JISSI The International Journal of Scientometrics and Informetrics

Volume 2 Number 1 March 1996

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

Business Correspondence to :
Brzark Information Systems (P) Ltd
112 Humayun Pur
(Near NCC Office)
Safdarjung Enclave
New Delhi 110 029, India
Fax/Phone : +91 11 688 2366

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

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

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

Publishers :
Prabhat Pandey
Bina Pandey

Cover design :
Bibhas Dutta

Price : Single Copy
US $ 55.00 / Rs. 700.00

Introductory Note

Is there an Information Imperialism and What Informetrics can do about it!

The other day I was browsing through a book on Australian comics (Comics in Australia and New Zealand : The collections, the collectors, the creators. *Ed.* by Toby Burrows & Grant Stone. New York: The Haworth Press, 1994). The articles (and also a few illustrations provided) show the overall impact and influence of the North American comics not only on the readers but also on the artists and publishers of comic strips. It is said that US citizens are identified with three C's: Comics, Colas and Cookies. The first two items have invaded almost every corner of the world. Just a few months back there was a comic strip story of the Phantom (adapted in Bengali) in a Bengali news daily which involved a US cattle trader, his daughter (to become wife of the Phantom), the Phantom and an African tribal prince in the 18th century set up. The US trader was an importer of cattle from Africa to US, and not a slave trader. It just appeared to be a tricky way of altering the image of the white men of the 18th century America for better. This apart, the point I want to make is this. Look at any comic strip — Superman, the Phantom, Mandrake, the magician — everywhere the image and culture of the Yankee are projected to be the only acceptable great thing! As God created man in his image, the comic- creators of North America are creating every possible cultural trait and super humans and saviors in the image of the Yankee.

Take Superman — the child lands in the US, not in Africa, Asia, Australia or Latin America — he is white, having caucasian features, though he comes from a distant planet. Magnon is a super-emperor or the monarch who again looks like an American in all features (and the comic-writer cannot think of a better system of administration, say the *Panchayati* of some ancient Indian states or the *Senate* of the Romans — where there is not a single all-powerful soul). Lesser and evilish stellar races have features like monkeys, dragons, mice or are pink, yellow or dark!

One may take this as an extreme example of information imperialism. But again, throughout the ages it has happened in some way or other — burning of books, killing of intellectuals, promoting books and ideas of the dominating people to the domination. To this day the South East Asian cultures bear the influence of the Indian epics of the Ramayana and the Mahabharata!

At the other extreme are the Dunkel proposal or the GATT, or the NPT. These are debatable issues. But the majority of intelligent citizens in the third world countries look upon them as the means of information imperialism or technological imperialism of the West.

I always have an argument with many of my friends (in India and elsewhere) as to the universality or globality of the scientific (and technological) enterprise. Most of my friends consider sciences as a whole to be global or universal. I consider them highly contextual and regional. A small proportion is very much global and the 'basics' are universal, but most of science is not. Cosmology, astrophysics, the fundamental theoretical physics, the basic principles of chemistry, pure mathematics and researches in them are global. Yet one may probably find cultural distinctions even in approaching these basic sciences. The biomedical and earth sciences are largely contextual.

And lastly, the bibliographic tools, and among them the most omniscient one of citation indexing are also under attack for biases. But we can probably keep such global data bases as INSPEC, CAS, BIOSIS, MATHSCI etc out of any criticism of being biased. The soul and coverage of available citation indexes are always being debated. Maričic in a *Letter to the Editor* (Three Faces of Citation-Janus) in this issue of JISSI reasserts his proposal that ISI® should only include core journals in its data bases and not the journals from the periphery. Otherwise there would not be any scientific enterprise worth its name in the less developed countries in the near future. The monopolistic or imperialistic pervasion effects no growth for indigenous local comics, local features or local science and local technology — only import, follow or perish! Can not a thousand flowers bloom? Any country's S&T endeavour is sure to be an islandish affair to a large extent.

Can there not be a few empirical studies taking all the journals published from a country and producing citation index and impact factors and ranked list on such a sample set? What would be the result – the same pattern as that of JCR® or the journals from that particular country forming the core group? There are also not sufficient data to know the research emphasis or topical interests in different subject fields in different countries.

Informetrics may probably show by developing different data files and indicators the extent of islandish affairs and national priorities and interests in science, technology, social sciences and other fields on one hand, and the extent of information imperilism through bibliometric and other informetric methods by studying the proportion of the number of indigenous comics and foreign comics, TV programmes, news paper column lengths and the like.There may be many approaches and indicators which may be developed by the informetricians.This is a virgin field of research needing immediate attention.Extent of information imperialism may also be understood by knowing the research topics a third world scientist is motivated to take up to maintain survival as a scientist. There are examples when good but unfashionable (to the West) science done by a third world scientist could not be easily publihed.

This is both a sensitive and debatable issue but needs immediate attention. We invite all our readers and any scholar interested in this issue to send comments, observations and research findings to us so that they may be published together in a special issue. The deadline for submissions is 31 August 1996.

I acknowledge the benefits of discussions on several occasions with Mr S Arunachalam who, however believes in universality and globality of S&T activities.

Subir K Sen

A Keyword Analysis of Joint Biotechnology Projects

Pépin Cabo and Margreet Boersma
University of Groningen, Faculty of Management and Organization
P.O. Box 800, 9700 A V Groningen, The Netherlands
Fax : 31503633850

This paper shows how keywords can be used for the study of technological innovation in biotechnology at two levels of aggregation : projects and program (as collections of projects). Keyword analysis, based on word patterns and co-word structures, is applied to records in which participants of R&D projects describe their activities. The analysis of projects that fall under the ECLAIR programme of the European Union shows how the innovation activity of public and private sector organizations can be compared, how degrees of similarity can be computed, and how differences can be a qualified.

Keywords : Keyword analysis; Joint R&D; Biotechnology; European Union

1. Introduction

Biotechnology is one of the areas in which technological cooperation between firms is most prevalent. It is also an area in which universities and research institutes play an important role. In fact, science and technology are highly interwoven in biotechnology [6][8]. Our main focus of attention in this paper is on the methodology of technology studies : how can we characterize the content of the technology in R&D projects? How can the activities of participants be described? A method that answers these questions is useful because we can then deal with research questions such as : to what extent are there differences between the activities of different types of organizations? Are these differences specific for particular types of R&D projects? etc. These questions seem relevant for high-tech fields in general because of the *strategic* nature of R&D in these fields. This means that results from basic science have immediate commercial effects, and that science based firms are a significant part of the research community.

The empirical basis for this paper are projects from the ECLAIR programme (the European Collaborative Linkage of Agriculture and Industry through Research). The size of this programme, 42 projects, permits an illustration of the detailed analysis that we advocate, but, at the same time, there are a relative large number of joint projects of firms and research organizations. Furthermore, the public databases from the European Commission provide excellent documentation about R&D activities.

The type of innovation in the projects is determined by analysing the keywords that each participant uses to characterize their activities in the project. One advantage of using keywords is that they reflect the content of the research, or *problem field*, as perceived by researchers. To make typologies of innovations is always difficult: typologies are either too general or the social scientist has not enough expert knowledge about the type of R&D. The use of keywords removes these problems. Another advantage of keywords is that they are standardized, as opposed to title words or abstract words. If researchers have to choose from a fixed pre-determined list, it becomes easy to compare their patterns of keywords.

First, we combine all keyword files for firms, research institutes, and universities. Our goal is to analyze the differences between these three organizations with respect to three classes of keywords : result types, subject descriptors, and market applications. Since the projects are international, we also check for differences between countries. Secondly, we take a more detailed look at five joint projects in order to examine, by using correspondence analysis, the *problem structure* of the cooperation between firms, research institutes and universities.

2. ECLAIR

The European Collaborative Linkage of Agriculture and Industry through Research (ECLAIR) programme was launched under the second Framework programme, subactivity 4.2 : *Agro-industrial technologies* for agricultural and industrial advancement through technological development, using biotechnology and other advanced technologies. This programme was aimed at improving links between agriculture and industry by taking advantage of Europe's strength in biotechnology and the life sciences. The programme can be characterized as joint R&D or as cooperative research. Under this programme 42 projects were initiated, starting in 1988 and finishing in 1993. In 1994, 100 files described the results obtained by participants. Under this programme, research institutes and/or universities should participate in a group together with one or more firms. Most of the time the research groups were initiated by one participant who contacted other participants with appropriate scientific skills. The research carried out in the ECLAIR programme is pre-competitive in nature, in that any products or processes developed in the programme will require a further two to three years development before commercialization. One of the main priorities of the programme is to encourage the transfer of technology in the areas where it will have the most practical and economic benefit. In the programme, three areas of research are distinguished :

a) Evaluation trials and production of new species or organisms : test trials at the approproiate scale and under various conditions into novel or modified species or organisms (plants, livestock and other);

b) Industrial products and services : more precise and effective inputs to agriculture, more precise and effective extraction, transformation and production processes;

c) Integrated approaches : development of systems for harvesting the whole of a crop, studies and development projects for the integrated use of new technologies.

Data about ECLAIR was obtained from ECHO (the European Commission Host Organization in Luxembourg). From the PROJECTS database we obtained information about the participants in ECLAIR projects. From the RESULTS database we downloaded information about results of the projects. The BIOREP database was used for additional information.

Our main variables, organizations and results, are operationalized in the following way :

Organizations are categorized as firms, universities, research institutes, or other.

For the *results* of R&D projects we use three items that originate directly from the database.

a) *Type of result* : this item has two discriminating categories : process or prototype, and methodology, skill, know how.

b) *Subject discriptors* : this is a set of keywords that indicate the disciplines covered by the result.

c) *Market application* : this is a set of keywords that indicate the commercial and industrial sectors in which the result may have commercial potential.

Together, type of results, subjects, and markets define what we call type of innovation.

Each project has a file that contains the names of the participating organizations. Each project has also a number of result records. Each result record contains the name of the organization that is most responsible for the result. The project and result files can be linked by a common code.

In the 42 project descriptions firms were identified 114 times, research institutes 136 times, and universities 76 times[1]. This means that a typical project consists of three firms, three research institutes and two universities. Actually, most

[1] 11 participants are classified as "other", 14 participants are undefinable. This makes a total of 351 participants. However, in this total it's possible that participants are double counted. A mid-term assessment of the ECLAIR programme [4] gives a total of 336 participants.

projects (24) contain at least one organization from each category. Seven projects contain only firms and research institutes and seven projects contain only firms and universities. There are two projects with only research institutes and universities and two projects with only firms[2]. In 57% of the projects a firm is the main participant (research institute 24%, university 19%).

Most projects have several result files and there can be different types of results from one project. Surprisingly, while research institutes are the most common organization, they have the smallest number of results (26%). Most results are assigned to Industry (45%).

3. ResultType,Subject Descriptors and Market Applications

Table 1 shows that the three types of organizations have about the same types of results. A relative comparison between industry and research intitutes shows that institutes have significantly more methods results. The other comparisons are not significantly different.

If research institutes and universities were to be combined in one *research* category, the chi-square would be 4.16 with p = 0.041.

Table 2 shows the main interest on each type of organization. Next to agriculture, in which all three types of organizations are interested, it seems that most firms are focusing on Biochemistry (28%) and Microbiology (28%) for research institutes the sector Bioengineering seems important, university's main concern lies within Genetics. Regarding the different patterns of subject descriptors, there's not a significant correlation between the three participants[3].

In addition to the disciplinary background, an overview can be given for the main commercial and industrial sectors in which the result may have commercial potential. For Industry, the Farming sector is prominent; Animal, plant breeding is an important sector for Industry as well as for

Table 1. Result type per organization (number of records)

Prime contractor	Process, Prototype	Methodology, skill, know-how	Total
Industry	12	33	45
Research Institute	2	24	26
University	4	25	29
Total	18	82	100

Chi - sq = 4.509, p = 0.105

[2] These last two combinations can appear, because of the total of undefinable participants. Except for one project, the category "other" is always joined with the cooperation between firm, university and research institute together.

[3] Correlations :

	Industry	Research Institute	University
Industry	1.00	0.4613	0.5566
Research Institute	0.4613	1.00	0.2946
University	0.5566	0.2946	1.00

Table 2. Subject descriptors per type of organization (percentages)

Subject descriptor	Industry 45=100%	Research institute 26=100%	University 29=100%
Agriculture	**0.53**	**0.54**	**0.45**
Biochemistry	**0.28**	0.25	0.31
Bioengineering	0.24	**0.46**	0.21
Biotechnology	**0.44**	0.25	**0.41**
Botany	0.13	0.21	0.28
Cell biology	0.18	0.25	0.28
Food science	0.16	**0.29**	0.17
Genetics	0.11	**0.29**	**0.41**
Microbiology	**0.28**	0.13	0.24

Table 3. Market applications per type of organization (percentages)

Market application	Industry 45=100%	Research institute 26=100%	University 29=100%
Animal, plant breeding	**0.27**	**0.46**	0.31
Biotechnology	**0.6**	**0.67**	**0.69**
Farming	**0.36**	0.17	0.14
Fementation	**0.27**	0.04	0.17
Genetic, protein engineering	0.16	0.29	**0.34**
Horticulture	0.11	**0.33**	**0.48**

Research Institutes. For Universities, the main sector is Horticulture.(Table3)

4. National Differences

In this section we want to test whether differences between the results of projects exist not only at the level of organizations but also at the level of nations. First, we will present some general data about this international dimension.

Eleven countries participate in ECLAIR. Table 4 shows how the total number of organizations, projects and results are distributed among countries. It can be seen that France has the highest share in all categories.

The number of countries per project ranges from two to seven, with an average of four. Both

Table 4. Involvement in ECLAIR by country (percentages)

Country	Number of organizations 334=100%	Number of projects 42=100%	Number of results 100=100%
Belgium	0.09	0.36	0.14
Denmark	0.04	0.19	0.04
France	0.28	0.67	0.27
Germany	0.06	0.24	0.07
Greece	0.04	0.19	0.01
Ireland	0.05	0.26	0.09
Italy	0.12	0.36	0.05
Netherlands	0.06	0.26	0.07
Portugal	0.04	0.26	0.01
Spain	0.11	0.5	0.05
United Kingdom	0.11	0.55	0.2

Table 5. Type of results for selected countries

Country	Process, Prototype	Methodology, skill, know-how	Total
France	5	22	27
U.K.	5	15	20
Total	10	37	47

Chi - sq = 0.288, p = 0.591

absolutely and relatively, most (40%) joint projects are between France and Spain. Cooperation between France and Belgium, France and Italy, Italy and Spain, and Denmark and the Netherlands is also relatively frequent.

Since France and the United Kingdom have a large number of result records we will make a comparison between these two countries.(Table 4).

Although the UK have relatively more process/prototype results than France there is no significant difference between the two countries.(Table 5).

As we did with the organizations, we determined the most frequent subject descriptors (Table 6) and market applications (Table 7) for France and the UK. As a rule, a keyword is included if it appears at least twice in one country. *Biotechnology* and *agriculture* are considered redundant words and are not included.

Chemical engineering, *food science*, and *chemicals* are the typical French words. *Polymers*, *horticulture*, and *pollution* are typical British words. The correlation between the patterns is 0.56 and not significant. This indicates that France and the UK have results on different subjects.(Table 6)

Again, the two countries have highly dissimilar patterns of market applications (correlation is 0.078 and not significant). France specializes in *food products (animal, human)*, *fermentation*, and

Table 6. Subject descriptors for selected countries (percentages)

Subject descriptor	France (27=100%)	United Kingdom (20=100%)
Aquaculture	0.07	0.15
Biochemistry	0.22	0.3
Bioengineering	0.3	0.35
Botany	0.11	0.2
Brewing	0.07	0
Cell biology	0.15	0.15
Chemical emgineering	0.22	0.05
Chemicals	0.15	0.05
Food science	0.33	0.2
Genetics	0.22	0.25
Horticulture	0.11	0.25
Immunology	0.11	0.05
Microbiology	0.3	0.25
Nutrition	0.07	0
Pathology	0.07	0
Pharmacy, Pharmacology	0.07	0.05
Physiology	0.15	0.1
Pollution	0	0.1
Polymers	0	0.20

nutritional products. British markets are : *agricultural chemicals*, *genetic, protein engineering*, and *organic agriculture*.(Table 7)

5. Discussion

Although the data do not allow us to compare the innovations of all countries, the comparison of France and the United Kingdom indicate that specialization is not only a matter of differences between organizations but also of differences between nations.

