ISSI
international society for scientometrics and informetrics

issi newsletter #7

quarterly e-zine of the international society for scientometrics and informetrics
vol. 02 / nr. 3
september 2006

# THE 9TH INTERNATIONAL CONFERENCE ON SCIENCE AND TECHNOLOGY INDICATORS (LEUVEN, BELGIUM)

Unlike the ISSI conference series, well-known to the Society Members, the Science and Technology Indicators conference series focuses on policy relevant aspects of quantitative S&T research, on the application and applicability of scientometrics, technometrics, informetrics and webometrics in a science policy and research management. And unlike the ISSI conferences, which are organised almost all over the world, in Europe, in North and Central America, in the Middle and Far East and in Australia, the S&T Indicator conferences are at home in Leiden (the Netherlands). The organisation of this series is an initiative by Professor *Anthony van Raan* and his research centre CWTS at the Leiden University. Nonetheless, there is a good tradition to alternately co-organise these conferences with other research centres in Europe. Previous conferences have thus been organised in Leiden (1988), Bielefeld (1990), Leiden (1991), Antwerp (1995), Cambridge (1998), Leiden (2000), Karlsruhe (2002) and again in Leiden

## CONTENTS

## Editorial Board

(2004). This year, on 7–9 September, the 9th STI conference has been hosted and organised by the Katholieke Universiteit Leuven (Belgium). The motto of this meeting was "New Challenges in Quantitative Science and Technology Research". Electronic communication, the worldwide emphasis on the knowledge-based society, increasing internationalisation and globalisation, on one hand, and the strengthened role of regions, on the other hand, result in more complex and dynamic S&T systems with the demand for more sophisticated instruments of measurement. Taking into account these latest developments in our discipline and its applications, the programme comprised several plenary and special sessions organised by experts in the corresponding 'hot topics', above all, the sessions on "Webometrics for science and technology indicators" organised by *Liwen Vaughan* (CAN), "Indicators for Science and Technology linkage" organised by *Martin Meyer* (GBR/BEL) and "Statistical Properties of Bibliometric Indicators and Applications of the Hirsch-Index" organised by Anthony van Raan (NLD) should be mentioned in this context.

The conference was opened by *Fientje Moerman*, the Vice Minister-President of the Flemish government and Flemish Minister for Economy, Enterprise, Science, Innovation, and Foreign Trade. In her speech she stressed among others the tasks and challenges the western democracies characterised as being in transition from industrial to knowledge-based society are faced with. This transition is accompanied by growing competition by the emerging nations of Asia. Another important issue addressed by the minister is the transformation in the European Union and the resulting tasks for education and R&D policies. Finally, typical questions of strengthening policymaking in a medium-sized European region like Flanders were discussed.

This opening lecture was followed by a plenary session tackling the issues addressed by Fientje Moerman; the emergence of new, formidable competitors in science and technology, benchmarking the European Research Area's integration and the development paths of knowledge-based technologies were among the topics of this session.

Most sessions of the conference as well as the social events took place in the "Groot Begijnhof" [link] with its marvellous ambiance of the late Middle Ages and early modern era. The walk in the evening of the first conference day took the participant from the Beginage through the old city to the magnificent city hall [link] where the mayor of Leuven welcomed the participants at a reception. Beyond any doubt, the ultimate highlight of the conference was the musical accompaniment at the conference dinner on Friday night; two participants of the conference, Dr. András Schubert (clarinet) and Drs. Balázs Schlemmer (piano) made it a wonderful and unforgettable event.

The conference was sponsored by the Research Foundation Flanders (F.W.O.), K.U. Leuven, the Municipality of the City of Leuven, the Ministry of the Flemish Community, the Science & Innovation Administration (AWI) and the Steunpunt O&O Statistieken van de Vlaamse Gemeenschap.

**Wolfgang Glänzel**
Programme Chair

## CORRECTION TO THE PREVIOUS NEWSLETTER

It were Rory Wilson **and John Lancaster**, from Swansea University (Wales) who proposed the referee factor. Indeed, in the 29 June 2006 issue of Nature a correction has been published on page 1048 stating that "The name of the co-author, John Lancaster, was accidentally left off the Correspondence letter 'Referee factor' would reward a vital contribution" (Nature 441, 812; 2006).

