of the method of Vaughan and Shaw proved to be very useful.

168 publications (= 73% of the 223 bibliographic multi- authored publications) became visible at least once in Google and 141 publications (= 63%) at least once in Alltheweb.

As mentioned above, the Web visibility rate of a bibliographic multi-authored publication (WVP) is measured as a frequency of the different Web sites, on which this publication is mentioned.

The Spearman correlation- coefficient between Google and Alltheweb amounts to R=0.67, statistically significant on the 0.01 level with 223 pairs.

The distributions of multi-authored publications with definite Web visibility rates (WVP), by Google and Alltheweb are won in Figure 1 represented (Google: full line, Alltheweb: broken line), whereby the value for WVP is limited to 10 in the figure since only very few publications received higher Web visibility rates.

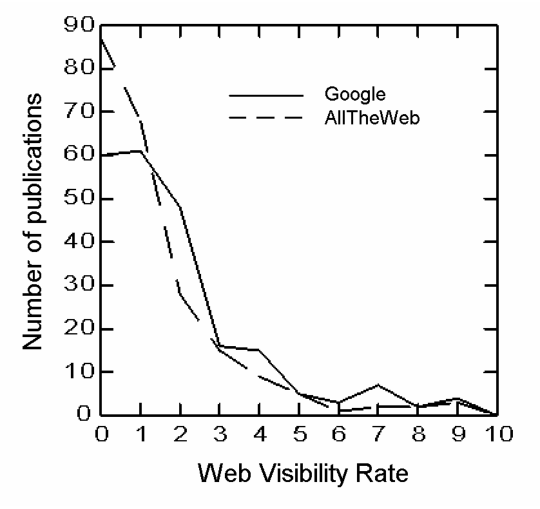

Figure 1. Distribution of publications

The distribution of multi-authored publications with definite WVP is represented also in a Two-Way frequency Table, see Figure 2. The WVP data are classified (class 1: WVP=1, class2: WVP=2, class3: WVP=3, class4: WVP=4, class5: WVP≥5).

Publications which achieve a lower value for WVP in Google than in Alltheweb, are arranged on the right of the diagonal (right corner), the other publications are on the left. On the right of the diagonal there are, however, fewer publications than on the left. That means reverse that the WVP per publication is higher in tendency in Google than in Alltheweb. This is in agreement with the positive evaluation of "Google" in relation to other search engines by Vaughan and Shaw. In the diagonal in Figure 2 we find those publications whose WVP is similar in Google and in Alltheweb.

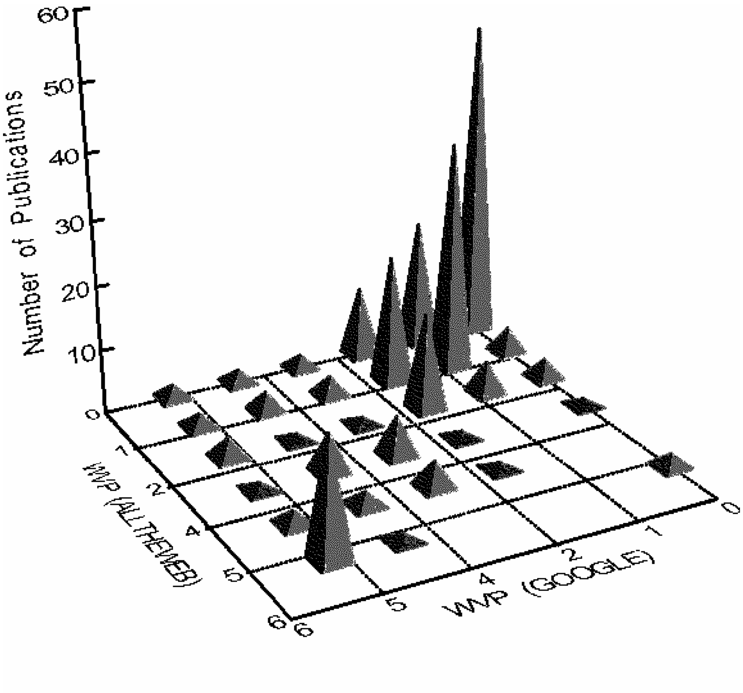

Figure 2. Two-way frequency table

Subsequently, the results of both the search engines per bibliographic multi-authored publication were summarized combining the results from both search engines and deleting the repeats. A higher value for WVP per bibliographic multi-authored publication results on average from this summary than by application of only one of the search engine. This is because the results of various search engines are not always identical.

It follows that a higher number of bibliographic multi-authored publications are also visible in the Web, here 173 (=78%).

As a conclusion, these results show that it is better to use several search engines at the same time in future investigations for Web visibility rate per multi-authored publication.

Forty of these 173 (= 23%) Web visible multi- authored publications are visible under other kind of Web sites also on the COLLNET Web site: www.collnet.de, and 115
(= 66%) also on the personal homepages of the COLLNET members or on the Web sites of their departments.