When we examined the type of innovations we found that research institutes have the smallest number of result records. So, although these organizations participate a lot, their role could be characterized as *supportive*. An answer to the question whether the nature of industrial R&D is similar to public sector R&D depends on the specific dimension of innovation. We defined innovation as a combination of subject, market, and type. The last dimension, which differentiates between process/prototype and knowledge/skill/know-how, is not very strongly affected by the organizational dimensions. First of all, almost all results are in the methodology category. This is in line with the objectives of ECLAIR, and more general EC policy, which favour precompetitive R&D. As could be expected, firms have relatively more prototype results. This is significant if we compare firms to research institutes, but not significant if they are compared to universities. Interestingly, when we looked at projects consisting

Table 7. Market applications of selected countries

Market application	**France (27=100%)**	**U.K. (20=100%)**
Agricultural chemicals	0	**.20**
Animal, plant breeding	**0.26**	.20
Biological pest control	0.11	0.15
Farming	0.22	**.30**
Fermentation	**0.3**	0.1
Fishing industry, fish farming	0.07	0.05
Food products	**.33**	0.1
Foods, drinks	0.15	0.1
Genetics, protein engineering	0.11	**.20**
Horticulture	0.18	**.20**
Industrial chemicals	0.22	0.2
Measurement systems	0.07	0.05
Nutrinional products	0.18	0.05
Organic agriculture	0.07	0.2
Paper, paper products	0.07	0.05
Pharmaceutical products	0.11	0.05
Physiological monitoring	0.11	.10

of all three types of organizations, the number of prototype results is even smaller. This may indicate an effect of the sheer number of participants : the more participants the less likely are applied results. A reason for this could be that a large number of organizations, especially firms, causes appropriability problems.

The subject and market indicators show that there are large differences between industry, research institutes and universities, and various combinations of the actors. Universities are clearly focused on the more strategic fundamental areas : genetics, and protein engineering. These results support the idea that there is some complementary of the different types of organizations. This would be a rationale for cooperation. Of course, this does not exclude complementary within a category : two firms (for instance a large pharmaceutical firm and a small new biotechnology firm) can have complementary assets as well. An answer to this question could be obtained by comparing the innovations per project. This will be discussed in the next section.

6. Correspondence Analysis of Joint Projects

In order to analyze differences between organizations in one project we selected the projects that contain result files from at least one unversity, one research institute, and one firm. Five cases satisfied this criterion. In each case we made a matrix in which the organizations are the columns and the keywords, from the subject and market categories, are the rows. Since the result type is almost always *methodology, skill, know-how*, this category has been left out. Correspondence analysis is a method for the graphical representation of the rows and columns in one map [5][10]. This enables us to determine, at the same time, the differences between the organizations, and the keywords that account for these differences. Applied to joint R&D projects, this technique gives an impression of the task and organizational complexity of alliances [7]. We used the correspondence analysis module of UCINET IV software [1] with the default options. We also used UCINET to compute the Euclidean distances between organizations, as a relative dissimilarity measure.

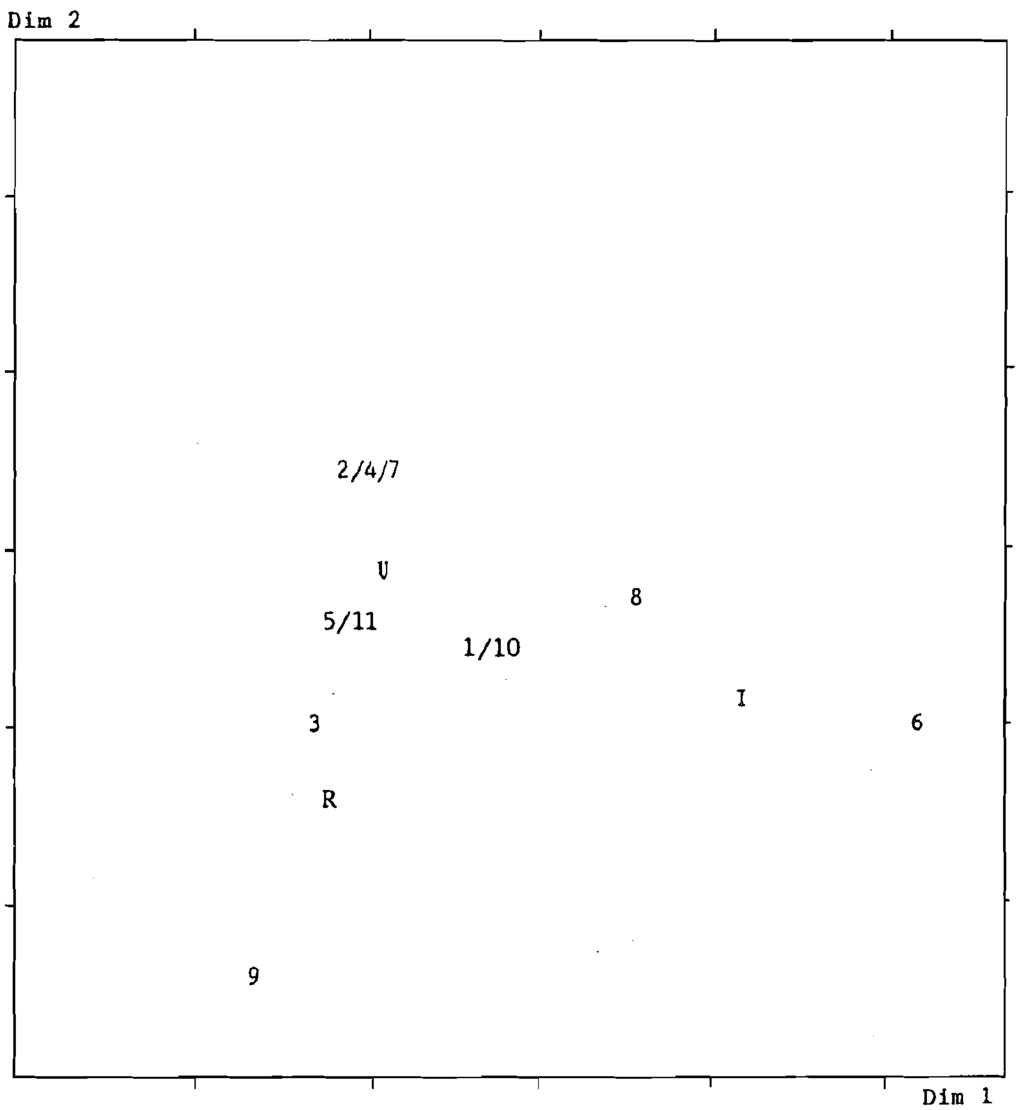
Case 1. Novel Antifungal Proteins : Applications in Crop Protection

In this project there are two result files from a university (U, the main participant) one from a firm (I) and one from a research institute (R).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM%	RATIO
1	0.634	58.1	58.1	1.385
2	0.458	41.9	100.0	
	1.092	100.0		

Scatterplot of Scores on First Two Non-Trival Factors.



Keywords.

Subjects

1. Agriculture
2. Biochemistry
3. Botany
4. Genetics

Markets

5. Animal, plant breeding
6. Biological pest control
7. Farming
8. Genetic,protein engineering
9. Measurement systems

Subjects/markets

10. Biotechnology
11. Horticulture.

DISSIMILARITIES

	U	R	I
U	0.00	1.10	1.65
R	1.10	0.00	1.90
I	1.65	1.90	0.00

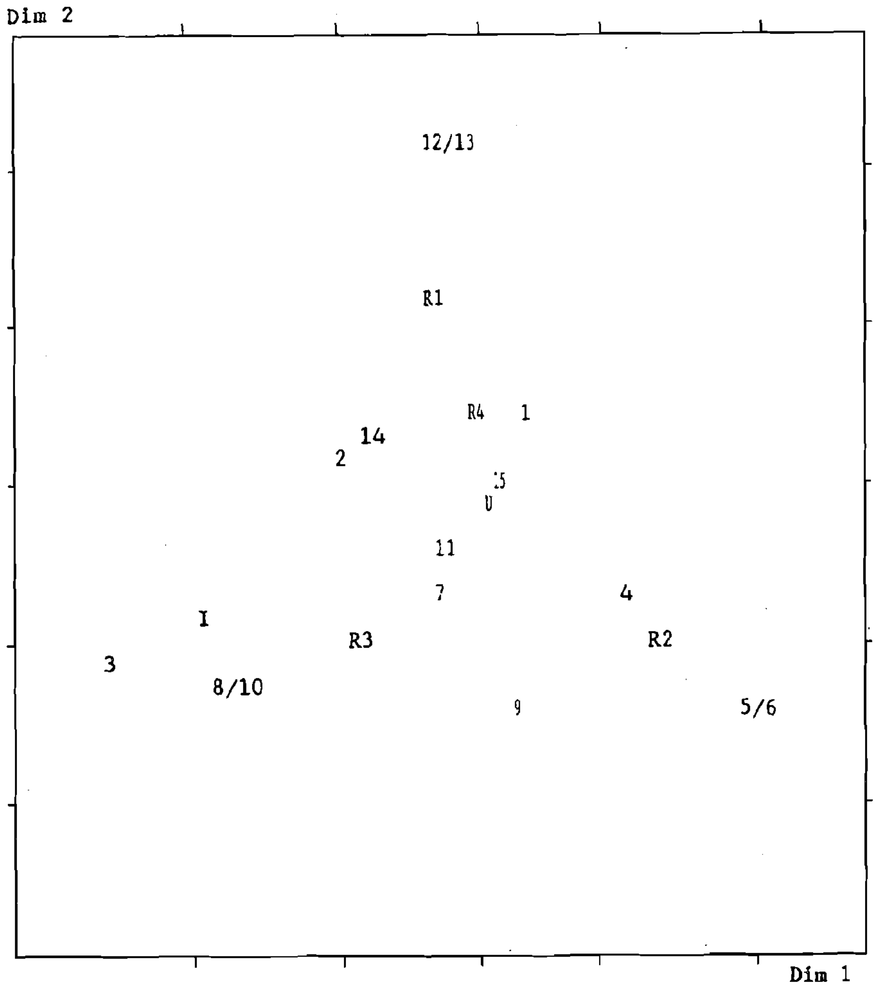
Case 2. The Development of Environmentally Safe Pest Control Systems for the European Olive

In this projects there are result files from four research institutes (R 1 to R 4), one firm (I), and one university (U) (the main participant).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM%	RATIO
1	0.704	27.6	27.6	1.128
2	0.624	24.4	52.0	1.389
3	0.449	17.6	69.6	1.085
4	0.414	16.2	85.8	1.138
5	0.364	14.2	100.0	
	2.554	100.0		

Scatterplot of Scores on First Two Non-Trival Factors.



Subjects
1. Agriculture
2. Biochemistry
3. Bioengineering
4. Biological interactions
5. Expert systems
6. Information analysis
7. Microbiology
8. Toxicology
9. Zoology

Markets

10. Agricultural chemicals
11. Biological pest control
12. Organic agriculture
13. Physiological monitoring

Subjects/Markets :

14. Biotechnology
15. Horticulture

DISSIMILARITIES

	R1	R2	R3	R4	U	I
R1	0.00	2.00	1.65	0.63	1.09	1.84
R2	2.00	0.00	1.50	1.39	0.99	2.22
R3	1.65	1.50	0.00	1.17	0.93	0.73
R4	0.63	1.39	1.17	0.00	0.45	1.57
U	1.09	0.99	0.93	0.45	0.00	1.51
I	1.84	2.22	0.73	1.57	1.51	0.00

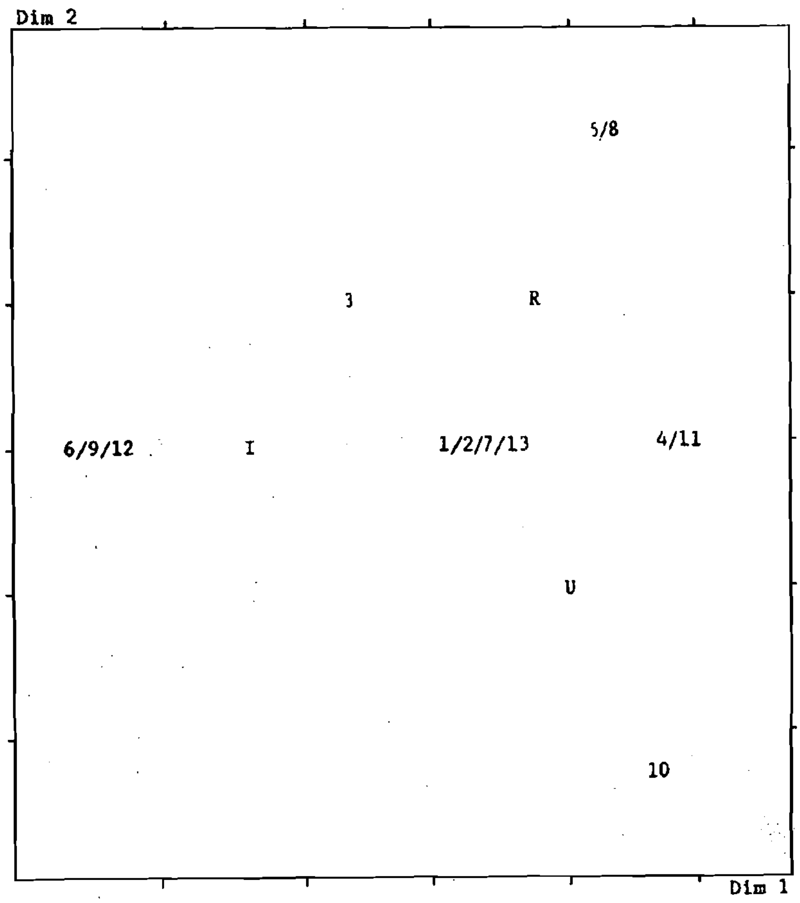
Case 3. The Development, Refinement and Commercialization of a Biotechnology Based on the in VitroProduction, Sex Determination, Freeze-Storage and Transfer of Bovine Embryos.

In this project there are result files from a firm (I, the main participant), a research institute (R), and a university (U).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM%	RATIO
1:	0.638	58.2	58.2	1.390
2:	0.459	41.8	100.0	
	1.097	100.0		

Scatterplot of Scores on First Two Non-Trival Factors



Subjects
1. Agriculture
2. Animal husbandry
3. Bioengineering
4. Cell biology
5. Physiology
6. Veterinary science

Markets
7. Animal, plant breeding
8. Farming
9. Genetic, protein engineering
10. Measurement systems
11. Physiological monitoring
12. Veterinary equipment

Subject/Market
13. Biotechnology

DISSIMILARITIES

	U	R	I
U	0.00	1.13	1.52
R	1.13	0.00	1.42
I	1.52	1.42	0.00

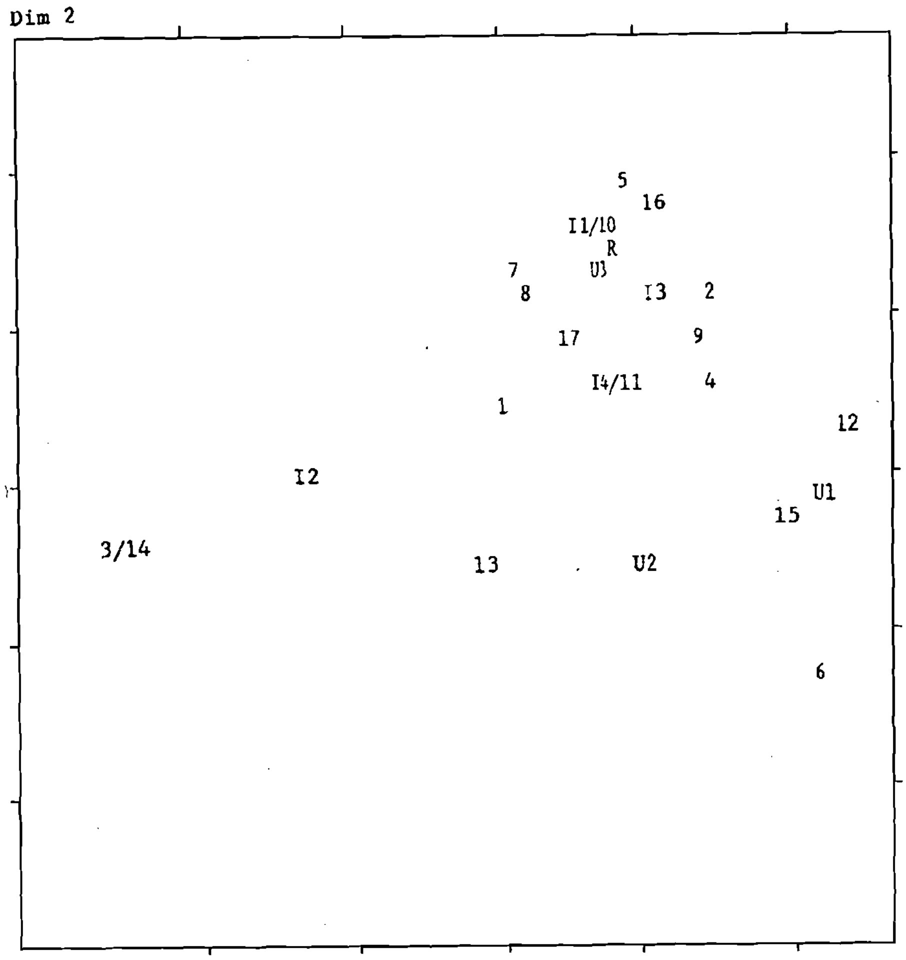
Case 4. Biological Inoculants for Seed / Plant Establishment.