Hence the reference to my editorial "Biologist Rory Wilson proposes a referee factor" and my comments "After the journal impact factor and the web impact factor a referee factor enters the fray: some comments" both published in the 2(2) issue of the ISSI Newsletter should have been:

Wilson, R. and Lancaster J. (2006) "Referee factor' would reward a vital contribution. Nature, 441, p. 812.

Actually, if I had known, the title of the editorial would have been " Biologists Rory Wilson and John Lancaster propose a referee factor".

**Ronald Rousseau**

# THE 9TH STI CONFERENCE IN FIGURES

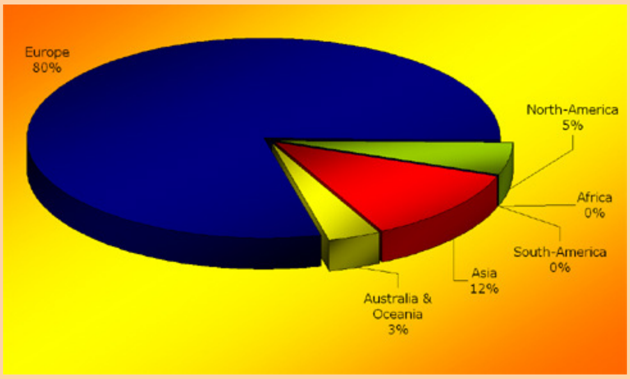
(Balázs Schlemmer)

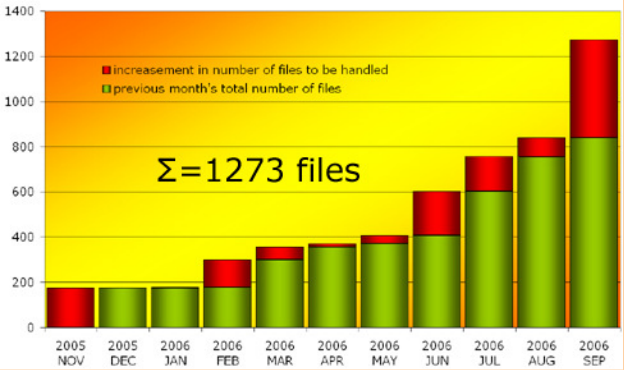
### On the one side:

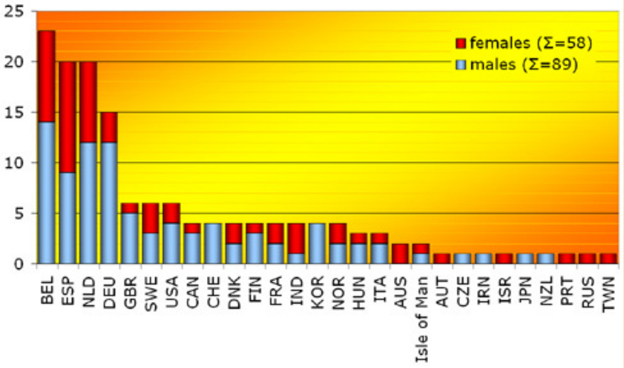
- over one thousand incoming and outgoing e-mails;
- several kilograms of printed papers;
- a website consisting of 649 files at the moment (still growing);
- further 624 files related to the conference, that is:
- 1273 files altogether gone through the hands of the conference organisers;
- 2 fat ring binder dossiers full with contracts, offers and other important documents;
- 125 abstracts to be sent out to minimum 2 referees each;
- still over 100 abstracts to be cleaned up after the refereeing procedure;
- countless hours of briefing, planning, phoning, e-mailing, acquiring, double-checking, re-calculating, refining;
- and a few false registrations even from criminals.