In summary it can be stated that bibliographic multi-authored publications which were investigated in the pilot study are visible to a high percentage in the Web and that it follows, therefore, that collaboration between scientists is well reflected in the Web. It is important to examine in the following whether the original bibliometric collaboration structures remain intact, or whether any specific changes develop in the Web regarding gender or countries.

Dependence of Web visibility rates on types of bibliographic multi-authored papers. The 223 bibliographic multi-authored publications are classified according to their type

– books,
– articles in peer reviewed journals,
– contributions in monographs,
– articles in conference proceedings or manuscripts

into the following categories:

1. articles in *Scientometrics,*
2. articles in *JASIS,*
3. both papers in monographs and articles in other journals than *Scientometrics* and *JASIS,* (The number of articles in other journals is less than 6 per journal),
4. papers from conference proceedings and manuscripts,
5. books.

The difference between these categories and Web visibility rates is studied Table 1.

Table 1. Classification of bibliographic multi-authored publications and visibility in the Web

Categories	Number of bibliographic multi-authored publications, n	Sum of Web visibility, Σ WVP	Average Web visibility, ΣWVP/n	Number of non-Web visible publications with WVP=0, m	Percentage of non-Web visible publications, 100m/n	Number of forthcoming publications (All are non-Web visible publications)
1	55; (0)	151	2.75	9; (0)	16	6; (0)
2	13; (0)	71	5.46	1; (0)	8	1; (0)
3	68; (15)	175	2.57	19; (7)	28	0; (0)
4	77; (12)	88	1.14	21; (5)	27	2: (0)
5	10; (2)	149	14.9	0; (0)	0	0; (0)
Total sum	223; (29)	634	2.84	50; (12)	50/223=0.22	9; (0)

Note: The numbers of non-English publications are in brackets

There is a statistically significant difference between the distribution of bibliographic multi-authored publications (n) and the distribution of Web visibility rates (Sum WVP) along the categories (Chi-square test: $p<0.01$). The average Web visibility rate is highest for books (14.9) and lowest for conference proceedings and manuscripts (1.14). It is an empirical proof for the dependence of Web visibility rate on type of bibliographic multi-authored papers

The highest percentage of non-Web visible publications (WVP=0) can be found in categories 3 and 4. It is related to papers from conference proceedings and manuscripts, and to papers in monographs and articles in other journals than *Scientometrics* and *JASIS*.

The sum of non-English publications with WVP=0 and forthcoming publications with WVR=0 (12+9=21) is equal to 42% of the total number of non-Web visible publications.

A high percentage (=67%) of the non-Web visible articles in *Scientometrics* are forthcoming articles (6/9=0.67). This phenomenon is not valid in categories 3 and 4. Only 5 % of the non-Web visible publications are forthcoming.

Social Network Analysis (SNA)

OTTE & ROUSSEAU (2002) recently showed that social network analysis (SNA) can be used successfully in the information sciences, as well as in studies of collaboration in science. The authors showed interesting results by the way of an example of the co-authorship network of those scientists who work in the area of social network analysis.

Otte and Rousseau refer in their paper to the variety of the application possibilities of SNA, as well as to the applicability of SNA to the analysis of social networks in the Internet (webometrics, cybermetrics).

Therefore, this paper examined, the extent to which scientific collaboration in the Internet becomes visible. Thus it deals with:

- Examinations using SNA to establish the extent to which the bibliographic COLLNET co- authorship network gets reflected in the Web and how similar the networks are.
- Examinations of the development of both the bibliographic COLLNET co-authorship network and the Web network over a specific time period. The results are presented in a separate chapter.

Bibliographic Co-authorship network. The methods of social network analysis (SNA) are related to WASSERMANN & FAUST (1994) and to OTTE & ROUSSEAU (2002).

- There are 64 ‘nodes’ (= 64 COLLNET members) in the network.

- 48 of these COLLNET members (= 75%) have published in co-authorship at least once with at least one of the other COLLNET members. That means, at least 1 ‘edge’ is adjacent to each of these 48 ‘nodes’.
- Differently expressed: Between two COLLNET members A and B, there exists an edge if both have published at least one publication in co-authorship. The members A and B are called *pair of collaborators* (A,B).
- There are L_B=63 edges between the nodes or in other words 63 different pairs of collaborators, respectively.
- A path from node X to node Y is a sequence of distinct edges between pairs of collaborators: (X, A_1), (A_1, A_2), …, (A_j, Y). The length of the path is equal to the number of distinct edges. The shortest path from X to Y is called *distance* d_{XY}.
- The co-authorship structure of COLLNET is a ‘disconnected graph’, i.e., there is not a ‘path’ between each pair of nodes X and Y. However the COLLNET members can be divided into several ‘connected subsets’. A path also exists between all pairs of nodes in a ‘connected subset’. The ‘connected subsets’ are denoted as ‘components’ or ‘clusters’.
- However between a pair of nodes from different components there exists no path.
- The COLLNET co-authorship network consists of 23 components:
  - 1 large central component of 32 members
  - 1 component of 4 members
  - 2 components of 3 members
  - 3 components of 2 members
  - 16 singletons

The largest cluster covers 50% of the COLLNET members. In addition there are 22 small and very small (singletons) clusters.