This project has result files from four firms (I to 4, 4 is the main participant), three universities (U1 to U3), and one research institute (R).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM%	RATIO
1:	0.612	21.8	21.8	1.177
2:	0.520	18.5	40.2	1.108
3:	0.469	16.7	56.9	1.157
4:	0.405	14.4	71.3	1.224
5:	0.331	11.8	83.1	1.261
6:	0.262	9.3	92.4	1.228
7:	0.214	7.6	100.0	
	2.812	100.0		

Scatterplot of Scores on First Two Non-Trival Factors



Subjects
1. Agriculture
2. Biochemistry
3. Bioengineering
4. Botany
5. Cell biology
6. Genetics
7. Microbiology
8. Pathology
9. Physiology

Markets :
10. Animal, plant breeding
11. Biological pest control
12. Biotechnology
13. Farming
14. Fermentation
15. Genetic, protein engineering
16. Physiological monitoring

Subjects/markets
17. Horticulture

DISSIMILARITIES

	I1	I2	I3	I4	R	U1	U2	U3
I1	0.00	1.59	0.53	0.68	0.29	1.65	1.54	0.23
I2	1.59	0.00	1.79	1.38	1.73	2.33	1.60	1.56
I3	0.53	1.79	0.00	0.48	0.25	1.14	1.20	0.33
I4	0.68	1.38	0.48	0.00	0.56	1.13	0.86	0.47
R	0.29	1.73	0.25	0.56	0.00	1.38	1.38	0.17
U1	1.65	2.33	1.14	1.13	1.38	0.00	0.85	1.42
U2	1.54	1.60	1.20	0.86	1.38	0.85	0.00	1.32
U3	0.23	1.56	0.33	0.47	0.17	1.42	1.32	0.00

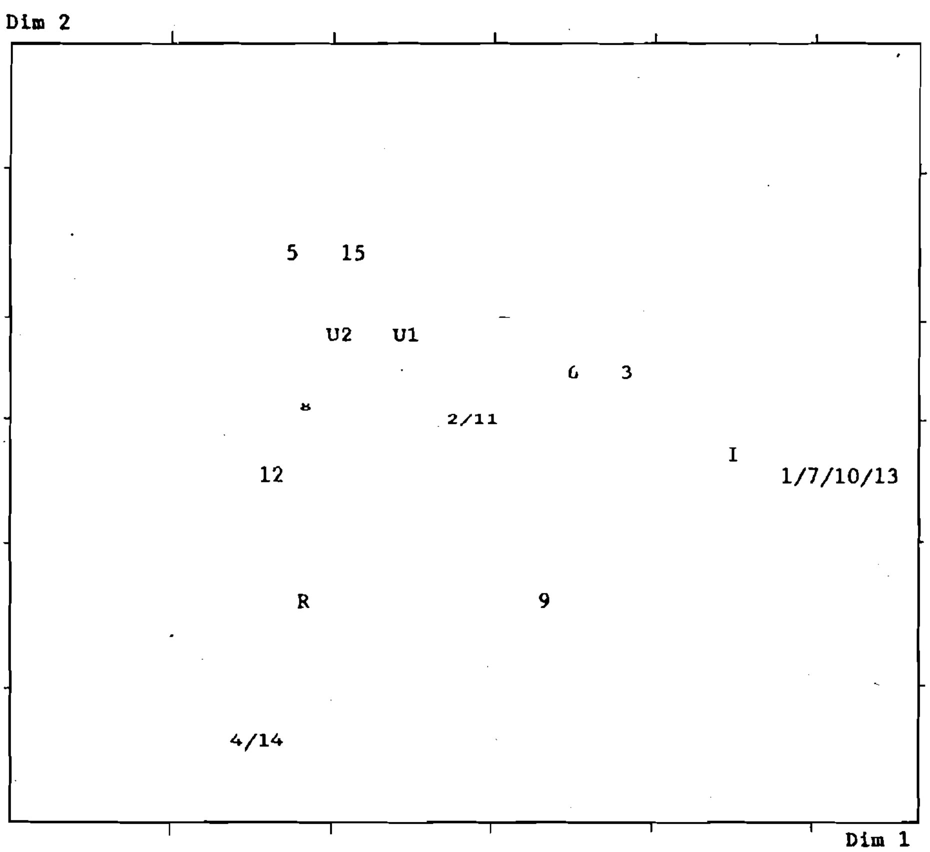
Case 5. Improvement of Yield and Feed Conversion in Salmonids and Maricultured Fish.

This project has result files from one firm (I, the main participant), one research institute (R), and two universities (U1 and U2).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM%	RATIO
1:	0.718	41.2	41.2	1.205
2:	0.596	34.2	75.3	1.386
3:	0.430	24.7	100.0	
	1.745	100.0		

Scatterplot of Scores on First Two Non-Trival Factors



Subjects:
1. Agriculture
2. Aquaculture
3. Biochemistry
4. Bioengineering
5. Biotechnology
6. Food science
7. Marine science
8. Physiology

Markets:

9. Animal, plant breeding
10. Farming
11. Fishing industry, fish farming
12. Food products (animal, human)
13. Foods, drinks
14. Measurement systems
15. Physiological monitoring

DISSIMILARITIES

	U1	U2	R	I
U1	0.00	0.27	1.52	1.45
U2	0.27	0.00	1.51	1.70
R	1.52	1.51	0.00	1.83
I	1.45	1.70	1.83	0.00

7. Discussion of the cases

We shall explain the results from the correspondence analysis by examining the first case. This case, like case 3, is a simple one because there are only three actors involved. This means that the overlap between organizations in terms of keywords could also be represented in a Venn diagram. In such a diagram, overlapping circles, representing organizations, would show the keywords that are common to all participants, keywords that are common to two participants, and keywords that are specific for one organization. If there are more than three actors (such as in cases 2, 4, and 5) the picture becomes too complicated and correspondence analysis will help. Also, if keywords occur more than once with one actor (this happens with the university in case 1) correspondence analysis will take account of this *weight*.

The singular values that we report with each plot indicate the amount of variation that each axis accounts for. Thus, in case 1, the horizontal axis accounts for 58%, and the vertical axis for 42% of the variation in the data. If we take a look at the plot we see that word 1 and word 10 are in the center of the space delimited by the three organizations. This means that these words occur with each of the three organizations. In this case, this is not surprising since the words (*agriculture* and *biotechnology*) are general for the programme. We can also see which words are organization specific : 2, 4, and 7 are specific for the university, 6 is specific for the firm, and 9 is specific for the research institute. The other words are 'in between' organizations. We suggest two measures for the degree of commonality or specificity of a *problem field* (the set of keywords). A problem field is *centrifugal* when there is a large number of specific words (in case 1, 36%), and a problem field is *centripetal* when there is a large number of common words (18% in case 1).

The dissimilarity scores, at the end of each case, measure the exact difference between organizations in the two-dimensional plot. So, the largest difference exists between firm and research institute (1.90), and the smallest difference is between university and research institute (1.10).

How can the cases be characterized in general terms? One of the research questions we mentioned in the introduction was about the specialization of diferent types of organizations : is public sector research different from private sector research? Cases 1, 3, and 5 would support this view. Universities and research centres are on one side of the map, firms are on the other side. However, this does not mean that there is absence of overlap between the keyword domains of the organizations (problem fields are never completely centrifugal). Also, there are important differences between research institutes and universities. Cases 2 and 4 do not support the differentiation hypothesis. Case 2 shows a clique of research institutes with a university in the centre, and a firm attached to this clique. Nevertheless, one research institute (R3) is closer to the firm than to all other organizations. In case 4, universities, firms and the research

[4] In case 4, the two-dimensional plot is problematic because it accounts for only 40% of the variation. A three dimensional plot would account for 57%. This means that the two-dimensional distances have to be interpreted with care.

institute all mix[4].

One hypothesis that came up when we plotted the data is about the role of the main participant. Does the fact that an organization is project leader has a greater effect on its position in the 'problem domain' (the set of keywords) than its organizational origin? In cases 1, 2 and 4 we find the main participant in the centre of the plot. So, being project leader means that one occupies an *average* position in the problem field, linking all other participants together. However, cases 3 and 5 do not support this view. The firms that are project leaders in these cases not have the most central position (these are occupied by R and U1 respectively). May be, the hypothesis should be that public sector organizations tend to occupy central positions in problem fields.

Two other sources of variation were considered and found not to be significant : the difference between market words and subject words, and the effect of nationality of the participants.

8. Conclusion

This paper showed how keyword analysis can be used for the study of technological innovation in biotechnology at two levels of aggregation : projects and programmes (as collections of projects). Keyword analysis, whether as word patterns [2] or co-word structures [9] is not new but its application to records in which participants of R&D projects desccribe their activities has been developed more recently [3].To our knowledge it is the first time such an analysis has been performed at the level of single R&D projects. The analysis of ECLAIR showed how the innovation activity of public and private sector organizations can be compared, how degrees of similarity can be computed, and how differences can be a qualified.

References

1. Borgatti, S.and D. Everett and L.Freeman (1992). *UCINET IV.* Columbia:Analytic Technologies.
2. Braam, R. (1991). *Mapping of science : foci of intellectual interest in scientific literature*, Ph.D. thesis, Leiden, ISBN 909003903-1.
3. Cabo, P. G. and W. Van Rossum (1994). A *Socio cognitive network analysis of research programs: the case of ESPRIT*, Research Report 94336 Graduate School/Research Institute Systems, Organisations and Management, University of Groningen
4. O'Driscoll, M. (1992). *European collaborative linkage of agriculture and industry through resarch. A mid-term assessment of the ECLAIR programme*.Commission of the European Communities, EUR 14471 EN, Luxembourg.
5. Greenacre, M.J. (1993). *Correspondence analysis in practice*. London : Academic Press.
6. Kenney, M. (1986). *Biotechnology : the industrial-university complex.*. London : Yale University Press.
7. Killing, J.P. (1988). Understanding alliances : the role of task and organizational complexity (in: P.J. Contractor, P.Lorange.*Cooperative strategies in intenational business,* Lexington Books) 55-67.
8. Orsenigo, L. (1986). *The Emergence of Biotechnology.* London : Pinter Publishers.
9. Rip, A. and J. P. Courtial (1984). Co-word maps of biotechnology : an example of cognitive Scientometrics. *Scientometrics* 6(6), 381-400.
10. Weller, S.C.and A.K.Romney (1990). *Metric scaling,correspondence analysis*. Sage Newbury Park, CA.

Some Remarks Concerning Misshelved Books

Lillian Beltaos
Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M 2J6
e-mail: lillian.beltaos@msvu.ca
Ronald Rousseau *
Katholieke Industriële Hogeschool West-Vlaanderen, Zeedijk 101, B-8400 Oostende, Belgium
and
Informatie - en Bibliotheekwetenschap, UIA, Universiteitsplein 1, B-2610 Wilrijk, Belgium
e-mail: rousseau@kihwv.be

Shelf reading is studied and practical observations concerning misplaced books are reported. Hypotheses are formulated to explain these observations. Further, an algorithm is proposed to determine the number of misplaced books on a shelf. It is observed that although this number is well-defined, the books that are actually misplaced are not. Suggestions are made for further research.

1. Shelf Reading

The ultimate usefulness of any library depends on the ability of the staff and the public to find books on the shelves with ease and assurance[1], as cited in [2]. Nowadays when *the network is the library* [3] this statement is only partially true, yet shelf reading remains an important, though time consuming, library service. Most librarians consider it of the greatest importance that stacks receive constant attention and are maintained in proper order.

Mathematical models describing the amount of time and/or money a library is willing to spend on shelf reading, were developed, among others, by Bookstein [4], Bookstein and Swanson [5] and, Cooper and Wolthausen [6]. These models try to minimize so-called *frustration* functions to obtain the best results for the least amount of input, i.e. time or money. Susan Lowenberg, circulation coordinator at California State University, Northbridge, on the other hand developed a comprehensive shelf reading programme to identify stack areas with high rates for misshelving errors and to determine the frequency with which each area should be read every year [7]. Applying the 80/20 rule she proposes to read the top twenty percent of the most heavily used areas on a monthly basis. Lowenberg further gives precise instructions on how shelf reading and reshelving out of place books must be done in order to minimize the time needed to do the task. Another well-documented proposal to organize shelf reading is given in Pedersen [8].

We further note the relation between effective browsing and shelf reading. Although browsing does not depend on perfectly ordered stacks, for browsing to be effective books may not migrate between different sections in the library (assuming, of course, that a subdivision in sections corresponds to an intelligent library classification scheme) [9].

2. Practical Observations

As part of their training, students of the Antwerp Library and Information Science Education Programme have made a number of observations concerning misshelved books in public and academic libraries[10]. These observations are based on small samples and hence are only

*To whom all correspondence should be addressed.

based on small samples and hence are only indicative. No statistical testing of these results has been done. We hope, however, that publication of these observations here, will prompt others to take a more scientific approach and use our observations as hypotheses to be formally tested.

Cooper and Wolthausen [6] found that the higher the shelf is from the floor, the more likely it is that books will be misplaced. They related this finding to the amount of use each shelf receives. This finding was corroborated by a test in the library of RR's institute (Oostende). Yet, other tests, mainly in public libraries could not confirm these results. On the contrary students found that misplacements occurred evenly over all heights, with an exception for the bottom shelf where often more misplaced books were found. We hypothesize that these divergent findings could be explained by different shelf reading and reshelving policies and practices. If shelf reading is done in a systematic way the observation and explanation from the Cooper-Wolthausen study seem logical. Yet, if shelf reading is never done or at most once a year, which is often the case in public libraries in Belgium, and library assistants relocate misplaced books observed while reshelving, it seems that two effects balance each other. On the one hand, the more books are placed at eye-level the more they are used; on the other hand, the higher the books, the more misplaced books are noticed and put back at their proper place by the library assistant. This would explain the fact that generally, misplacements follow a uniform distribution. Finally, when the bottom shelf is really located at the floor level, the physical inconvenience associated with shelf maintenance leaves misshelved books for longer periods of time. This observation would suggest that for efficient use of the library it is not recommended that the bottom shelf is on floor level. Perhaps the lowest shelf should better be used for oversized books.