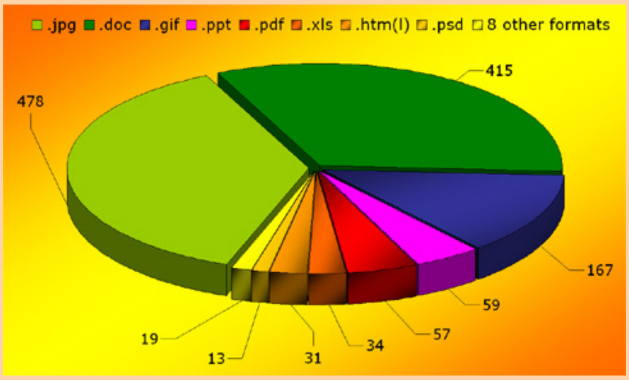
### On the other side:

- approximately 150 participants;
- a welcome reception in the Faculty Club;
- an opening speech by Fientje Moerman, Vice Minister-President of the Flemish Government and Flemish Minister for Economy, Enterprise, Science, Innovation, and Foreign Trade;
- 61 oral presentations;
- 32 poster presentations;
- 3 special sessions;
- 1 ISSI board meeting;
- a 304-page-long book of abstract;
- a marvellous trip for the accompanying persons to the Meuse Valley (Gardens of Annevoie, Maredsous Abbey, Dinant Citadel);
- a guided city tour in Leuven;
- a reception at the City Hall, given by Louis Tobback, mayor of Leuven;
- and a cosy conference dinner in the most prestigious Faculty Club.









# A new Hirsch-type index saves time and works equally well as the original h-index

Marek Kosmulski

Department of Electrochemistry, Lublin University of Technology
PL-20618 Lublin (Poland)
e-mail: mkosmuls@hektor.umcs.lublin.pl

**Abstract**

A scientist's h(2) index is defined as the highest natural number such that his h(2) most-cited papers received each at least $[h(2)]^2$ citations. The advantage of h(2) as the index to characterize the scientific output of an individual over the original h index (HIRSCH, 2005) is that less work is required to verify the authorship of the relevant papers.

(HIRSCH, 2005) introduced an index h defined as the number of papers of certain author with citation number of at least h, and calculated the values of the h-index of Nobel-winners and of other top physicists and biologists. Typical h-values in this group were about 50, and the highest h of an individual was 191. The h-index gained popularity after a publication in Nature (BALL, 2005). BORNMANN & DANIEL (2005) discussed the usefulness of the h-index in assessment of relatively young scientists (h-index of about 3), but in the opinion of the present author the h-index is more suitable for assessment of mature scientists who have published at least 50 papers and have h-indexes of at least 10.

The h-indexes of all full professors affiliated at a department of chemistry of one university in Poland have been calculated and analyzed in detail. Sufficient information was available to assess if the papers found in the database (Thomson Scientific) were authored by the individual of interest of by someone else who happens to have the same family name and first name initial(s). Even the maiden names of the professors who took their husbands' names after marriage were know, but (in this particular case) the pre-marriage publications did not improve their h-indexes. For 14 out of 19 professors, the automatic search of the database immediately produced the correct h-index. For the other 5 professors, the automatic search of the database produced an overestimated h-index. This was because the other scientists, who accidentally have the same family name and first name initial(s) have also published frequently cited papers. The verification of the authorship was time-consuming, and without sufficient knowledge of the scientific CVs of the individuals of interest, the results would be very uncertain. In other words, the calculation of the h-index would be very easy provided that each scientist had an unique combination of family name and initials, but obviously this is not the case. This problem was not mentioned by HIRSCH (2005), although he must have encountered it. For example the automatic search for M.L.Cohen produces a combined h-index of the physicist from California (third-highest h-index among physicists, according to HIRSCH) and of another M.L.Cohen representing medicine, who also has an outstanding citation record. Hopefully HIRSCH solved this problem properly, but for a non-expert user of the database it is even difficult to establish how many different M.L.Cohens contributed to the fantastic result produced by the automatic search of the database.