This structure of clusters, which contain a single very large cluster and also a large number of small clusters, is in agreement with the existing findings in the literature (NEWMAN, 2001; GENEST & THIBAULT, 2001; KRETSCHMER, 2003; OTTE & ROSSEAU, 2002). It is possible this could denote a general rule in a special type of co-authorship network.

The density of a co-authorship network (D) is an indicator for the level of connectedness of this network:

D = Number L of edges divided by the maximum number L_{max} of edges in the network.
It is a relative measure with values between 0 and 1.
L_{max}=V (V-1)/2
D = 2L / V(V-1)

The studied bibliographic co-authorship network is a network with low density of D_B=0.031.

Web co-authorship network. The bibliometric network indicates that 48 COLLNET members have published at least once in co-authorship with at least one of the other COLLNET members. 44 of them (92%) are visible as co-authors in the co-authorship network obtained from the Web.

There are L_W= 56 edges (56 pairs of collaborators) in the Web network, i.e. 89% of the edges obtained from bibliographies.

The Web visibility of a pair of collaborators (WVC) may possibly grow exponentially with the number of their bibliographic co-authored publications (Figure 3). The connection mentioned should be examined in future investigations on larger samples. There is no statistical significance for the present data.

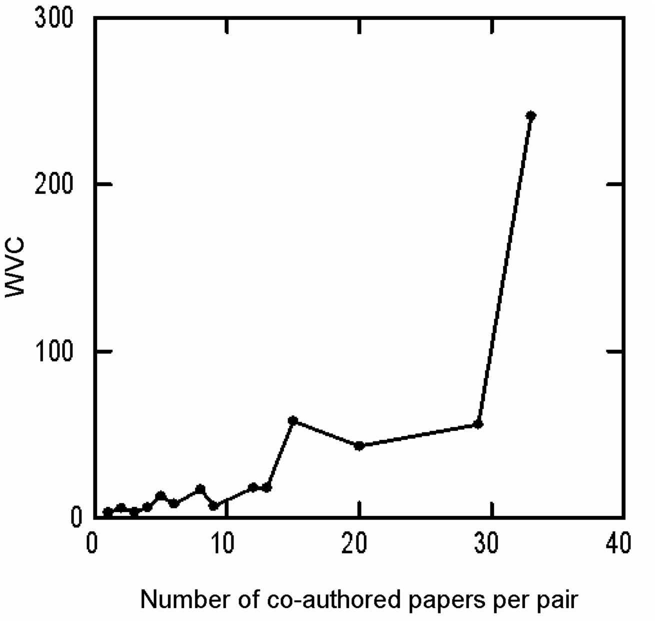

Figure 3. Web visibility rate of a pair of collaborators

The structure of the network obtained from the Web is similar to the structure of the network obtained from the bibliographies (Figures 4 and 5). The large central component obtained from the bibliographies is slightly reduced on the Web by 4 members (28/32: 88%), producing a new component of 4 members (cf. Figure 5 down on the right side). 2 components of 2 members each fall to pieces, i.e. 4 singletons.

Thus, the disconnected graph of the Web co-authorship network is partitioned into 26 components:

- 1 large central component of 28 members
- 2 component of 4 members
- 2 components of 3 members
- 1 component of 2 members
- 20 singletons

The largest cluster covers 44% of the COLLNET members. In addition there are 25 small and very small clusters (singletons).

From another point of view the change from the bibliographic to the Web network consists of 7 missing pairs of collaborators only. These missing edges in the Web are related to edges with a maximum of 1 multi-authored publication each obtained by the bibliographies. Four of these publications are forthcoming. We expect to find these publications in the Web some time after appearance of the present forthcoming publications.