3. How to Determine the Number of Misplaced Books

As a preliminary definition, we say that the number of misplaced books on a shelf is equal to the least number of moves required to bring the books back in their proper order. Here, a move must be understood in the physical sense : when a book is inserted between two other ones, the books on this shelf are pushed aside to give room to the inserted book. Such an operation will be called a deletion-insertion operation. This way of sorting differs from ranking objects in a row or from sorting data by computer. Inserting objects in a row often implies that all objects larger (or smaller) than the object to be inserted must be moved in order to make room for the new item (they do not move by themselves as books on a shelf do). When sorting data by computer either data must be moved to other places, or an elaborate system of pointers must be updated.

Assume now that a shelf of books is thoroughly mixed up. What is the number of misplaced books on this shelf? A moment's reflection will reveal that this is not an easy question. To determine this number one first counts the number of correctly placed books. Note that the notion *correctly placed books* is a relative notion. Books are correctly placed relative to one another, while (perhaps) a number of incorrectly placed books can be found in between. Robert Floyd has proved that this number is equal to the length of the longest increasing subsequence of books on the shelf. Note that books in a library form a total order with respect to the call-number order. Hence it is perfectly legitimate to use the term *increasing (sub) sequence*. Floyd's assertion can be shown as follows :

Proposition 1 ([11], p.105 & p.605) . *The number of correctly placed books on a shelf is equal to the longest increasing subsequence of books.*

Proof.

Assume that there are n books on ashelf. Any arrangement of these books is called a permutation. Let $\pi = a_1 a_2 \ldots a_n$ be a permutation of $\{1, 2, \ldots, n\}$. A deletion-insertion operation changes π to $a_1 \ldots a_{i-1} a_{i+1} \ldots a_j a_i a_{j+1} \ldots a_n$ or to $a_1 \ldots a_j a_i a_{j+1} \ldots a_{i-1} a_{i+1} \ldots a_n$, for some i and j. Let *dis* (π) be the minimum number of insertion-deletion operations which will sort π into order.

Now we note that a deletion-insertion operation essentially moves only one a_i. In a sequence of such operations, unmoved elements retain their relative order. Therefore if π can be sorted with k deletion-insertions, it has an conversely. Hence *dis* (π) = n - (*length of longest increasing subsequence of* π).

We now are able to define precisely what we mean by *the total number of misplaced books that belong to a particular shelf.* This number is defined as the number of books that should have been on the shelf (that is : those that are placed correctly, plus those that are on the shelf but are misplaced, plus those placed on the wrong shelf) minus the number of correctly placed books, minus the number of books on the shelf, but which actually should be placed on a different shelf. Books that are loaned out are not taken into account when determining the number of misplaced books on a shelf.

An example. Assume that the call-numbers of books on a shelf are given by the following ordered sequence :

A12, A22, A67, X99, A11, A34, A68, A23, A33, A49, A89

Moreover, the book A99 is placed on another shelf. Assume further that no other books are available with a call-number beginning with A. Then we consider all correctly placed subsequences such as :

A12, A67, A89

A12, A22, A67, A68, A89

A11, A23, A33, A49, A89

A12, A22, A23, A33, A49, A89

Obviously, the longest correctly ranked subsequence has length six. As there is one book that is on the wrong shelf (X99) and there is another book which should be on this shelf but is not (A99) the total number of misplaced books on this shelf is equal to :

$$(6 + 5 + 1) - 6 - 1 = 5. \qquad (1)$$

Note also that reversing the correct order such as :

$$A5, A4, A3, A2, A1 \qquad (2)$$

yields only one correctly placed book.

It is often thought that a book is either shelved accurately or it is not (see e.g. [2]). Unless some special rule is followed this is not true. Indeed although the number of misplaced books is well-defined, the actual books which are misplaced are not. Even if two books are interchanged it can not be determined whether it is the first one encountered which should be placed after the second one, or that it is the second one which should be placed in front of the first one. Consider, for example :

$$A1, A3, A2, A4 \qquad (3)$$

Is A3 or is A2 the misplaced book ?

More generally, in the sequence

$$A4, A5, A1, A2, A6, A7, A3, A8, A9 \qquad (4)$$

there are two correctly placed subsequences of length six, namely A1, A2, A6, A7, A8, A9 and A4, A5, A6, A7, A8, A9. Unless special rules are followed, it is not possible to say which books are correctly placed and which are misplaced. An obvious way out is to follow this rule : compare the items in the subsequences one by one. From the moment the subsequences differ, that subsequence is considered to be the *correct* one whch contains the smallest item (smallest in the call-number order). Note that this is only a formal rule which may well bear no relation at all with the way the shelf actually became disordered. If this rule is followed the subsequence A1, A2, A6, A7, A8, A9 is considered to be the correct one, hence the books A3, A4 and A5 are considered to be misplaced. Similarly, in case of a simple permutation of two books, the one first encountered is considered to be misplaced. Observe that also the counting rule followed by Cooper and Wolthausen[6] is ambiguous. They searched the shelf from left to right, and then each misplaced volume was mentally reshelved, counting the number of volumes to the left or right of its proper position, to determine the degree of misplacement. As we have shown above, one must consider the whole shelf and follow an extra rule to determine which books are *misplaced* and which are at their *correct place*.

Note finally that we have assumed that it is always possible to determine on which shelf a book is located (if placed at its correct place). In practice it may be that only when all books are returned to the library it is found that a book which was placed at the far right of a shelf should actually be placed on a lower shelf.

4. Conclusion and Suggestions for Further Research

We have given a short review on shelf reading and its influence on the number and location of misplaced books. From a theoretical point of view misplaced books. From a theoretical point of view we made an interesting observation that, although the number of misplaced books can be determined, the books that are actually misplaced can not. For this to be determined an additional formal rule is needed.

Suggestions for further research :

(a) Develop a model for reshelving and, based on this model, determine a practically useful shelving policy. This model should, in particular, incorporate the influence on misplacements for both cases where users or library staff do most of the shelving. In practice, the more users browse, the more they reshelve themselves (as in most — open stack — public libraries).

(b) Is it possible to develop an automatic scanner which determines quickly how many books are misplaced ? Such a device could then determine the status of shelves and the appropriate need for shelf reading.

(c) What is the influence of the classification scheme used by the library (Dewey, UDC, LC)? What happens if a rougher classification scheme is used e.g. one which only requires that books written by the same author are placed together (without caring about the exact order within this group).

Studying misplaced books in a library is, in a sense a form of pattern recognition. The authors hope that this article will encourage other scientists to look for, and explain patterns of misplaced items.

Acknowledgement

We thank Leo Egghe who made a number of good suggestions to improve an earlier draft of this article.

References

1. Flexner, J.M.(1927). *Circulation work in public libraries.* Chicago : ALA.
2. Kendrick, C.L.(1991). Performance measures of shelving accuracy. *The Journal of Academic Librarianship* **17**, 16-18.
3. Kibbey, M and Evans N.H.(1989). The network is the library. *Educom Review* **24**, 15-20.
4. Bookstein, A.(1973). Models for shelf reading. *Library Quarterly* **43**, 126-137.
5. Bookstein, A. and Swanson, D.R. (1973). A stochastic shelf-reading model. *Library Quarterly* **43**, 138-141.
6. Cooper, M.D. and Wolthausen, J. (1977). Misplacement of books on library shelves : a mathematical model. *Library Quarterly* **47**, 43-57.
7. Lowenberg, S.(1989). A comprehensive shelf reading program. *The Journal of Academic Librarianship* **15**, 24-27.
8. Pedersen, W.A.(1989). Statistical measures for shelf reading in an academic health sciences center library.*Bulletin of the Medical Library Association* 77, 219-222.
9. Morse, P.M. (1970). Search theory and browsing. *Library Quarterly* **40**, 391-408.
10. Rousseau, R.(1992). Opleiding in bibliometrisch veldwerk : werkwijze, resultaten en suggesties voor verder onderzoek. (A bibliometric field work course : methods,results and suggestions for further research). *Bibliotheek- & Archiefgids* **68**, 39-45.
11. Knuth, D.(1973). *The art of computer programming. Volume 3. Sorting and searching.* Reading : Addison-Wesley.

World Structure of Physics : A Cross-National Comparison

Sujit Bhattacharya
National Institute of Science, Technology and Development Studies
Dr. K.S. Krishnan Marg, New Delhi-110012, India

In view of the huge investments which are required in the field of Science and Technology, there is a need for cross-country collaborations for pooling resources and talents together. This is possible only if we know which countries have similar research directions and also their priorities in different fields/subfields of a discipline. This paper tries to find these patterns in the field of physics.

Fifteen major countries accounting for almost 80% share of the World Physics output among themselves are taken for analysis. The data for the study is of the year 1993 taken from INSPEC database (CD-ROM version). The research activities of each country are seen in different subfields of Physics as classified by IUPAP using Activity Index as proposed by Frame, et. al. The Science Strategy Index, a distance measure of the rectilinear type proposed by Bonitz, et. al. is used to cluster the countries using multidimensional scaling method. The map depicts the countries which have similar profile and their deviations from the World pattern. The countries having unique profile are also observed. The activity profile helps us to interpret the map better. The implications of this study are discussed.

Keywords : Bibliometric Indicators, Science Strategy Index, Activity Index, Multi-Dimensional Scaling.

1. Introduction

Different types of bibliometric indicators have been formed from time to time to measure scientific activities [1,2]. These indicators get refined as they increasingly try to provide a better direction to research foci, strength and weakness of research efforts, etc. Indicators are being increasingly rallied upon by science policy administrators to provide funding for science activities like the Science indicators reports of the US National Science Board and the official R&D Indicators Series of countries like Australia and Canada. Indicators are playing an important role in assessing the *health of science*, for strategic planning and decision making in Science & Technology.

Indicators on absolute values of the publication and citation counts are simplest to construct and interpret but they are too elementary to base decisions on them. They do not convey much information as they are confounded by the size of the countries and the size of the subject fields. Relative numbers, i.e. shares of the fields within a country's field distribution or within the world's field distribution are relatively better as a comparative analysis is possible. This concept of relative measure and the need for proper normalization has led to the construction of Activity Index. This index was proposed by Frame [3] and used by Schubert [4], et. al. and others. It provides a method of eliminating these confounding factors. AI is computed as :-

$$(AI)_{ij} = [(n_{ij}/n_{io}) / (n_{oj}/n_{oo})] * 100$$

Where

$n_{ij} \rightarrow$ is the no. of publication's of country *i* in sub-field *j*;

$n_{io} \rightarrow$ is the no. of publication's of all country in all subfields,

$n_{oj} \rightarrow$ is the no. of publication's of all countries in sub-field *j*;

n_{oo} is the total no. of publications of all countries in all subfields.

AI takes world share of a field as a kind of standard for that field. It conveys how much priority a particular country is giving to a field or a subfield of its research. Thus AI is also called Priority Index. Relating this standard to the field's share of any country, one can compare countries in each fields/subfields of science.

It is desirable to have a possibility to compare a country's science activities in all fields as a whole with corresponding activities of other countries. This desire to obtain a cumulative measure led to Frame, Narin & Carpenter [5] using factor analysis to study the national pattern of scientific publishing on the basis of data from SCI. R Todorov [6] applied factor and cluster analysis to study the distribution of physics over ten subfields for 36 countries. This approach has some serious limitations. Cluster analysis precludes the possibility of a country to belong to more than one cluster. Also using this method one can either cluster rows or columns of a data matrix but not both simultaneously. Factor analysis leads to the classification of the countries into clusters which may not be mutually exclusive. The determination of factor meanings is also not straightforward (These are detailed further in Notes 1). It is therefore desirable to design an appropriate indicator which uses the cumulative output of all fields/subfields of a discipline for clustering countries with similar research profile together.

This led Bonitz, et. al. [7] to develop a structural indicator Science Strategy Index (SSI) based on the scattering of a country's science activity. This indicator is similar to city block distance measure. Through this indicator he calculated similarity profile of countries based on their total science output : Life Sciences, Physical Sciences, Chemistry, Engineering and Mathematics.

The **Science Strategy Index** is constructed as

$$SSI_{ij} = 100 - 1/2\,(\sum_{f} | P_{if} |)\ (\%)$$

For a field f and a country i, the country share of the field, P_{if} (%) is subtracted from the same field's share for a country j, i.e. P_{jf} (%), taken as positive deviation. This is summoned over all fields.

$P_{jf} = P_{wf}$, gives the scattering of the country's science activity with respect to the world.

This indicator allows us to use the total output of a country in calculating its similarity with another country. The contribution of each field of a country is taken care of in calculating the similarity index between any pair of countries. This indicator can therefore be used for generating similarity profile of a country with another in a particular discipline which also reflects the contribution of each subfields of a discipline.

As a result of such pattern comparison, we get SSI values between zero and unity. For pattern comparison of a Country vs. World, unity means exact matching of the country pattern to the world pattern, zero indicates that there are no matching at all. Similarly for comparison between pair of countries, these value indicates exact matching or no matching of their profile. Values between zero and one indicates how much two country profiles have deviations from each other. Low value of SSI_{ij} indicates large deviation whereas similarly high value implies that the two countries have similar profiles.

In this paper, fifteen countries which account for over 80% of the World production in Physics have been chosen for study. We have used two indicators, Activity Index and Science Strategy Index to analyse their research pattern. This pattern is indicative of the World pattern/World Structure as they are accounting for major publications in Physics. Multi-Dimensional Scaling (MDS) is applied on a similarity matrix constructed using SSI taking all the ten subfields of physics together. MDS has advantage of both mathematical and of interpretability over the other two main Multivariate Data Analysis method that can furnish spatial representation, Cluster Analysis and Factor Analysis approach. (These are detailed further in Notes 1).

2. Justification of this Study

Appraisal and assessment of research priorities is of fundamental concern to national science policy. Decision-makers frequently need answers to such questions : What are the priorities of different fields and subfields of science? How do they compare with those of other countries? What are the areas which are receiving low priority in the country, but are accorded high priority elsewhere and vice versa? Can we classify the countries into typologies based on the similarity of their priority profiles? Thus it is important to identify the priorities and potential holes in the research agenda of countries. This article seeks to identify research priorities and similarity profiles of fifteen countries in physics. These countries account for the bulk of output in physics. Their efforts give direction to the development of theories and applications in physics. The underlying hypothesis behind this is that the output of publications in different fields or subfields or Science by a country is not a random event. It indicates more researchers, more resources, more institutions and more facilities in a country in that field. Also from methodological reasons it is easier to analyse and interpret the results. Which countries are doing research in similar directions and which are unique in their research effort? This depiction can be a pointer to linkages between different countries working in similar thrust areas.

Along with other type of non-bibliometric indicators and constraints of each country, this type of map can at best supplement the process of R&D funding and cross-country cooperation in different areas.

3. Methodology and Data

In this paper we have analysed activity in physics over ten subfields as classified by IUPAP (International Union of Pure and Applie Physics). This classificationn system is depicted in **Table 1.** The following fifteen countries were identified as major contributors in physics : USA, JAPAN, GERMANY, USSR, UK, FRANCE, CHINA, ITALY, INDIA, CANADA, SPAIN, NETHERLANDS, SWITZERLAND, AUSTRALIA, POLAND. These countries account for the bulk of output in physics (80% in 1993). Hence they are designated as *major countries*. The data for this study is taken from 1993 INSPEC database (CD-ROM version). INSPEC provides an indepth coverage for physics literature in comparison to SCI or other database.

We have used the Activity Index (AI) to measure the relative impact of these countries in these subfields. SSI is then used to provide a cross-national comparison taking all the subfields together. We have constructed a matrix based on SSI_{ij} for comparing the structure of the country i with the structure of the country j, each SSI_{ij} value provides a measure of the degree of the similarity of the two countries. We have departed from Bonitz's method in displaying the configuration (See notes 2).Using MDS, each country is projected as a point in space. Each SSI_{ij} is a distance measure in 10 dimensional space. Close proximity for country points in this space depicts the structural similarity of countries in are Physics.
The AI helps us to interpret the spatial map better i.e. the clusters and differentiated profile.

4. General Overview of the Data

In all tables and figures, the Countries are identified by their ISO codes and subfields by their abbreviations given in **Table 1(a).**

Table 1(a). Names of Countries With Their ISO Triliteral Codes

Country	Code
WORLD	WLD
USA	USA
JAPAN	JPN
GERMANY	BRD
USSR	SUN
UK	UKD
FRANCE	FRA
CHINA	PRC
ITALY	ITA
INDIA	IND
CANADA	CAN
SPAIN	ESP
NETHERLANDS	NDL
SWITZERLAND	CHE
AUSTRALIA	AUS
POLAND	POL

Table 1(b) Presents the classification system for Physics (first - level headings)

Table 2(a) indicates the output and world share of publications in different subfields.