The verification of the authorship is the most difficult and time consuming step in the calculation of the h-index. The time required for the analysis can be reduced when a new index is used rather than h. A scientist's h(2) index is defined as the highest natural number such that his h(2) most-cited papers received each at least $[h(2)]^2$ citations. The h(2)-index can be established by looking at the list of papers of an individual ordered by number of citations in the Thomson Scientific database. Most natural scientists and engineers know the squares of natural numbers up to 16 by heart, and $h(2)>16$ occurs very seldom. For example a h(2) of 10 denotes that 10 papers were cited at least 100 times each. The total number of citations of a scientist is at least $[h(2)]^3$, and usually it is higher than $[h(2)]^3$ by a factor of about 5 (analogous problem for the original h-index was discussed in detail by HIRSCH). Thus, the h(2)-index is roughly proportional to the cube root of the total number of citations.

The h(2) of 25 was obtained by automatic search for citations of E.Witten (the highest h-index among physicists, according to HIRSCH), and the same search produced h=112. As expected $[h(2)]^3$ is on the same order of magnitude as $h^2$. Witten's h(2) is lower than his h by a factor of 4.5, and so is the time necessary to verify the authorship of papers contributing to the corresponding indexes. Yet, the original h-index still needs much less time for the verification of the authorship than the total number of citations, and in the opinion of the present author this is the most significant advantage of the h-index over the total number of citations.

Table 1 presents the ranking of 19 chemistry professors from the one university (*vide ultra)* in terms of the following factors

**max** – the number of citations of the most cited paper. This is not necessarily the best method to assess the scientific output of an individual, but in contrast with other methods considered by HIRSCH (total number of papers, total number of citations, etc.), it can be quickly established.

**h(2)** – index. The average h(2) of an individual was 5.11 (st. dev. 1.10).

**ch(2)** – h(2) index corrected for self citations. For each frequently cited paper, the number of self-citations was subtracted from the number of citations. Then the papers were ranked according to the corrected number of citations and h(2) was established as described above. It should be emphasized that citations corrected for self-citations are not necessarily independent citations, because citations by a coauthor were not corrected for. The average h(2)-index dropped by 18 % after correction for self-citations.

**h** – index. The average h of an individual was 15.42 (st. dev. 5.39).

**ch** - h index corrected for self-citations as described above for h(2). The average h-index dropped by 26 % after correction for self-citations.

The shared ranks are expressed by averages, e.g., shared 1st and 2nd place = rank 1.5.

Table 1 Ranking of 19 individuals in terms of different criteria

| Individ. | max | h(2) | ch(2) | h | ch |
|---|---|---|---|---|---|
| A | 1 | 1.5 | 2.5 | 8.5 | 5 |
| B | 2 | 5 | 2.5 | 3.5 | 3.5 |
| C | 3 | 5 | 2.5 | 2 | 2 |
| D | 4 | 1.5 | 2.5 | 1 | 1 |
| E | 5 | 5 | 5.5 | 6 | 7.5 |
| F | 6 | 5 | 5.5 | 5 | 3.5 |
| G | 7 | 10.5 | 10.5 | 12 | 10.5 |
| H | 8 | 10.5 | 10.5 | 10 | 10.5 |
| I | 9 | 5 | 10.5 | 3.5 | 7.5 |
| J | 10 | 16 | 10.5 | 14 | 14 |
| K | 11 | 10.5 | 10.5 | 8.5 | 7.5 |
| L | 12 | 10.5 | 10.5 | 12 | 14 |
| M | 13 | 10.5 | 16.5 | 12 | 16.5 |
| N | 14.5 | 16 | 10.5 | 15 | 12 |
| O | 14.5 | 16 | 16.5 | 17 | 16.5 |
| P | 16 | 10.5 | 10.5 | 7 | 7.5 |
| Q | 18 | 16 | 16.5 | 17 | 14 |
| R | 17 | 16 | 16.5 | 17 | 18 |
| S | 19 | 19 | 19 | 19 | 19 |

Table 2 presents the correlations between the rankings based on different criteria.