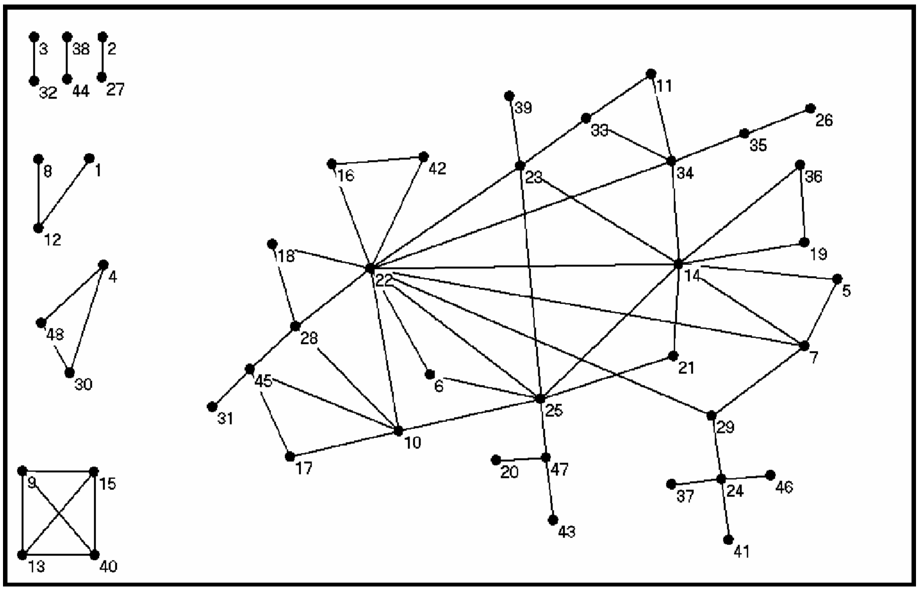

Figure 4. Network obtained from the bibliographies (2003)
(For explanation see the Appendix)

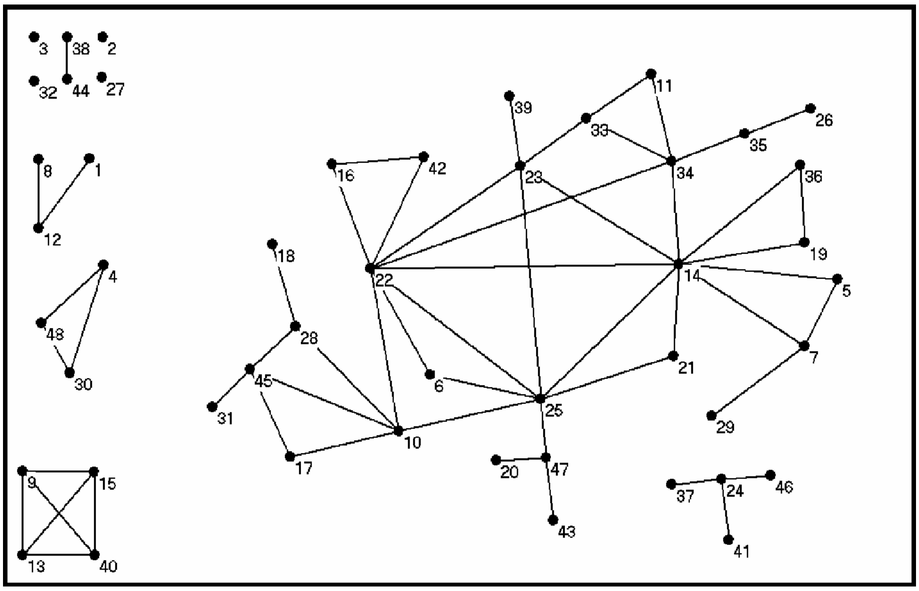

Figure 5. Network obtained from Web (2003)
(For explanation see the Appendix)

The Web co-authorship network is a network with the density of D_W=0.028. The densities of both networks are loose but the difference is very low.

Development of bibliographic and Web networks and structure formation processes

The research question of the next part of this paper is to which extent collaboration structures visible in the Web are changing their shape in similar ways as the bibliographic collaboration networks over a specific time period. Is there any similarity between structure formation processes in bibliographic and in Web networks?

Is there some explanation for, or background information to, special changes of the structures over time, and is there any explanation for the arising slight differences between bibliographic and Web networks over the last time stage?

In answer to these questions the development of COLLNET was used as presupposition for the division of the studied time period into 4 stages.

Development of COLLNET in brief and corresponding 4 stages of the studied time period.

First Step of the Development of COLLNET (1998–1999). An important trigger in the creation of COLLNET was the first Berlin Workshop on Scientometrics and Informetrics/Collaboration in Science that took place at the Institute of Library Science of the Humboldt University, Berlin, in August 1998. This workshop was organized by the Society of Science Studies (Gesellschaft fuer Wissenschaftsforschung e.V., Berlin), and supported by the Free University Berlin, and DFG (German Research Foundation).

Second Step (2000–2001). Two years later in September 2000, in conjunction with the Second Berlin Workshop on Scientometrics and Informetrics/Collaboration in Science and in Technology, the first COLLNET meeting was held at the Free University Berlin (A special issue of the journal *Scientometrics* is published in 2001 about selected papers). From this time on, COLLNET meetings have been held regularly: the Second COLLNET Meeting at the National Institute of Science, Technology and Development Studies, New Delhi (India). Again, COLLNET used the synergy of conjoint activity with the international workshop "Emerging Trends in Science and Technology Indicators: Aspects of Collaboration".

A third COLLNET Meeting took place in July 2001 in Sydney (Australia) in conjunction with the 8th International ISSI conference on Scientometrics and Informetrics.

Third Step (2002–2003). Future strategies were discussed at the 4th COLLNET Meeting held in conjunction with the 9th International Conference on Scientometrics & Informetrics in Beijing, China during August 2003. At that time, further measures of the effectiveness of these collaborative engagements among members and productivity in the field of 'collaboration in science and in technology' were discussed.