Three large subfields condensed Matter I, Condensed Matter II and Crossdisciplinary physics account for about one half of the total output (50.6%).

Table 1(b). Classification System for Physics (First Level Heading)

Name of the Subfield	Abbreviations used
General	GEN
The Physics of Elementary particles and fields	ELEM
Nuclear physics	NUCL
Atomic and Molecular Physics	ATOM
Classical areas of phenomenology	CLASS
Fluids, Plasmas and Electric Discharges	FLU
Condensed Matter : Structure, Mechanical and Thermal Propertis	CON 1
Condensed Matter : Electronic Structure, Electrical, Magnetic and Optical Properties	CON 2
Cross-disciplinary physics and Related Areas of Science and Technology	CROSS
Geophysics, Astronomy and Astrophysics	GEO/ASTR

Table 2(a). Publication Output and World Share of Ten Subfields of Physics

Subfields	No.of Publication	World Share(%)
GEN	24820	11.28
ELEM	5460	3.85
NUCL	14345	6.52
ATOM	10370	4.71
FLU	5201	2.36
CON1	36549	16.62
CON2	36478	16.58
CROSS	38194	17.36
GEO/ASTR	18574	8.44
CLASS	26947	12.25

Table 2(b) indicates the output and world share of the countries, ranked according to their output. The top six countries (USA, Japan, Germany, USSR, UK and France) account for 64% of the total output.

Table 2(b). Publication Output & World Share of Major Countries in all Physics Fields Coimbined

Country	No. of Publication	World Share (%)
USA	62326	28.34
JPN	24651	11.20
BRD	15965	7.26
SUN	13916	6.33
UKD	12547	5.70
FRA	11448	5.20
PRC	7671	3.49
ITA	6590	3.00
IND	6401	2.91
CAN	6188	2.81
ESP	3071	2.00
NDL	3034	1.38
CHE	2923	1.33
AUS	2830	1.29
POL	2715	1.23

Table 2(c) presents the distribution of publications of major countries in different subfields of Physics in 1993. On their own these raw counts of publications do not convey much information. For example, Germany has more publications in *Elementary Particles and Fields* than Italy. Does it mean Germany gives more priority to Italy in this field? Italy has more publications in *General Physics* than *Elementary Particles and Fields*, but what inference we can draw from these figures? Can we say Italy is more active in *General Physics* then *Elementary Particles and Fields*? Thus we compute the Activity Index to draw useful inference from this publication output data.

Table 2(c). Publication Output of Major Countries in Different Subfields of Physics

Country	Gen	Elem	Nucl	Atom	Class	Flu	Con	Mat	Cross	Geo/Ast	Total
WLD	24820	8460	14345	10370	26947	5201	36549	36478	38194	18574	21993
USA	6499	2311	4506	3208	7988	1407	9893	8229	11742	6543	62326
JPN	2109	569	1601	654	2771	583	5685	5223	5244	212	24651
BRD	1735	846	1117	1029	1418	321	3120	3090	2401	888	15965
SUN	1417	563	1068	641	2045	708	2343	2598	1569	964	13916
UKD	1563	367	630	590	1665	388	2123	1638	2376	1207	12547
FRA	1267	344	571	555	1328	243	2495	2175	1628	842	11448
PRC	859	288	396	217	1083	153	1570	1392	1240	473	7671
ITA	972	578	612	284	587	146	922	978	787	724	6590
IND	592	217	320	320	628	118	1250	1276	1090	590	6401
CAN	715	258	420	443	821	111	868	757	1091	704	6188
ESP	422	175	110	197	276	43	600	557	463	228	3071
NDL	318	81	132	175	311	53	566	574	585	239	3034
CHE	361	439	327	145	219	45	436	463	369	119	2923
AUS	390	78	75	139	431	59	433	286	563	376	2830
POL	304	97	96	153	257	31	548	753	321	155	2715

5. Priority Profiles of Major Countries

The profiles of research priorities of these of these countries are presented in **Table 3**. These tables indicate the differences in the priority accorded to different subfields by different countries. We shall adopt a systematic procedure suggested by Barre [10] for fixing bench marks for qualitative description of the relative status of a subfield within a country.

Let X denote the value of Activity Index of a country for a subfield. We use the following seven point scale for fixing the bench marks :

Scale	Value	Description
$X < 25$	1	Field of neglect
$25 < X < 55$	2	Field of very low activity
$55 < X < 85$	3	Field of low activity
$85 < X < 115$	4	Field around the mean position of the country
$115 < X < 145$	5	Field of marginal activity
$145 < X < 175$	6	Field of major activity
$X > 175$	7	Field of thrust

The subfields of relative activity in different

Table 3. Activity Profile and SSI (with respctive to the world profile) of Major Countries in Physics

Country	SSI	GEN	ELM	NUCL	ATOM	CLASS	FLU	CON	MAT	CROSS	GEO/AST
WLD	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
USA	94.77%	92.00%	96.40%	110.85%	109.17%	104.61%	95.46%	95.52%	79.61%	108.49%	124.31%
JPN	85.05%	75.81%	60.01%	99.58%	56.27%	91.75%	100.01%	138.78%	127.75%	122.50%	10.18%
BRD	90.65%	96.30%	137.76%	107.27%	136.70%	72.49%	85.03%	117.60%	116.70%	86.60%	65.86%
SUN	91.18%	90.23%	105.18%	117.67%	97.69%	119.94%	115.15%	101.32%	112.56%	64.93%	82.03%
UKD	94.04%	110.39%	76.04%	76.98%	99.73%	108.31%	130.77%	101.82%	78.71%	109.05%	113.91%
FRA	92.28%	98.07%	78.12%	76.47%	102.82%	94.68%	89.76%	131.15%	114.55%	81.89%	87.09%
PRC	92.73%	99.23%	97.60%	79.15%	60.00%	115.23%	84.34%	123.16%	109.41%	93.08%	73.01%
ITA	86.31%	130.70%	228.02%	142.39%	91.40%	72.70%	93.69%	84.19%	89.48%	68.77%	130.09%
IND	92.69%	81.95%	88.13%	76.65%	106.03%	80.08%	77.96%	117.51%	120.19%	98.06%	109.14%
CAN	92.49%	102.39%	108.39%	104.06%	151.84%	108.29%	75.86%	84.41%	73.76%	101.53%	134.72%
ESP	94.77%	121.77%	148.15%	54.92%	136.05%	73.35%	59.21%	117.57%	109.36%	86.82%	87.91%
NDL	92.66%	92.88%	69.41%	66.70%	122.33%	83.66%	73.87%	112.26%	114.07%	111.03%	93.28%
CHE	82.85%	109.44%	390.45%	171.52%	105.21%	61.15%	65.10%	89.76%	95.50%	72.69%	48.21%
AUS	86.96%	122.12%	71.65%	40.63%	104.17%	124.30%	88.16%	92.07%	60.93%	114.56%	157.32%
POL	84.36%	99.22%	92.88%	54.21%	119.52%	77.26%	48.28%	121.46%	167.22%	68.08%	67.60%

Table 4. Subfields of Relative Activity of Major Countries in Physics

Country	Thrust	Major Activity	Marginal Activity	Subfields of Relative Average	Low Activity	Very Low Activity	Neglect
USA			Geo/Astr	Gen, Elem, Atom, Nucl, Class, Flu, Con 1, Cross	Con 2		
JPN			Con 1, Con 2, Cross	Nucl, Class, Flu	Gen, Elem, Atom		Geo/Astr
BRD			Elem, Atom, Con 1, Con 2	Gen, Nucl, Flu, Cross	Class, Geo/Astr		
SUN	Flu		Nucl, Class	Gen, Elem, Atom, Con1, Con 2	Cross, Geo/Astr		
UKD			Flu	Gen, Atom, Class, Con1, Cross, Geo/Astr	Elem, Nucl, Con 2		
FRA			Con 1	Gen, Atom, Class, Flu, Con 2, Geo/Astr	Elem, Nucl, Cross		
PRC			Class, Con 1	Gen, Elem, Con 2, Cross	Nucl, Atom, Flu, Geo/Astr		
ITA	Elem		Gen, Nucl, Geo/Astr	Atom, Flu, Con 2	Class, Con 1, Cross		
IND			Con 1, Con 2	Elem, Atom, Cross, Geo/Astr	Gen, Nucl, Class, Flu		
CAN		Atom	Geo/Astr	Gen, Elem, Nucl, Class, Cross	Flu, Con 1, Con 2		
ESP		Elem	Gen, Atom, Con 1	Con 2, Cross, Geo/Astr	Class, Flu	Nucl	
NDL			Atom	Gen, Con 1, Con 2, Cross, Geo/Astr	Elem, Nucl, Class, Flu		
CHE	Elem	Nucl		Gen, Atom, Con 1, Con 2	Class, Flu, Cross	Geo/Astr	
AUS		Geo/Astr	Gen, Class	Atom, Cross, Con 1, Flu	Elem, Con 2	Nucl	
POL		Con 2	Atom, Con 1	Gen, Elem	Class, Cross, Geo/Astr	Nucl, Flu	

subfields for *major countries* are listed in **Table 4**, which shows large difference in the profiles of the countries. If all the subfields are concentrated in the middle three categories of the 7 point Scale, then the profile is considered as `homogeneous' i.e. the research effort is diffused and there are no clear cut priorities. On the other hand, if none of the subfields are in the middle three categories, then the profile is considered as `differentiated' i.e. there are clear-cut priorities/activities. From **Table 4** it is clar that only a few countries conform to these typologies of clear-cut priorities. Most of the countries have a mixed profile. Only three countries Italy (ITA), USSR (SUN) and Switzerland (CHE) have major thrust areas. Only Japan has a major area of neglect. These tables are self-explanatory and any further elaboration would be redundant.

6. Similarity Patterns of Countries

Table 3 depicts the similarity a country profile has with respect to the world profile in physics. One can derive some interesting results from this table. USA, Germany (BRD) and Spain (ESP) have similar world pattern whereas Poland (POL), Italy (ITA) and Switzerland (CHE) show large deviation from the world pattern. **Fig. 1** acts as supplementary for showing the similarity or deviations a country has with the world pattern.

Table 5 depicts the SSI_{ij} matrix of 16 major countries in Physics. This matrix shows the similarity these countries have among each other. The sixteen countries (world included) form 16*15 = 240 partnerships. It is not possible to interpret the results from this matrix. Spatial representation through multi-dimensional scaling helps us to visualize the pattern and shape of the structure inherant in this matrix.

Fig. 2 represents the two-dimensional configuration of relationship obtained using MDS scaling to the matrix of profile similarity.

Guttman's algorithm was used for obtaining the optimal configuration. The value of the stress (standardized residual sum of squares) obtained was 0.12197.

Thus we were able to depict approx. 88% of the information in two dimensional space with only 12% loss of information.

7. Discussion of the Map

The use of the Science Strtegy Index to construct a country by country matrix and using Multi-dimensional Scaling Algorithm to project this configuration allows us to interpret the distance between countries directly as a measure of dissimilarity of the research direction of a country with another.

In **Fig. 2**, we depict the two dimensional configuration of countries. It must be stressed that this map is specific for this particular set of countries - adding or deleting a single country may alter the structure significantly. This spatial representation is a fairly accurate representation of the similarity of research activity of a country in physics with the other major countries since it accounts for as much as 88% of the relational information in the data array. The axes in the space are merely mathematical constructs; a description must therefore be derived from the location of the points in the configuration on these axes by means of projection. By combining the information of both axes, one can interpret the structure of the map. The map can be interpreted using the following keys :-

i) The origin of the axes corresponds to average profile. A country closer to the origin shows that its physics activity is uniformly distributed over the seven point scale of thrust - neglect. As world profile is at the origin, it also indicates that the country near the centre has its science activity similar to it.

ii) The country's in space are related to each other. If they are near each other would mean a strong linkage. The greater the distance between them, weaker is the linkage. Thus countries which are near each other in a cluster indicates similar research efforts in physics.

The structure obtained can then be easily interpreted using the keys mentioned above. Some of the interesting and useful features are elaborated below :

Japan and Italy have diametrically opposite profiles as they are projected on two opposite side of the horizontal axis.

Australia, Italy, Japan, Switzerland, Poland are having a different research direction then the world profile.

Three clusters are identifiable from the configuration :

In one cluster, there are India, China, France and

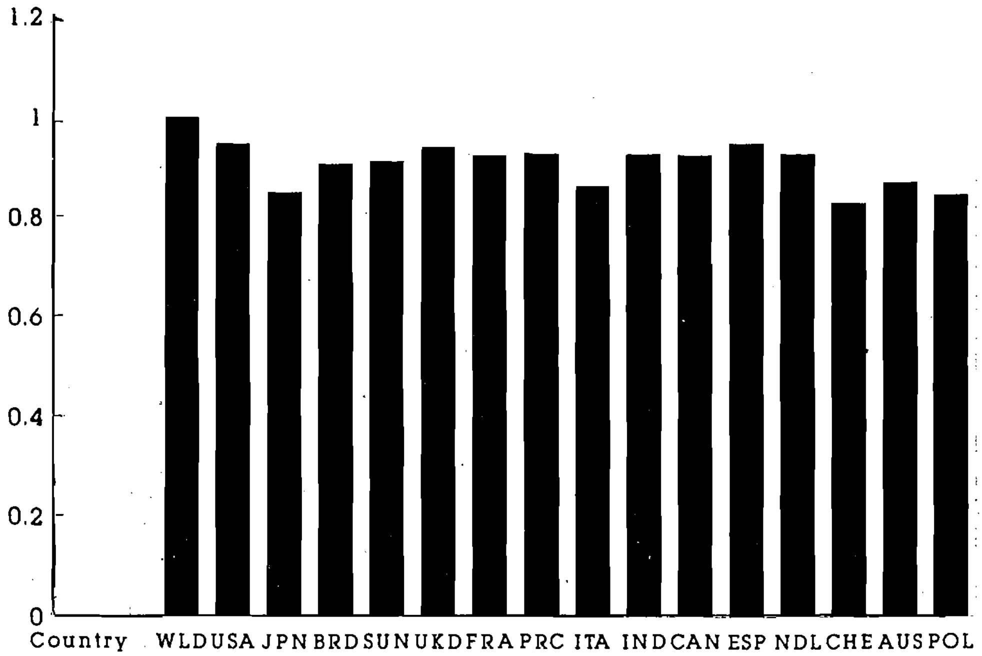


Fig. 1. Countries Ranked by SSI Values

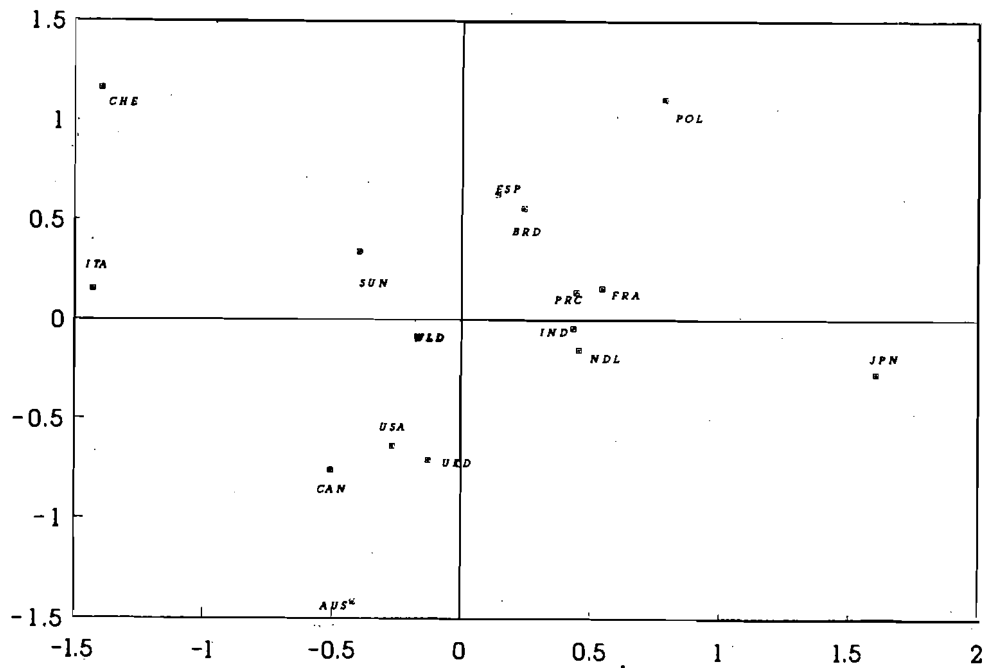


Fig. 2. Multidimensional Map of Major Countries In Physics Based On SSI Matrix

Netherlands.