Table 2 Correlation coefficients

| | max | h(2) | ch(2) | h | ch |
|---|---|---|---|---|---|
| **max** | 1.0000 | 0.8738 | 0.9145 | 0.7871 | 0.8183 |
| **h(2)** | 0.8738 | 1.0000 | 0.8643 | 0.9138 | 0.8786 |
| **ch(2)** | 0.9145 | 0.8643 | 1.0000 | 0.8497 | 0.9266 |
| **h** | 0.7871 | 0.9138 | 0.8497 | 1.0000 | 0.9289 |
| **ch** | 0.8183 | 0.8786 | 0.9266 | 0.9289 | 1.0000 |

Tables 1 and 2 indicate that different criteria produce a similar order with a few exceptions. The time-consuming correction for self-citation has induced rather insignificant changes in the rankings. Interestingly, although the contribution of self-citations to h (26 %) is substantially higher than to h(2) (18 %), the h-ch correlation coefficient is higher than the h(2)-ch(2) correlation coefficient. This result is against expectations. In other words, replacement of h by h(2) did not result in reduction of the effect of self citations on the record of the individual. For instance two individuals (I and M) clearly improved their position by frequent auto-citation in terms of h- and h(2)-index as well.

The most substantial difference between the rankings based on the h- and h(2)-index is in relatively better correlation of the later with the max-based ranking. Clearly h favors a type of "hard worker" (many papers with moderate number of citations per paper, the individual P is an example) over a type of "genius" (few papers with a high number of citations per paper, the individual A is an example). This problem was discussed by EGGHE (2006), who proposed the g-index, which gives even more credit to a "genius" than the present h(2)-index.

The popularity of the h-index may be due to the fact that "hard-workers" and more numerous than "genii".

Although the h(2)-index failed in elimination of self-citations as the means to artificially improve own record, it succeeded in reducing the number of papers in the sample of interest from 293 (h-index) to 97 (h(2)-index), and the time necessary to verify their authorship.

Certainly the h(2)-index has the same intrinsic disadvantages as the other indexes based on the number of citations that is:

- the indexes of scientists working in different fields are not comparable
- the indexes of scientists of different age are not comparable
- own record can be easily improved by self-citations or mutual citations
- own record can be easily improved by publishing review papers

Nevertheless, the h(2)-index offers an attractive alternative to the h-index and to the total number of citations as the means to assess the scientific output of a chemist.

The idea coined in the present paper can be further generalized, by defining a h(x) index as the number of papers of certain author with citation number of at least $[h(x)]^x$. The original h-index corresponds to x=1, the h(2)-index introduced in the present paper corresponds to x=2, and the total number of papers corresponds to $x=-\infty$. The original h-index is probably appropriate in the fields, where the typical number of citations per article is relatively low, e.g., in mathematics or astronomy. The h(2)-index is favored in chemistry and physics. In medicine and biology, where the typical number of citations per article is higher than in chemistry, x=2.5 may be more appropriate.

## Acknowledgement

An anonymous referee is acknowledged for helpful suggestions.

## References

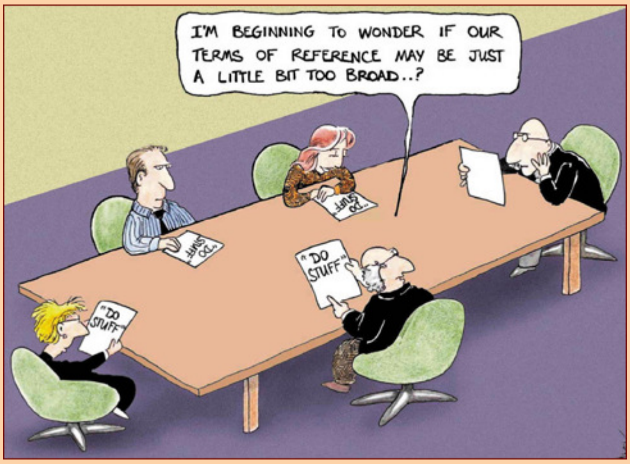
BALL, P. (2005), Index aims for fair ranking of scientists. *Nature*, 436 : 900.

BORNMANN, L. DANIEL H.D. (2005), Does the h-index for ranking of scientists really work? *Scientometrics*, 65: 391-392.

EGGHE, L. (2006), An improvement of the h-index: the g-index. *ISSI Newsletter*, 2:8-9.

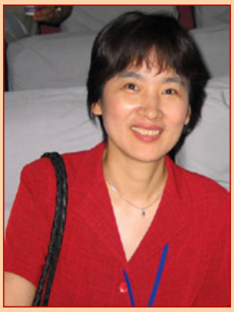
HIRSCH, J.E. (2005), An index to quantify an individual's scientific research output. *P.Natl. Acad.Sci.USA*, 102: 16569-16572.