Thus these 3 steps, along with the additional inclusion of the preliminary stage, will be incorporated to show the development of both the bibliographic COLLNET co-authorship network and the Web network in 4 stages:

- Until 1997: Collaboration of the future COLLNET members before 1998 (preliminary stage)
- Until 1999: Collaboration until 1999 (cumulative, including collaboration until 1997, i.e. preliminary stage and first step of COLLNET development)
- Until 2001: Collaboration until 2001 (cumulative, including collaboration until 1997, first and second steps)
- Until 2003: Collaboration until 2003 (cumulative, including collaboration until 1997, first, second and third steps)

Development of bibliographic and Web networks. The growth of the number of edges (pairs of collaborators), the decreasing number of components, the growth of the large component and the decreasing number of singletons along the 4 stages are of interest.

In addition, we shall also focus on some selected indicators of centrality describing the structure of networks and the role played by particular nodes (in analogy to OTTE & ROUSSEAU, 2002; WASSERMANN & FAUST, 1994):

– Degree Centrality

– Betweenness

Degree Centrality of a node A is equal to the number of nodes (or edges) that are adjacent to A:

$DC_A=E_A$.

The Degree Centrality of a node A is equal to the number of his/her collaborators or co-authors. An actor (node) with a high degree centrality is active in collaboration. He/she has collaborated with many scientists.

Mean Degree Centrality (MDC) of the network is the ratio of the sum of the Degree Centralities of all the nodes in the network to the total number of nodes:

$MDC=2L/V$.

Betweenness Centrality BC_A is the number of shortest paths (distance d_{xy}) that pass through A. Otte and Rousseau mention actors (nodes) with a high betweenness play the role of connecting different groups or are ‘middlemen’. WASSERMAN & FAUST (1994, p. 188) mention: *Interactions between two nonadjacent actors might depend on the other actors in the set of actors who lie on the paths between the two. These “other” actors potentially might have some control over the interactions between the two nonadjacent actors.* A particular “other” actor in the middle, the one *between* the others, has some control over paths in the network.

$BC_A=\Sigma_{X,Y}\ G_{XAY}/\ G_{XY,}$

G_{XAY} is the number of shortest paths from node X to node Y passing through node A. G_{XY} is the number of shortest paths from node X to node Y (X,Y≠A).

The general formula:

$C_{NETWORK}=(\Sigma_X\ (C_{max}-C_X))$/max value possible

can be applied for determining degree, closeness or betweenness centrality for the whole network (In detail, cf. OTTE & ROUSSEAU, 2002). These measures are relative measures with values between 0 and 1.

The development of collaboration between COLLNET members is studied in connection with the visibility of this network on the Web.

The indicators density, mean degree centrality and betweenness centrality are applied in analyses of both bibliographic co-authorship network and Web network. The general formula is applied for Betweenness. Furthermore the development of number of edges, number of components, number of singletons and the size of largest component (number of nodes in the largest component) are studied (Table 2).

The values of the indicators describing the structure of networks (density, mean degree centrality and betweenness) increase from 1997 to 2003 with a particular rise from 1999 to 2001 (cf. Figure 6 as example). The probability is high that both the foundation of COLLNET and first COLLNET meeting in 2000 maybe the reasons for this increase.

Table 2. Development of bibliographic and Web networks

	1997	1999	2001	2003
Number of edges or of pairs of collaborators	16	25	47	63
	14	22	45	56
Number of components	48	44	30	23
	51	47	32	26
Number of singletons	39	36	22	16
	44	40	25	20
Size of largest component	7	11	23	32
	6	9	22	28
Density	0.008	0.012	0.023	0.031
	0.007	0.011	0.022	0.028
Mean degree centrality of the network MDC	0.53	0.78	1.47	1.97
	0.44	0.68	1.38	1.75
Betweenness	0.008	0.028	0.101	0.22
	0.005	0.017	0.096	0.11

Note: The first value in each cell is the bibliographic value and the second value is the Web value

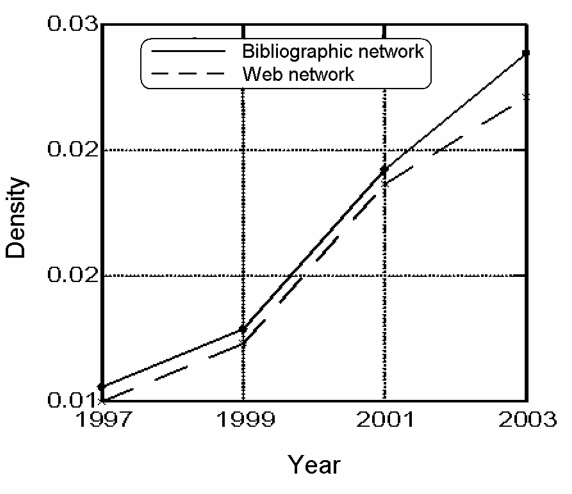

Figure 6. Density

The values for the Web structure increase in parallel to the bibliographic network.