In the second cluster there are Germany and Spain.

USA, Canada and United Kingdom form the third cluster.

It is clear from the map that the neighbourhood with respect to size of a country has nothing to do with respect to structural neighbourhood; for ex. USSR, Japan, USA and Germany are neighbours with respec' to publication output are scattered with respect to similarity profile in physics.

The Activity Index table constructed in **Table 4**, helps us to interpret this to an extent; the correlation although is not perfect. The major reason is that, A1 charaacterises the relative research effort a country devotes to given subfield within physics whereas SSI is the total research effort of all subfields of physics taken together in calculating its similarity with another country. The contribution of each field of a country is taken care of in calculating the similarity index. Thus the countries having similar areas of thrust/neglect should come near each other in SSI map. This underlying hypothesis stimulates us to look at AI table for understanding the clusters better. The countries which do not form any clusters and have differentiated, isolated and Unique profile - Switzerland (CHE), Poland (POL), Japan (JPN), Australia (AUS), and Italy (ITA) have there physics activity uniformly distributed in the 7 Point Scale of Thrust/Neglect.

The three identifiable clusters have somewhat similar AI profile.

But there are large number of other factors which are responsible for countries to form one cluster with geographical, historical, cultural and political being some of them [7].

8. Concluding Remarks

Research performance is essentially a multi-dimensional concept which can not be encapsuled into a single universal indicator. In this paper we have used to statistical function : Activity Index (AI) and Science Strategy Index (SSI) derived from the basic quantity, the publication output. These two indicators help us to understand the pattern of world research in physics. Another set of statistical functions can be derived from another basic bibliometric indicator, citation counts. Citation based indicators are used mainly in evaluating the performance of different countries in particular fields or subfields of a discipline. Thus we have refrained from using this indicator profile in different countries.

Fifteen major countries which account for maximum production in physics are taken for analysis. Micro level indicator, AI helps us to look at the priorities each country gives to a particular subfield/subfields. Athe macro level, SSI allows us to compare the structure of the publication output of countries. Multi-dimensional Scaling (MDS) helps us to explore and infer the underlying structure.

Since the structure of the real life data is rarely perfectly reflected in a one or two - dimensional MDS solution, certain units might be projected near each other in such a low - dimensional space which are not actually close to each other in a full (higher) dimensional space. Inspite of these limitations, the distance matrix provides an intuitively easily understandable yardstick for assessment of the degree of relationship between any pair of elements in the configuration.

AI profile of each country helps indicate to some extent why some countries have similar research pattern and why some have differentiated profile.

As pointed out by Glanzel, et. al. in their study of physics in the European Union [12] that the comparison of publication output indicators of one given subfield of different countries is correct, but the comparison between different (sub)fields may lead to flaws which make evaluation unreliable. SSI which we have used takes care o. this erroneous calculation by subtracting subfield or field with the same subfield or field of a country and then taking summation over all fields or subfields.

Thus use of two indicators at two different levels provide a deeper level of understanding of the linkage between countries and their strengths and weaknesses then by using the simple indicator of publication count to rank the countries.

This study presents a simple framework for looking at the research efforts of different countries, based on comparative analysis of bibliometric data. Tracking of imbalances in the structure of research priorities is of fundamental concern to policy makers. Identification of areas that need to be emphasized and its dual, viz. the identification of areas that need to be deemphasized are of crucial importance. Identification of countries having similar research profile can help them augument their research by having more cooperation with each other. Countries which are looking foreward to more high level interaction with another can see how far their research profile is away from the other country.

Ideally, besides bibliometric indicators, one should use other indicators for cross-national comparison e.g. distribution of scientific manpower among different fields/subfields and allocation of financial resources to different fields/subfields. But unfortunately, the data on these indiators are not available for several countries, especially at the level of subfields. On the other hand bibliometric indicators can be collected easily at different levels of breakdown of scientific fields.

Acknowledgements

We are grateful to Dr. Ashok Jain, Director, NISTADS for his encouragement and guidance. Thanks are due to Sh. P. S. Nagpaul and Dr. Prajit Basu for his comments and suggestions and to Sh. Inder Pal Singh for typing out the material.

Notes

1. These three methods have in common that the underlying statistical models allow a description of the relational structure in terms of distances based on a certain metric that is applied uniformly across all elements.Each method finds the *best* summary description of the entire (high-dimensional) network of relations between all pairs of elements given the model-imposed restrictions. Each element is represented as a point in vector space. The location of the points is meaningful in terms of structure and strength of the similarities; a small distance between two elements in general, will reflect a relatively high degree of similarity, whereas a large distance indicates a large difference. Hierarchical cluster analysis produces representations in which the degree of similarity is usually expressed as the distance between the levels of the nodes in hierarchical representation (e.g. tree like structures). An important weakness of the clustering approach per se is that this type of representation is often far too limited to capture all the relevant features of the underlying structure. A very important drawback is that hierarchical clusters do not permit multiple or cross-classifications.

   Factor analytical approaches are defined, for the sake of brevity, as a broad class of matrix approximation methods. Factor analysis and principle components analaysis are probablay two most prominent members of this class. The methods have in common that they use linear algebra (the singular value decomposition - SVD) to construct a low-dimensional space wherein the elements are placed as a set of dimensionless points (one for each element in the data set). The distance between these points is a reflection of the degree of relatedness with one another. Both factor analysis and principal components analysis approximate the matrix in terms of mutually independent dimensions, where each dimension represents a distinct factor/component in the data. In contrast to principal component analysis, the factor analysis model assumes the presence of error in the data. Both the approaches have the potential of much greater richness than cluster models.

   The M.D.S. model is based on distances between points whereas the F.A. model is based on the angles between vectors. M. D. S. has the advantage over F. A. in that it is easier to interpret distances between points than angles between vectors. Also, F. A. often results in a relatively larger number of dimensions mainly because most procedures are based on the assumption of linear relationships between the variables. This is a severe assumption with regards to perceptual data. The M. D. S. approach does not contain this assumption, and the result is that it normally provides more readily interpretable solutions of lower dimensionality. Apart from these reasons our preference for M. D. S. over F. A., there is a fundamental reason for preference concerned with type of data normally collected for analysis by each procedure. Data for M. D. S., when collected by direct judgement of dissimilarities, are least subject to experimenter contamination and most like to contain relevant structure. Data for F. A. generally contain scores for each stimulus on shopping list of attributes which may or may not be relevant. Our preference, therefore, from experimental, mathematical, and interpretative viewpoint for the Data described is for M. D. S. over F. A. or Cluster approach. Even if the experimental procedure has not promised direct proximities, it is still possible to convert the derived proximities for subsequent analsis by M. D. S.

The rapid development of M. D. S. methods which led to the present day level of computational sophistication was largely spurred on by the success of these distance model in providing useful, sufficiently robust and economical descriptions of the patterns of relationships in a host of empirical studies throughout the natural and behavioral sciences.

For interpretation and dissemination of structural properties underlying relational data, M. D. S. generated maps of science have often given a better performance compared to the other two types of M. D. A. methods, in terms of a facilitation data reduction and finding meaningful clusters.

2. Bonitz has done clustter analysis of the country-country matrix with SSI_{ij} values in (%) as matrix elements. We have used Multi-dimensional Scaling in generating a map deviating from the cluster approach used by Bonitz (See Notes 1).

Further, Bonitz has generated graphs by using SSI versus percentage of publications (the total world publication taken as 100%). We have used activity index in abscissa as it normalises the size of the field and gives a uniform scale for comparison.

References

1. Van Raan, A.F.J., Nederhof, A.J. and Moed, H.F.(1989). Science and technology indicators : Their use in science policy and their role in science studies (*Select Proceedings of the first International Workshop on Science and Technology indicators, Leiden, 1988*). DSWO Press : Leiden.
2. Carpenter, M.P., et.al. (1988). Bibliometric profile for British Academic Institutions : An experiment to develop research output indicators. *Scientometrics*, **14**, 213-233.
3. Frame, J.D. (1977). Mainstream research in Latin America and Caribbean. *Interciencia*, **2**, 143-148.
4. Schubert, A. and Braun, T.(1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, **9**, 281-291.
5. Frame, J.D., et.al. (1977). The distribution of world science. *Social Studies of Science*, 7, 501-516.
6. Todorov, R. (1985). Distribution of physics literature. *Scientometrics*, 7, 195-209.
7. Bonitz, M., et.al. (1993). The science strategy index. *Scientometrics*, **26**, 37-50.
8. Glänzel, W., Schubert, A., Braun, T. (1994). Physics in the European Union in the *80*'s - A scientometric study. *FOM-94*. 1483. **FOM**. Utrecht.
9. Nagpaul, P.S. and Pant, N. (1993). Cross-national assessment of specialization patterns in chemistry. *Scientometrics*, **27**, 215-235.
10. Barre, R. (1987). A Strategic assessment of the scientific performance of five countries. *Science and Technology Studies*, **5**, 32-38.
11. Tijssen, R.J.W. 1992). *Cartography of science : scientometric mapping with multi-dimensional scaling methods*. DSWO-Press : Leiden.

Scientometrics and Science Studies : From Words and Co-words to Information and Probabilistic Entropy

Loet Leydesdorff
Department of Science and Technology Dynamics, University of Amsterdam
Nieuwe Achtergracht 166, 1018 WV Amsterdam,
The Netherlands

The tension between qualitative theorizing and quantitative methods is pervasive in the social sciences, and poses a constant challenge to empirical research. But in science studies as an interdisciplinary specialty, there are additional reasons why a more reflexive consciousness of the differences among the relevant disciplines is necessary. How can qualitative insights from the history of ideas and the sociology of sciece be combined with the quantitative perspective? By using the example of the lexical and semantic value of word occurrences, the issue of qualitatively different meanings of the same phenomena is discussed as a methodological problem. Nine criteria for methods which are needed for the development of science studies as an integrated enterprise can then be specified. *Information calculus* is suggested as a method which can comply with these criteria.

1. Introduction

The topical relations of science studies with issues in the philosophy of science make the clarification of the relation between theory and methodology urgent for the further development of *empirical* research in this interdisciplinary field. In my opinion, the delineation of empirical science studies from older traditions of theory and philosophy of science can only be warranted if speculative reflection can be systematically supplanted by methodological exposition.

Furthermore, science studies develop in a science policy context so that analytical and programmatic questions can easily be confounded. For example, in evaluation studies the primary question is typically whether the stimulation of a research programme has led to substantive developments, and not whether the additional funding has increasingly generated measurable activities [21]. From this perspective science studies can afford to ignore philosophical questions about science and its progress only but at the price of becoming a rather trivial enterprise in itself [11]. However, one needs methods for bringing together systematically the results from more normatively oriented types of analysis of science with those from more empirically oriented ones, *yet without confounding the two*.

The methodological reflection enables us to introduce the necessary distinctions in terms of domains and research questions [13]. I shall argue in this communication that the discussion in scientometrics about words and their co-occurrences (*co-words*) has provided us with a model for relating different types of theory with respect to the data. This model for interdsciplinary science studies is then evaluated on its methodological implications.

2. From *Words* to *Information*

Word occurrences and their distributions are susceptible to statistical interpretation, and word occurrences in sentences have a semantic meaning. Both the information content of the distribution and the meaning of the words may change over time, independently and/or in relation to each other. Thus, we encounter the problem of relating different types of theory with respect to the same data in the study of words and their co-occurrences, at a very concrete level (e.g. Callon et al. [3 and 4]).

Already in 1955, Bar-Hillel [1] hinted at the possibility of a single *infomation calculus* enabling us to understand these two types of change in relation to each other. The statistical interpretation takes the occurrence of a word as an instance of this word as a nominal variable. The occurrences and the patterns in their distributions inform us in a very basic sense about the text as a system of signals defined at the word level[1]. Bar-Hillel, however, argued that the smallest unit of meaning is not the word itself but the sentence. (*Half sentences often do not have meaning.*). Therefore, he wished to look at *words in sentences*. Words can have different meanings in sentences because of their different positions, argumentative functions, etc.

When sentences are taken as units of analysis, the unit of observation can still be the word. But the relevance of an occurrence is differently evaluated, since the systems under study are different. The occurrence of a word in a sentence is no longer the instance of that word itself as a nominal variable, but the instance of another category which must be specified in terms of a theory of meaning of words in sentences. While in the former case we were interested in structure in textual data, this latter theory refers to structure in meaning among language users. Thus, the scheme by which an author like Bar-Hillel would rate word occurrences is different : two different words may be instances of the same variable in a scheme which assesses word occurrences in terms of meaning (e.g., synonyms), and the same word may be rated on different variables in two different instances (e.g., because of its position in the sentence).

In summary, researchers using different theories can be expected to generate from the same data two (or more) different relative frequency distributions. This is the crucial point for the methodological intervention : the same data, but multiple relative frequency distributions! Two analysts see something different in the same data. The data are made relevant to two or more histories[2]. However, in each case the result is eventually a relative frequency distribution : both analysts are able to ascertain whether something, the relevance of which can be specified in terms of their respective theories, is the case or not; or has occurred or not; or is to be expected or not.

Otherwise incompatible theories do not have to contradict one another in terms of their results. Theories guide us in collecting the data, and in providing us with an interpretation of the results. But the results of the measurement are comparable even if the guiding theories are mutually incommensurate [cf. 8]. The formal basis of the respective inferences offers us a common ground for a data theoretical comparison in terms of the quality of the representation.

3. What is Indicated by the Indicators?

The relative frequency distributions are the results of (sometimes implicit) theoretical assumptions concerning the subject of study. For example, a co-word mapping can be considered as a representation of a field of science. As noted, the unit of analysis that is represented in the representation may be different from the representation itself. In general, a representation provides us with a window on the represented system by using the representing system. The various systems, however, can also be theoretically specified. By paying both attention to the (methodological) quality of the representation and the relevant substances, a dual perspective on theory and methods can be developed that enables us, among other things, to bridge the gap between qualitative and quantitative science studies.

On the qualitative side of the field, only contributions which involve at least the lowest measurement scale, i.e. description in nominal categories, can be considered as part of *empirical* science studies. The (e.g., historical) description provides us with a zero-order explanaion : it indicates a possible explanation. (Note that

[1] One could also have studied the system of signals at the character level (cf. Shannon [18]), but in science studies substantive reasons have been specified for looking at the aggregation process and the dynamics of networks in terms of words and their co-occurrences (e.g., [15]; [11]).

[2] See for the concept of *multi-historicity* in relation to various reference systems : Luhmann [14], at pp. 88ff.

philosophical contributions can be relevant for empirical science studies by using this criterion). Whereas measurement in terms of categories is only a minimal requirement, more precise measurement is often possible. For example, in an aggregate one may be able to specify not only whether something was the case, but also how often it was so. The number of nominal instances can then be counted at the interval level.

On the quantitative side of science studies, the need for analysis in terms of structural units of science and in terms of various dimensions is usually recognized (see, e.g., [15]; [19]), but hardly ever are these latent units made subject to systematic theorizing. The emphasis is on the organization of data, using various methods of multi-variate analysis, and on graphic representations of the results (e.g., *mappings*). Statistical methods are often black-boxed in sophisticated computer programmes, which the researchers - in many cases legitimately - use to analyse their data. However, the choice of various parameters in statistical methods, e.g., similarity criteria and clutering algorithms, is not always discussed with reference to the theoretical questions, and thereby a vision of methods as only a kind of magical toolbox tends to be reinforced.