## CARTOON



© Nick Kim (Nearing Zero). Reproduced with the permission of the author.

# A co-occurrence study of international universities and institutes leading to a new instrument for detecting partners for research collaboration



Liying Yang
National Science Library of CAS, Beijing, China
yangly@mail.las.ac.cn

Bihui Jin
National Science Library of CAS, Beijing, China
jinbh@mail.las.ac.cn

## 1. Introduction

Large universities and research institutes such as Harvard University in the USA, the Max Planck Institute in Germany, CNRS in France and the Chinese Academy of Sciences (CAS) play an important role in improving a country's innovation ability. Their publications often reveal the focus of their actual research. By analyzing co-occurrence phenomena observed in articles published by twenty leading universities and institutes, we intend to demonstrate the feasibility of a new instrument for policy makers. This instrument detects possible partners for research collaboration. When it is not important to make a distinction we will use the term 'institutes' for universities as well as institutes.

## 2. Methods

Two co-occurrence phenomena are analyzed in this article: institute - institute co-occurrence in the byline of the article, reflecting collaboration between institutes, and institute-keyword co-occurrence revealing the topic studied by members of the institute.

Institutes included in this investigation are those occupying the top-20 in chemistry according to the Essential Science Indicators (ESI; Thomson Scientific), December 2005. A list is given in the appendix. Publication data of these institutes are taken from the Web of Science covering the period 2004-2005. If an author has two addresses: one at an institute, e.g. CNRS and one at a university, e.g. Université de Strasbourg 1, then the author is considered to belong the university or institute that is first mentioned in the address field.

Data are handled as follows: first raw co-occurrence (institute-institute co-occurrence, and institute-keyword) matrices are drawn. Normalization is performed using Pearson's correlation coefficient. Then a hierarchical clustering method (between-group linkage) is applied and a dendrogram is produced. Finally, a multi-dimensional scaling (MDS) map is generated. Tools used in the study include SPSS, Derwent Analytics and Excel.

## 3. Results

### *3.1 Institute-institute co-occurrence analysis (research collaboration net)*

Fig. 1 shows the collaboration dendrogram and Fig. 2 the corresponding MDS map. In this map we may discern four quadrants. Two American ones, a European one and an Asian one. It is interesting to see that the only British institute (Cambridge University) in the TOP-20 belongs to an American group, and not to the European one. The Russian Academy of Sciences, on the other hand belongs to the European quadrant. Japanese universities are mapped very close to one another, and also the Chinese Academy of Sciences belongs to this quadrant, be it somewhat half-way between the Japanese cluster and the European one. Overall CAS is the most isolated institute on this map. There are three plausible reasons for this result: either CAS has less opportunities for collaboration (because of the special topics its scientists are interested in), or CAS does collaborate but with many dispersed partners, or CAS simply does not take advantages of existing opportunities. We will further show that the third reason is probably closed to the truth. It is also remarkable that the European institutes (CNRS, CNR, Max Planck Institute) cluster together.