On average, the values for the Web structure are slightly lower than the average values for the bibliographic structure. However, the Web values reach the bibliographic values or exceed them after two years. For example, the betweenness of the bibliographic network in 2001 is equal to 0.101 and the value of betweenness of the Web network is equal to 0.11 in 2003.

The growth in the number of pairs of collaborators (edges) is in correspondence with the growth of density.

Structure formation process measured by entropies. Whereas the size of the largest component grows, the number of components and the number of singletons diminish (cf. Table 2). This kind of structure formation processes in both the bibliographic and the Web networks can be measured by entropies H:

There is a series of numbers $K_f(f=1,2,\ldots z)$, $K_f \neq 0$

$$h_f = K_f / \sum_{f=1}^{z} K_f$$

$$H_f - \sum_{f=1}^{z} h_f \times lg_2\, h_f$$

K_f is the size of a component f. The number of components in the network is called z.

The entropy H is decreasing with increasing size of the components and with decreasing number of components. The maximum entropy H is reached in a network under the condition there are singletons only. The minimum entropy is reached under the condition where there is one large cluster only and there are not any other components.

The structure formation processes both in the bibliographic network and in the Web network are shown in Figure 7.

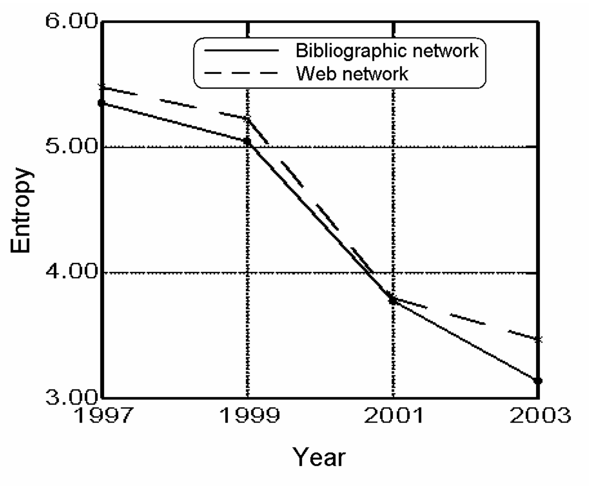

Figure 7. Structure formation process measured by entropies

The development of structure in both networks can be visualised by the maps drawn with Pajek (Figure 8). The co-authorship networks from bibliometric data are presented in the left column and the networks from webometric data in the right column.