Methods specify the procedures by which data can be related to theories. Thus, they provide us with means to reflect on the quality of the representation. Most methods, however, have been developed for theoretical purposes; they are based on specific assumptions. For example, co-citation analysis and co-word analysis use a *relational* algorithm like single linkage clustering [3; 4; 19]. A relational algorithm, however, cannot be used for indicating the structural dimensions of a network, since this requires an analysis in terms of *positions* [2]. Thus, the relational co-word and co-citation maps should not be considered as a tool for studying the structure of science [9]. In other words, the choice of a method implies a (sometimes implicit) hypothesis with respect to the data.

The variety in clustering algorithms and similarity criteria signals the variety of reconstructions that is possible on the basis of scientometric data. Each reconstruction refers to a theoretical appreciation, and therefore remains a hypothesis. In this respect the scientometric reconstruction has a similar status as a narrative reconstruction [cf. 6]; it differs mainly in the extent to which it can be systematically tested given a theoretical perspective. But as noted, the data can be interpreted in different theoretical contexts. For example, the scientific journal system by operating produces a yearly distribution of publications (citations, etc.) over nations. What do these distributions indicate? This scientometric data can be assessed with reference to the journal system, the international division of scientific labour, and/or the development of the international science system. With reference to which system(s) do the indicators exhibit change, and which systems were (sometimes implicitly) assumed to remain stable during the period under study? Or, in terms of the example of the previous section, does a change in co-word structure reflect a change in language usage or does it reflect a change in the conceptual apparatus that is indicated by these co-words?

The observed distributions inform the reflexive analyst with reference to the *hypothesized* systems under study. The hypothesized systems appear in the reconstruction as the grouping variables, the factor designations or the cluster structure that the analyst attributes (or instructs the computer to attribute) to the data. One does not study aggregated journal- journal citations as data, but one analyses them as distributions in terms of, for example, specialties. Accordingly, one is not able to observe the specialties in the data without an analytical assumption, i.e. an expectation that the assumed existence of specialties may induce a structure in the data.

In other words : the data exhibit the mutual information or the co-variation between what is being grouped and the grouping variable, or between the represented system and the representing system[3]. For example, a body of scientific literature can be analysed in terms of words and co-words. The occurrences indicate both the documents under study, and the vocabulary used in these documents. The observable occurrences can be considered as events that inform us about a

[3] Although differently defined the mutual information and the co-variance are both measures of the uncertainty in the co-variation.

co-variation between the document set and the available vocabulary. However, information about the co-variation *plus* information about the remaining variation in each of the co-varying systems is needed for improving our prediction about the respective system's future behaviour. The remaining variation is not visible in the event, but it can be specified as an expectation by the analyst, either directly or mediated through the choice for a computer program. The observed distributions enable us to test the quality of the expectations given these specifications. Thus, the measurement informs us, if we are able to appreciate the information in theoretical terms.

4. Methodological Implications

A model has to show its strength in the prediction. The computer, however, allows us to test sets of models by using, for example, different algorithms in analysing the data. Since all the models remain informed conjectures, one may wish to develop methods which enable us to compare not only results with expectations given one model or another, but also among different representations. While the various recostructions are based on different assumptions, the data analyst may wish to uncouple as far as possible the heuristic function of theory in data collection and its appreciative function in the interpretation of empirical results, on the one hand, and the formal analysis, on the other hand. Is it possible to develop second-order methods which are content-free with reference to the reconstructions under study?

Note that the development of such methods would not only serve the comparison with hindsight. Content-free methods might have functions also in the empirical research process. A researcher often develops his/her research question when actually performing the empirical part of the research. Reflection on preliminary results easily leads to a reformulation of the research question. The parameters which were appropriately chosen for one specific research question, however, will almost never be the most appropriate choice for the next case, especially when one not only addresses a different domain, but also different variables, or when one extends the analysis to other levels of aggregation (i.e. to use different grouping rules). As noted, these parameters often refer to theoretical assumptions, and thus, the question may rise again of how to compare among results based on different assumptions.

For example, the most appropriate or at least, the currently best available technique for analysing the relations among actors, is in terms of various types of network analysis. By doing so, one is able to compute the densities of the networks, centralities, and structural equivalencies in the positions of the various actors involved [e.g., 2]. However, as soon as one addresses the question of how scientific journals relate to each other, it becomes problematic, or at least metaphorical, to think of journals as actors who cite one another, and to analyse aggregated citation networks among them in terms of the same measures of centrality, density, and structural equivalence [24]. For example, the question can be raised how to handle diagonal values in journal–journal citation matrices (are these *self-citations*?); but this difficulty indicates the underlying conceptual problem of the difference between self-citation on the part of individuals and of (articles within) journals (see e.g., [16]). These methodological issues indicate conceptual problems [12].

One strategy to follow is to use only one type of analysis for the various problems involved. Actually, many researchers have an inclination to address a new problem with methods that are familiar from previous research. However, the problems of different dimensions, different levels of aggregation, and heterogeneous units of analysis will predictably lead to a faiture of this approach in more complex instances. For example, if the researcher chooses one type of multi-variate statistics for the scientometric mapping of data-matrices composed from large document sets, he/she may thereafter wish to know whether one can decompose the total set into subsets which more pronouncedly exhibit the previously retrieved structure, and others which are more random. Which units contribute most to the structure? However, while the number of citations or the number of co-occurrences of words (co-words) in a database is a frequency, and therefore may be analysed as an interval scaled variable, at the level of each of the composing texts these same co-occurrences may be dichotomous : the citations, c.q. words, either occur or not. How can one then correlate the results of the metric and non-metric analyses? More generally : how can one compare reconstructions in which different types of analysis lead to the same results, and others in which this is not the case? The intrinsic problems of statistics may distract us from the substantive research questions.

In summary, although one is able to specify the limitations of a given method, one is usually not able to specify *a priori* whether one would like to extend the analysis on the basis of the results of a first analysis to a domain which is beyond the reach of that method. Any extension of the research may imply a shift in the level of aggregation, the measurement scale, and the relevant variables.

5. Methodological Requirements for Interdisciplinary Science Studies

The above considerations allow us reflexively to list some methodological requirements for science studies as an integrated enterprise. First, it follows from the considerations in the previous section, that methods should enable us to vary over levels of aggregation, measurement scale, and relevant variables. As noted, the specification of other relevant variables may imply the attribution of the same data to other possible units of analysis or the addition of other data with reference to the same unit of analysis. In the latter case, methods should preferably allow for the specification of the increment in the information.

In the case of another unit of analysis, methods should allow us to perform secondary analysis by using previous data collections and data analysis. Only if this latter requirement can be warranted, one is able to build on the results of the many case studies performed for other (e.g., policy) reasons. This is an urgent question for scientometrics, indeed, since data collection is often too expensive for fundamental research, while contract money is sometimes available for indicator work.

Let me more systematically summarise the methodological requirements that were derived above for integrating theoretical work in science studies with the quantative perspective provided by scientometric methods :

i) Methods should make it possible to actively import data and results (e.g., descriptions, facts, trendlines) from other types of studies. One can call this *the requirement of secondary analysis* : Data analys is should preferably support the translation among the various paradigms which are used in interdisciplinary science study.

ii) Second-order methods should allow for variation in the types of theories and methods which use the same or similar data. They should therefore be reflexive with respect to the research process, and not prescriptive in any strong sense. In particular, methods should allow for the appreciation of qualitative descriptions. This may be called *the requirement of multiple paradigms.* (Of course, if one wishes to use a particular method one should use it correctly).

In addition to these two requirements, one can also specify :

iii) *The requirement of aggregation and decomposition.* Methods should allow us to control for the relations among levels of aggregation [cf. 23].

This latter requirement, however, holds not only when we move among levels of aggregation in one dimension. In empirical science studies, the researcher may wish to import, for example, information about developments in literary structures at the field level (e.g., journal structures) into a research design which focusses on social processes at the level of research groups. The units of analysis at the different levels of aggregation are then heterogeneous. Thus, this leads to a fourth requirement which is a composite of the above requirements, i.e. :

iv) *The requirement of* ***heterogeneous nesting*** (cf. [3])

As noted, we do not *a priori* require measurement to be more precise than nominal, since we wish to allow for historical and explorative research. However, one would like to be able to use any information that can be achieved by more accurate measurement. (For example, in order to address questions like whether the described developments are *significantly* different from comparable ones). Therefore, in addition to the above specified requirements, we may specify as a fifth criterion for more integrative methods in science studies a permissive requirement with respect to the measurement technique :

v) Methods should allow for variation in the measurement scale of observations, but save any additional information from better measurement. This *requirement of neutrality in terms of the measurement scale* asks technically for a *non-parametric method.*

Actually, the use of non-parametric statistics is convenient for the import of scientometric data since the distribution of this data is often skewed. Most multi-variate statistics, however, is based on assumptions concerning normality in the distributions.[4] Therefore if one relaxes this assumption, one has additionally to specify the following requirements for the methods that one looks for :

vi) *Requirement of multi-variate statistics*, i.e., methods for science studies should allow us to develop non-parametric equivalents of clustering algorithms, etc., on datasets which we can also compose and disaggregate. Higher level results should be interpretable in terms of lower level results, and *vice versa.*

In science studies we are interested not only in these complex data structures at each moment in time, but also in their development over time. Therefore, in addition to providing us with a full equivalent of *multi-variate analysis*, methods should provide us with possibilities for studying time series of data, to make predictions, and to reconstruct. This leads to the formulation of the following two further requirements :

vii) *The requirement of dynamic analysis*, i.e., methods should allow us not only to analyse (multi-variate) data in slices at each moment in time, but systematically to account for change in the various dimensions, and in relation to overall development.

viii) *The requirement of reconstruction*, i.e. methods should enable us not only to analysedynamically and multi-variately, but also to investigate irreversibilities in the data (e.g., path-dependent transitions). Note that the formulation of this requirement is in itself neutral to the question of whether one analyses historical descriptions or scientometric data sets.

[4] Many scientometricians treat their data as if they were normally distributed. For example, in evaluation studies the mean is often used as a norm.

Finally, with respect to the data we may formulate one additional criterion which pertains to the specificity of the domain of science studies:

ix) *The requirement of virtually no systems limitations on the number of variables*, since methods should allow us to study complex phenomena and/or large communities and archives, i.e. many variables, at both aggregated and decomposed levels.

Methods for science studies should preferably not only meet one or a few of these requirements, but make it possible for the analyst to integrate results from studies in which only a subset of these requirements are needed. Therefore, methods should in principle comply with all these requirements. This means that one type of analysis is systematically relatable to another in terms of specifiable transformations.

As already conjectured by Bar-Hillel [1], Shannon's [18] information theory can be of help, indeed. The expected information content of a distribution or its probabilistic entropy is, among other things, non-parametrical, content-free, and definable in statical and dynamical measures. Theil [23] has extended Shannon's formulas to the multi-level and the multi-variate case (see also [7]). Information theory seems to provide us with a useful method for complying with the specified criteria. The recent application of this method to a large set of problems in science studies makes it increasingly plausible that the development of science studies as a specialty can become a more integrated enterprise [11].

6. Summary and Conclusions

The bridging of the gap between qualitative theorizing and the use of scientometric methods is only one among a set of requirements for the further integration of science studies [13] . On the one hand, both in qualitative and in quantitative research, the researcher has to specify categories (variables), levels of aggregation, and relevant time horizons, before any type of change can be described, or tested against the data. On the other hand, if the same data can reveal different dynamics, how are we to analyse these dynamics, both independently and in relation to one another?

By elaborating the example of the measurement of *meaning* in terms of word occurrences in the semantic tradition and the measurement of word distributions in the semiotic tradition [3], I specified how to relate the different meanings of data in science studies. A set of criteria could be derived for methods in science studies which aim at integration, despite the noted differences at the theoretical and the methodological levels. By further reflection on some methodological issues, and issues in relation to the type of data involved, additional criteria for this purpose could be specified.

The use of probabilistic entropy as an integrative measure in scientometrics refers to theories about dissipation in potentially self-organizing systems (e.g., [20]; [22]). Scientometric indicators provide us with a rich domain in terms of complex and longitudinal data for testing hypotheses concerning cultural evolution [25]. The elaboration of this issue, however, reaches beyond the scope of this communication [10].

References

1. Bar-Hillel, Y. (1955). An Examination of Information Theory. *Phil. Sci.* **22**, 86-105.
2. Burt, R.S. (1982). *Toward a Structural Theory of Action.* New York, etc.: Academic Press.
3. Callon, M., J. Law, and A. Rips, Eds. (1986). *Mapping the Dynamics of Science and Technology.* London : Macmillan. 4 Callon, M., J. P. Courtial, H. Penan. (1993). *La Scienométrie.* Paris : Presses Universitaires de France.
5. Hesse, M. (1980). *Revolutions and Reconstructions in the Philosophy of Science.* London : Harvester Press.
6. Hinton, G., J. L. McClelland, and D. E. Rumelhart, Distributed Representations. In : [27], 77-109.
7. Krippendorff, K. (1986). *Information Theory : Structural Models for Qualitative Data.* Beverly Hills, etc : Sage.
8. Kuhn, T.S. (1962/1970). *The Structure of Scientific Revolutions.* Chicago : Chicago University Press.
9. Leydesdorff , L. (1992). A Validation Study of LEXIMAPPE. *Scientometrics* **25**, 295-312.
10. Leydesdorff , L. (1994). The Evolution of Communication Systems. *Int. J. Systems Research and Information Science* **6**, 219-30.
11. Leydesdorff, L. (1995). *The Challenge of Scientometrics : the development, measurement, and self-organization of scientific communications.* Leiden : DSWO Press, Leiden University.
12. Leydesdorff, L., and O. Amsterdamska. (1990). Dimensions of Citation Analysis. *Science, Technology and Human Values* **15**, 305-335.
13. Leydesdorff, L., J. Irvine, and A. F. J. Van Raan, Eds. (1989). The Relations Between Qualitative Theory and Scientometric Methods in Science and Technology Studies. Theme Issue of *Scientometrics* **15**, 333-631.
14. Luhmann, N. (1990). *Die Wissenschaft derGesellschaft.* Frankfurt a. M. : Suhrkamp.
15. Mullins, N., W. Snizek, and K. Oehler. (1988). The Structural Analysis of a Scientific Paper. In : A.F.J. Van Raan (Ed.), *Handbook of Quantitative Studies in Science and Technology.* Amsterda:Elsevier..
16. Price, D. de Solla. (1981). The Analysis of Square Matrices of Scientometric Transactions. *Scientometrics* **3**, 55-63.
17. Rumelhart, D. E., J. L. McClelland, and the PDP Research Group (1986). *Parallel Distributed Processing* **Vol. I**. Cambridge, MA/London : MIT.
18. Shannon, C. E. (1948). A Mathematical Theory of Communication, *Bell System Technical Journal* **27**, 379-423 and 623-56.
19. Small, H., E. Sweeney, and E. Greenlee. (1985). Clustering the Science Citation Index Using Co-Citations II. Mapping Science. *Scientometrics* **8**, 321-40.
20. Smolensky, P. (1986). *Information Processing in Dynamical Systems : Foundation of Harmony Theory.* In : [27], 194-281.
21. Spiegel-Rösing, I. (1973). *Wissenschaftsentwicklung und Wissenschafts steuerung.* Frankfurt a. M.: Athenäum Verlag.
22. Swenson, R. (1989). Emergent Attractors and the Law of Maximum Entropy Production : Foundations to Theory of General Evolution. *Systems Research* **6**, 187-97.
23. Theil, H. (1972). *Statistical Decomposition Analysis.* Amsterdam-London : North Holland.
24. Todorov, R., and W. Glänzel. (1988). Journal Citation Measures : a concise review. *Journal of Information Science* **4**, 47-56.
25. Van Raan, A.F.J. (1991). Fractal Geometry of Information Space as Represented by Co-Citation Clustering. *Scientometrics* **20**, 439-49.

CENTRAL DRUG RESEARCH INSTITUTE

NATIONAL INFORMATION CENTRE
FOR
DRUGS & PHARMACEUTICALS (NICDAP)

ESTABLISHED IN OCT., 1977 UNDER THE NISSAT PLAN, DEPT. OF SCIENCE & TECHNOLOGY, GOVT OF INDIA.