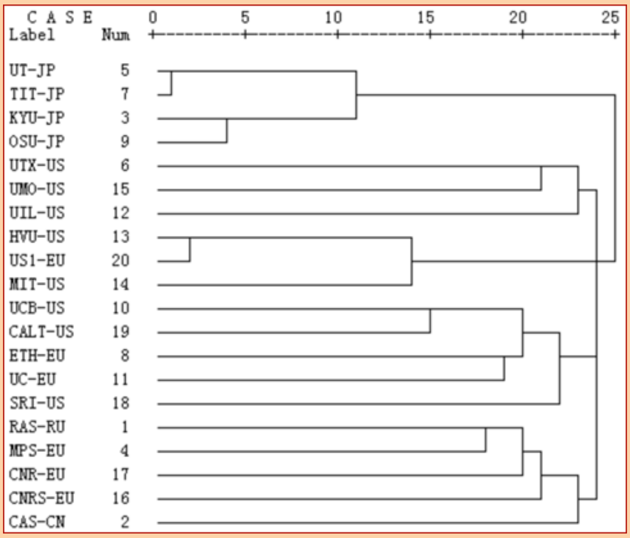


*Figure 1* Dendrogram for institute co-occurrence data

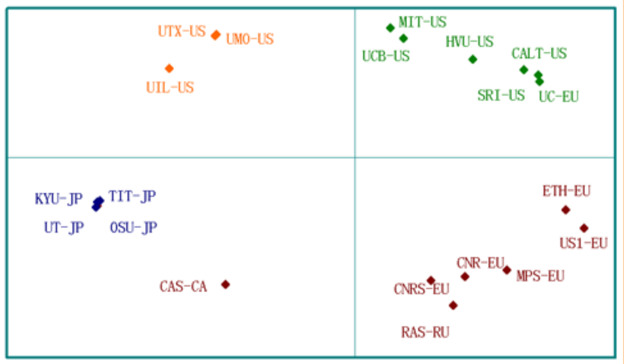


*Figure 2* Institutional collaboration: MDS map

*3.2 Results of institute-keyword co-occurrence analysis*

Figures 3 and 4 illustrate the results of this analysis. Based on the institute-keyword co-occurrence matrix, these twenty institutes form a central core with some satellites. The chemical topics studied by the Scripps Institute (mainly in the category of Biochemistry & Molecular Biology) and Caltech make these institutes outliers on this map. Also Harvard University (HVU-US) does not really belong to the core and the same observation holds for the Université de Strasbourg 1 (US1-EU). CAS is situated near Osaka University and the Max Planck Institute.

*3.3 A comparison between the two co-occurrence maps*

When institutes collaborate on a regular basis they automatically study the same topics, resulting in similar positions in the institute-keyword map. The Max Planck Institute in Germany and CNR in Italy being a good example. It is the other situation which is interesting. When institutes are situated near each other in the second map (Fig.4) and not in the first (Fig.2) then it is clear that there is an opportunity for collaboration as they are interested in similar topics. Such institutes may be called 'latent partners'. For instance, CAS and the Max Planck Institute are latent partners in this sense.

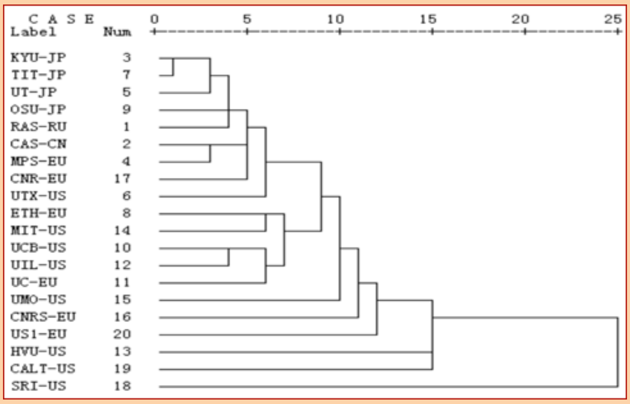


*Figure 3* Dendrogram of institute - keyword co-occurrence

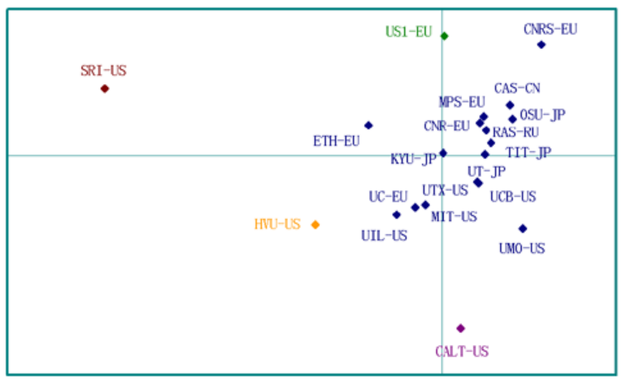


*Figure 4* Institute - Keyword MDS Map

## 4. Concluding Observations

Combining collaboration networks and institute-keyword mappings (perhaps also keyword-keyword mappings) seems to be a very promising idea for detecting latent partners. This is: institutes that are interested in similar topics but do not collaborate yet.