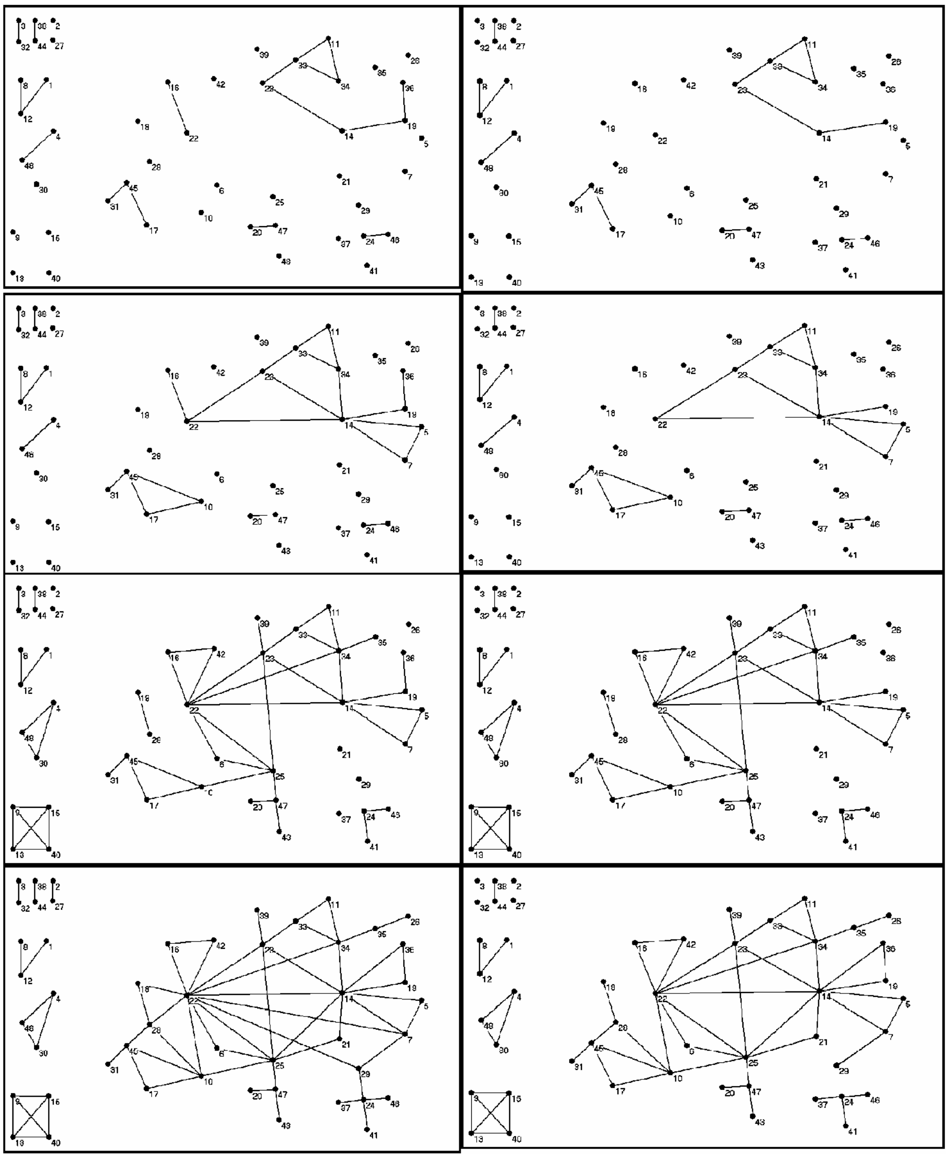

Figure 8. Development of structure in both the bobliographic and the Web network
(For explanation see the Appendix)

Both networks in the first row are from the stage of collaboration until 1997, in the second row from the stage until 1999, in the third row from the stage until 2001 and the 4th row from the stage until 2003.

Horizontal: The co-authorship network from bibliometric data and the corresponding co-authorship network from webometric data are very similar up to very slight deviations at the same stage.

Vertical: The structure formation process is characterized by the growth of the number of edges (pairs of collaborators), the decreasing number of clusters, the growth of the large cluster and the decreasing number of singletons. It is valid for both bibliometric and webometric data. Slight differences between bibliometric and Web networks arise in the last stage. The explanation for this phenomenon may be that forthcoming publications can be found in the personal bibliographies of the COLLNET members but these publications are visible some time later in the Web after publishing.

Discussion and conclusion

The following ideas are the result of a round table discussion with all members of the WISER project.

A new approach of Web visibility indicators of collaboration is examined. The ensemble of COLLNET members is used to compare co-authorship patterns in traditional bibliometric databases and the network visible on the Web. As Sylvan Katz pointed out, the question of quality control is also involved. It makes a difference if the collection of data is based on a database which includes peer reviewed journals only (as in the case of SCI) or only taken from the Web with a mixture of peer reviewed and not reviewed articles.

In the discussion, Mike Thelwall pointed out that the similarity of Web based structures and SCI based structures might also be used to replace (costly) ISI products with publicly available information on the Web.

The comparison of collaboration patterns on-line and off-line needs an off-line starting point, either in the form of a bibliography in the field or a list of names of authors. If one starts with a list of co-authored publications the “Web visibility indicator” gives an indication for the visibility of this collaboration on-line, and there might be more articles visible on the Web than in bibliometric databases. In this case the Web offers additional information.

If one would start from the list of authors one could expect to detect other forms of collaboration than in the form of co-authorship (Viv Cothey). If search engines are to be used, one could use the option to look for relations in specific file types. This approach could reduce the noise in the information on the Web (Isidro Aguillo, Mike Thelwall).

However, an unresolved discussion point remains the interpretation of the meaning of a hyperlink in terms of collaboration.

*

This paper is based on a presentetion given at the *International Workshop on Webometrics, Informetrics and Scientometrics* and *8th International COLLNET Workshop* held in Roorkee (India), on 2-5 March, 2004.

This work was supported by a grant from the Common Basis for Science, Technology and Innovation Indicators part of the Improving Human Research Potential specific programme of the Fifth Framework for Research and Technological Development of the European Commission. It is part of the WISER project (Web indicators for scientific, technological and innovation research) (Contract HPV2-CT-2002-00015) (www.webindicators.org).

We thank Mike Thelwall for his important comments and proposals as well as we thank for the round table discussion with all members of the WISER project. Editorial discussions with Colin Reddy are acknowledged.

References

BALABAN, A. T., KLEIN, D. J. (2002), Co-authorship, rational Erdös numbers, and resistance distances in graphs, *Scientometrics*, 55 : 59–70.

BASU, A., AGGARWAL, R. (2001), International collaboration in science in India and its impact on international performance, *Scientometrics,* 52 : 379–394.

BATAGELJ, V., FERLIGOJ, A., DOREIAN, P. (1992), Direct and indirect methods for structural equivalence, *Social Networks,* 14 : 63–90.

BEAVER, D. DEB., ROSEN, R. (1978), Studies in scientific collaboration. Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics,* 3 : 231–245.

BORGMAN, C. L., FURNER, J. (2002), Scholarly communication and bibliometrics. In: B. CRONIN (Ed.), *Annual Review of Information Science and Technology,* Vol. 36, Medford, NJ: Information Today,pp. 3–72.