The level of big institutes and a large field such as chemistry do not allow for precise collaboration proposals. In future investigations we intend to look at smaller units and subfields. In this way we intend to develop a useful instrument for research policy makers. We are convinced though that this article already proves the feasibility of this approach.

## Acknowledgements

The authors thank Prof. Ronald Rousseau for valuable advice and editorial help. They also acknowledge financial help from the Ministry of Science and Technology of China project 2004CCC00400.

## Literature

1. Small, H.. Co-citation in the Scientific Literature:A New Measure of the Relationship Between Two Documents. Journal of the American Society for information Science, 1973,24(4):28-31
2. Small H. Structural Dynamics of Scientific Literature. International Classification.1976, 3(2):67-74
3. Small H. A co-citation model of scientific specialty:a longitudinal study of collagen research. Social studies of science, 1977,7,139-166
4. Small H.. Clustering the science citation index using co-citations. Scientometrics, 1985,7(3)391-409
5. Small H. Tracking and Predicting Growth Areas in Science. Proceedings of ISSI 2005, vol(1):13-23
6. Glänzel W, Czerwon, H J. A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. Scientometrics; 37 (2) Oct 96, p.195-221
7. Sharada, B A; Sharma, J S. A study of bibliographic coupling in linguistic research. Annals of Library Science and Documentation; 40 (4) Dec 93, p.125-37
8. Callon, M; Courtial, J P; Laville, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics; 22 (1) Sept 91, 155-205
9. Law, J; Whittaker, J. Mapping acidification research: a test of the co-word method. Scientometrics; 23 (3) Mar-Apr 92, 417-461.
10. Van Raan, A F J, Tijssen, R J W. The neural net of neural network research: an exercise in bibliometric mapping. Scientometrics; 26 (1) Jan 93, p.169-92
11. Law, J; Courtial, J P; Bauin, S; Whittaker, J. Policy and the mapping of scientific change: a co-word analysis of research into environmental acidification. Scientometrics; 14 (3-4) Sept 88, 251-264.

## Appendix

List of 20 institutes and universities used in this study. Abbreviations and full names.

1. UCB-US. University of Berkeley, California, USA.
2. MIT-US. Massachusetts Institute of Technology, Cambridge, MA, USA.
3. HVU-US. Harvard University, Cambridge, MA, USA.
4. CIT-US. California Institute of Technology (CALTECH), Pasadena, CA, USA.
5. UIL-US. University of Illinois, IL, USA
6. UTX-US. University of Texas, TX, USA.
7. UMO-US. University of Minnesota, MO, USA.
8. SRI-US. Scripps Research Institute, California and Florida campuses, USA.
9. KYU-JP. Kyoto University, Japan.
10. UT-JP. University of Tokyo, Japan.
11. OSU-JP. Osaka University, Japan.
12. TIT-JP. Tokyo Institute of Technology, Japan.
13. MPS-EU. Max Planck institute, Germany.
14. CNRS-EU. Centre national de la recherche scientifique (National centre of scientific research), France.
15. UC-EU. Cambridge University, England, UK.
16. ETH-EU. Eidgenössische Technische Hochschule, Zürich (Swiss Federal Institute of Technology), Zurich, Switzerland.
17. CNR-EU. Consiglio Nazionale delle Ricerche (National Research Council), Italy.
18. US1-EU. Université de Strasbourg 1, France.
19. RAS-RU. Russian Academy of Sciences, Russia.
20. CAS-CN. Chinese Academy of Sciences, P.R. China.

ISSI Newsletter is published by ISSI (http://www.issi-society.info/).
Contributors to the newsletter should contact the editorial board **by email**.
Wolfgang Glänzel: wolfgang.glanzel@econ.kuleuven.be
Ronald Rousseau: ronald.rousseau@khbo.be
Liwen Vaughan: lvaughan@uwo.ca
Aparna Basu: basu.aparna@rediffmail.com
Balázs Schlemmer: balazs.schlemmer@econ.kuleuven.be
Accepted contributions are **moderated** by the board. **Guidelines** for contributors can be found at http://www.issi-society.info/editorial.html