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

DAVIS, M., WILSON, C. S. (2002), Elite researchers in ophthalmology: Aspects of publishing strategies, collaboration and multi-disciplinarity. *Scientometrics,* 52 : 395–410.

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

GENEST, C., THIBAULT, C. (2001), Investigating the concentration within a research community using joint publications and co-authorship via intermediaries. *Scientometrics*, 51 : 429–440.

HAVEMANN, F. (2001), Collaboration behaviour of Berlin life science researchers in the last two decades of the twentieth century as reflected in the Science Citation Index, *Scientometrics,* 52 : 435–444.

HERRING, S. C. (2002), Computer-mediated communication on the Internet. In: CRONIN, B. (Ed.), *Annual Review of Information Science and Technology,* Vol. 36, Medford, NJ: Information Today Inc., pp. 109–168.

INGWERSEN, P. (1998), The calculation of Web Impact Factors. *Journal of Documentation*, 54 (2) : 236–243.

KLING, R., MCKIM, G. (2000), Not just a matter of time: field differences in the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51 (14) : 1306–1320.

KRETSCHMER, H., LIANG, L., KUNDRA, R. (2001), Foundation of a global interdisciplinary research network (COLLNET) with Berlin as the virtual center, *Scientometrics,* 52 : 531–538.

KRETSCHMER, H., THELWALL, M. (2003), The development of information professionals: The European perspective- The way from librametry to webometrics. In: A. AMUDHAVALLI (Ed.), *Proceedings of the MALA Platinum Jubilee Celebrations, Seminar on Information Professionals for the Digital Era*, Madras, India, January 29-30, 2003, EFEX: Chennai, pp. 13–25.

KRETSCHMER, H. (2003), Author productivity and Erdős distances in co-authorship and in Web networks. In: *Proceedings of the 9th International Conference on Scientometrics and Informetrics,* Beijing, August 25–28, 2003 (forthcoming).

KUNDRA, R., TOMOV, D. (2001), Collaboration patterns in Indian and Bulgarian epidemiology of neoplasms in *Medline* for 1966–1999.

NEWMAN, M. (2001), The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the USA*, 98 : 404–409.

OTTE, E., ROUSSEAU, R. (2002), Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28 : 443–455.

PRICE, D. J. DE SOLLA (1963), *Little Science, Big Science*, New York: Columbia University Press.

SCHUBERT, A. (2002), The Web of Scientometrics. A statistical overview of the first 50 volumes of the journal. *Scientometrics*, 53 : 3–20.

TERVEEN, L. G., HILL, W. C. (1998), Evaluating Emergent Collaboration on the Web, In: *Proceedings of CSCW 1998* Seattle WA, ACM Press, pp. 355–362.

THELWALL, M. (2003), What is the link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8.

VAUGHAN, L., SHAW, D. (2003), Bibliographic and Web citations: What is the difference? *Journal of the American Society for Information Science and Technology*, 54 (14) : 1313–1322.

WAGNER-DÖBLER, R. (2001), Continuity and discontinuity of collaboration behaviour since 1800- from a bibliometric point of view, *Scientometrics*, 52 : 503–518.

WASSERMAN, S., FAUST, K. (1994), *Social Network Analysis. Methods and Applications*. Cambridge: Cambridge University Press.

WILKINSON, D., HARRIES, G., THELWALL, M., PRICE, L. (2003), Motivation for academic web site interlinking: evidence for the web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29 : 59–66.

Appendix

Explanation for Figures 4, 5 and 8

1. Aguillo Isidro
2. Ahrweiler Petra
3. Ambuja R.
4. Bassecoulard Elise
5. Basu Aparna
6. Beaver Donald deB.
7. Bhattacharya Sujit
8. Bordons Maria
9. Brandt Martina
10. Davis Mari
11. Egghe Leo
12. Gomez Isabel
13. Grosse Ulla
14. Gupta Brij Mohan
15. Hartmann Frank
16. Havemann Frank
17. Hood William W.
18. Jansz Margriet
19. Karisiddappa
20. Katz Sylvan
21. Kharbanda Ved Prakash
22. Kretschmer Hildrun
23. Kundra Ramesh
24. Leydesdorff Loet
25. Liang Liming
26. Liberman Sofía
27. Liu Zeyuan
28. Markusova Valentina
29. Meyer Martin
30. Okubo Yoshiko
31. Osareh Farideh
32. Raghavan Koti S.
33. Rao Ravichandra
34. Rousseau Ronald
35. Russell Jane
36. Sangam Shivappa
37. Scharnhorst Andrea
38. Schulze Annedore
39. Tomov Dimiter
40. Voss Rainer
41. Wagner Caroline
42. Wagner-Döbler Roland
43. Wang Yan
44. Wenzel Vera
45. Wilson Concepcion S.
46. Wouters Paul
47. Wu Yishan
48. Zitt Michel

49.-64. are singletons up to June 2003. The 16 singletons are not included in the figure.