- **RECOGNITIONS:**
  - WHO COLLABORATING CENTRE ON DRUG INFORMATION FOR SEARO REGION
  - BIOTECHNOLOGY INFORMATION SYSTEM - DISTRIBUTED INFORMATION SUB CENTRE OF DEPARTMENT OF BIOTECHNOLOGY
  - MARINE DATA CENTRE OF NATIONAL OCEANOGRAPHIC INFORMATION SYSTEM (NOIS), DEPARTMENT OF OCEAN DEVELOPMENT
  - NODAL AGENCY (PHARMACOLOGICAL & PHYTOCHEMICAL DATA) OF INDIAN MEDICINAL PLANTS DISTRIBUTED DATABASES NETWORK ORGANISED BY , FOUNDATION FOR REVITALISATION OF LOCAL HEALTH TRADITIONS, BANGALORE,

- **COMPUTER FACILITY :**
  - COMPUTERISED INFORMATION SERVICES (CIS) : Using
    - CD-ROM DATABASES [6 NOS]
    - ONLINE DATABASES [DIALOG, STN ETC.]
    - INHOUSE DATABASES :-
      - * NATURAL PRODUCTS; * MARINE FLORA & FAUNA ;
      - * RESEARCH PROJECTS; * LUSLIC CURRENT PERIODICALS;
      - * INDUSTRIAL LICENCES;
      - * PUBLICATIONS DATABASES
  - TRAINING / WORKSHOP FOR STAFF & OUTSIDERS

- **NICDAP PUBLICATIONS :**
  - CURRENT AWARENESS PUBLICATIONS:
    - DRUGS & PHARMACEUTICALS - R&D HIGHLIGHTS
    - DRUGS & PHARMACEUTICALS - INDUSTRY HIGHLIGHTS
    - OCEAN DRUGS ALERT [DOD SUPPORT]
  - ADHOC PUBLICATIONS: BOOKS, MONOGRAPHS, REPORTS ETC.

- **REPROGRAPHY FACILITY:**
- **MICROGRAPHICS FACILITY:**
- **TRANSLATION FACILITY :** Spanish/German/French to English
- **INTERNATIONAL TELECOMMUNICATION LINKAGES :**

ERNET, SIRNET, NICNET , RENNIC, INTERNET

N.B.

LIBRARY FACILITY :

BOOKS - 18,900 ; * PERIODICALS - 514 [426:FOR, 88: IND]; BOUND VOLUMES: 40,270;

For all your information needs please contact:

Scientist-In-Charge
National Information Centre for Drugs & Pharmaceuticals
D. L. S. Division
Central Drug Research Institute
P.O.Box 173, Lucknow - 226 001.
Gram: CENDRUG Phone: 213812 Fax: (91)-0522-223405
EMail:- SIRNET: root%cdrilk@sirnetd.ernet.in INTERNET : root @cscdri.ren.nic.in

*Amdahl's Law and Scientific Collaboration**

Leo Egghe,
LUC, Universitaire Campus, B-3590, Diepenbeek, Belgium
and
UIA, Informatie - en Bibliotheekwetenschap, Universiteitsplein 1, B-2610, Wilrijk, Belgium.
and
Ronald Rousseau
LUC, Universitaire Campus, B-3590, Diepenbeek, Belgium
and
UIA, Informatie - en Bibliotheekwetenschap, Universiteitsplein 1, B-2610, Wilrijk, Belgium.

We apply some of the ideas underlying Amdahl's law - a result which is well known in computer science - and derive results on the effectiveness of scientific collaboration. We propose a simplified model and concentrate on two aspects of scientific collaboration, namely speed and quality. We find formulae describing the gain and the relative gain in time when collaborating on a scientific project. Similarly we propose a formula expressing the gain in quality when a scientific project is done in collaboration.

1. Introduction

Collaboration in scientific research has always been one of the important topics in bibliometric research. It is defined as : a process of functional interdependence between scholars in an attempt to coordinate skills and tools (based on [1]).

Martha Harsanyi recently wrote a very interesting and comprehensive review on scholarly collaboration and multiple authorship [2]. She covered the following topics :

(a) How does collaboration affects prestige, productivity, visibility.

(b) Multiple authorship and citation analysis, e.g. how to count multiple authorship and how to determine credit (see also [3]).

(c) The relation between multiple authorship and quality, acceptance rates of publications, impact, funding, salary.

(d) The concept of 'authorship' : some authors provide no intellectual contribution at all; they only lend the prestige of their name, or provide funding or administrative services. This leads to the subject of the ethics of collaboration.

(e) The concept of name order : mathematicians, logicians, statisticians and theoretical physicists tend to favour alphabetical name order (this is why there only exist Egghe-Rousseau articles, and no Rousseau-Egghe ones); experimentalists are inclined to use name order to indicate the importance of each collaborator's contribution.

(f) Mathematical measures of collaboration, e.g. those based on Egghe's collaborative principles[4].

This article differs from all those reviewed by Harsanyi. Yet, it fits in the general framework, in the sense that it is related to the question of the relation between multiple authorship and productivity increase on the one hand; and the relation between multiple authorship and quality,

* Paper presented at the 5th International Conference of the International Society for Scientometrics and Informetrics (Rosary College, River Forest, IL, USA; June 7-10,1995).

on the other hand.

2. Amdahl's Law

Amdahl's `law' [5] in computer science is a mathematical finding against parallel computing. It states that even when the fraction of computations which have to be done serially is small, say s, then the maximum speedup obtainable from even an infinite number of parallel processors is only 1/s. Of course, reality has made it clear that the assumptions underlying this `law' are - partly - wrong and now-a-days it is a proven fact that parallel computing can enormously speed up computations, see e.g. [6], [7]. Moreover, parallel computing is a hot topic in recent investigations in information storage and retrieval, see e.g. [8], [9], [10].

In this article we will apply some of the ideas underlying Amdahl's law and derive results on the effectiveness of scientific collaboration. Note that we will restrict ourselves to a very simplified model and to only two aspects of scientific collaboration, namely speed and quality. The more work can be done in parallel the sooner a job is done, yet, the more time is spent in close collaboration (serially) the higher, probably, the quality gain. For a more comprehensive model of collaboration and productivity in research the reader is referred to [11].

3. The Time Factor

In this section we will present an elementary and simplified framework to describe the time aspect of a collaborative effort. More elaborate, hence more realistic, assumptions are discussed in the relevant literature, see e.g. [11], [12]. Moreover, our model is purely deterministic, no probabilistic aspects will enter into the picture. If they do then, e.g. queueing theory [13] or a similar technique must be applied, cf [12].

Assume that a research project can be subdivided into S steps, some of which can be tackled simultaneously by teams working in parallel, and that at all times, the project leader has at his or her disposal a - practically - unlimited number of possible collaborators. At any moment in time, it can be decided which is the optimal number of steps to be done in parallel. Yet, sometimes it is just impossible to work in parallel and then this step is done by one team (this can be a group of scientists, but may equally well be the project leader working on his/her own). Fig.1 depicts an optimal organization of a project : it begins with four teams working in parallel (8 steps), goes on serially (2 steps), goes through a stage where three teams work in parallel (3 steps) and ends serially (1 step). We will compare this way of completing the job with the case every step is done by one group as a whole (serially). Assume further that any step can be done in the same time (a standard unit of time) independent of the number of collaborators. What is then the advantage, from the point of view of duration of projects, to do parts of the work in parallel?

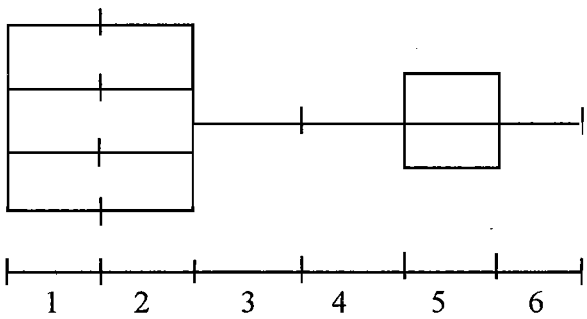


Fig.1. An optimal partition of a project in parallel and serial tasks where S(1) = 3, S(3) = 3 and S(4) = 8

Denote by S(i) the number of tasks done by i groups working in parallel (so S(i) must be a multiple of i) and by $\phi(i) = S(i)/S$ the corresponding relative number of tasks. Note that $\phi(i)$ can be zero for some i, and that $\sum \phi(i) = 1$.

3.1. Definition

The gain, denoted as G, obtained by doing some steps in parallel is defined as :

$$G = \frac{A}{C} \tag{1}$$

where C is the time it takes to complete the project in case of optimal parallel collaboration and A is the time it takes to complete this same project if done by exactly one group.

Then,

$$G = \frac{S}{\sum_{i=1}^{N} S(i)/i} = \frac{1}{\sum_{i=1}^{N} \phi(i)/i} \qquad (2)$$

where N is the maximum number of people working in parallel (at any time during the project). Equation (2) is essentially Theorem 7 of [12], i.e. the *generalized Amdahl law*. Note that we assume that there is no difference in the time it takes to finish a *serial* or a *parallel* task. Note further that we make no assumption on the order of the steps in the two different fractions. It may well be that one begins in parallel (e.g. data collection) and then continues serially, or vice versa; or that periods of serial and parallel work alternate, as shown in Fig. 1. In the siuation depicted by Fig. 1 the gain is 14/ (3 + 0 + 3/3 + 8/4) = 2.33.

3.2. Some Examples

1°) Let N be a natural number larger than one, and let $\phi(N) = 1 - \phi(1)$, while $\phi(j) = 0$ for $j \neq 1, N$. This is an arrangement in which steps are either being done N at the time (by teams working in parallel) or are done serially. In this case

$$G = \frac{1}{\phi(1) + \frac{1-\phi(1)}{N}} = \frac{N}{1 + (N-1)\phi(1)} \qquad (3)$$

In particular, if $\phi(1) = 0$, $G = N$ (the maximum gain in this situation), and if $\phi(1) = 1$, (all tasks are done serially) $G = 1$ (no gain). For $N \to \infty$, the gain in (3) becomes $1/\phi(1)$. This is the original content of Amdahl's law (cf Section 2).

2°) If $\phi(1) = \phi(2) = \ldots = \phi(N) = 1/N$, i.e. the relative number of tasks done by i teams working together, $i = 1, \ldots, N$, is equal. An example : $N = 3$, $S = 18$, $S(1) = S(2) = S(3) = 6$ (recall that $S(i)$ must be a multiple of i). Then :

$$G = \frac{1}{\sum_{i=1}^{N} 1/(Ni)} = \frac{N}{\sum_{i=1}^{N} 1/i} \approx \frac{N}{\ln(N)} \quad (N \text{ large}) \qquad (4)$$

For $N \to \infty$ the gain (4) tends to infinity.

3°) If $\alpha > 0$, and for all i, $1 \leq i \leq N$,

$$\phi(i) = \frac{1/i^{a}}{\sum_{j=1}^{N} 1/j^{a}}$$

i.e. a Lotka-like distribution, the gain G becomes :

$$G = \frac{\sum_{j=1}^{N} 1/j^{a}}{\sum_{j=1}^{N} 1/j^{a+1}} \qquad (5)$$

If $\alpha > 1$, the gain G in (5) tends to a finite value for $N \to \infty$; if $0 < \alpha \leq 1$, it tends to infinity.

3.3. Proposition

The gain G satisfies the inequality

$$1 \leq G \leq N \qquad (6)$$

Proof. On the one hand, we have

$$\frac{1}{G} = \sum_{i=1}^{N} \frac{\phi_i}{i} \leq \sum_{i=1}^{N} \phi_i = 1,$$

with equality if and only if $\phi(1) = 1$;

on the other hand, we have

$$\frac{1}{G} = \sum_{i=1}^{N} \frac{\phi_i}{i} \geq \sum_{i=1}^{N} \frac{\phi_i}{N} = \frac{1}{N},$$

with equality if and only if $\phi(N) = 1$.

4. Relative Gain

As, for fixed N, $1 \leq G \leq N$ and as it occurs that G tends to infinity for increasing N, it seems a good idea to introduce a relative (normalized) gain, denoted as RG and defined as follows.

4.1 Definition : the relative gain

$$RG = \frac{G - 1}{N - 1} \qquad (7)$$

As a function of the $\phi(i)$ the relative gain, RG, can also be calculated as follows :

$$RG = \frac{\dfrac{1}{\sum_{i=1}^{N} \phi(i)/i} - 1}{N - 1}$$

$$= \frac{1 - \sum_{i=1}^{N} \phi(i)/i}{(N-1)\left(\sum_{i=1}^{N} \phi(i)/i\right)} \quad (8)$$

As $1 \leq G \leq N$, we see that the relative gain satisfies the inequality $0 \leq RG \leq 1$.

The extreme cases $RG = 0$ and $RG = 1$ are characterized as follows :

4.2. Proposition

1°) $RG = 0$ if and only if $\phi(1) = 1$, i.e. everything is done serially.

2°) For fixed N, $RG = 1$ if and only if $\phi(N) = 1$, i.e. everything is done in parallel. (Note that in this case there is no collaboration, all groups work independently without any contact. This means that this situation should only be considered as a limiting case for more realistic ones.)

Proof.

This follows immediately from Proposition 3.3.

4.3 Examples

1°) Let N be a natural number larger than 1, and let $\phi(N) = 1 - \phi(1)$, while $\phi(j) = 0$ for $j \neq 1, N$. From Section 3, formula (3), we already known the value of G. Hence, using formula (7) we obtain:

$$RG = \frac{N - (1 + (N-1)\phi(1))}{(N-1)(1 + (N-1)\phi(1))} \quad (9)$$

In particular, if $\phi(1) = 0$, $RG = 1$, and if $\phi(1) = 1$, $RG = 0$. For $N \to \infty$, the relative gain becomes zero.

2°) If $\phi(1) = \phi(2) = \ldots = \phi(N) = 1/N$,

the relative gain is given as :

$$RG = \frac{N - \sum_{i=1}^{N} 1/i}{(N-1) \sum_{i=1}^{N} 1/i} \quad (10)$$

$$\approx \frac{N - \ln(N)}{(N-1)\ln(N)} \quad (N \text{ large}) \quad (11)$$

For $N \to \infty$ the relative gain (11) tends to zero.

3°) If for all i, $1 \leq i \leq N$,

$$\phi(i) = \frac{1/i}{\sum_{j=1}^{N} 1/j}$$

the relative gain RG becomes :

$$RG = \frac{\sum_{j=1}^{N} 1/j - \sum_{j=1}^{N} 1/j^2}{(N-1) \sum_{j=1}^{N} 1/j^2} \quad (12)$$

For $N \to \infty$, RG (12) tends to 0.

4°) We consider another special case which yields a limiting value equal to 1. Take $\phi(1) = \ldots = \phi(N-2) = 0; \phi(N-1) = \delta$ and $\phi(N) = 1 - \delta$; then, by (8),

$$RG = \frac{1 - \frac{\delta}{N-1} - \frac{1-\delta}{N}}{(N-1)\left(\frac{\delta}{N-1} + \frac{1-\delta}{N}\right)}$$

$$= \frac{(N-1)^2 - \delta}{(N-1)^2 + (N-1)\delta} < 1. \quad (13)$$

We note that in this case the limit for $N \to \infty$ is 1.

5. The All-or-Nothing Case

In this section we draw our attention to the special case that only $\phi(1)$ and $\phi(N)$ can differ from zero. This means we consider a team of scientists which either work in parallel or in series (in the latter case we will make no distinction between the case that only the project leader is activeor that all scientists work together on the same task; we will reconsider this situation when investigating the quality of the collaborative effort in the next section). Basic results are already obtained as Example 1 in Section 3 and Example 1 in Section 4. Denoting $\phi(1)$ by f (the fraction of work done serially) the formulae for G and RG are now :

$$G(f, N) = \frac{N}{1 + f(N - 1)} \quad (14)$$

and

$$RG(f, N) = \frac{1 - f}{Nf + 1 - f} \quad (15)$$

Actually, because of time spent for communication among collaborators and coordination of steps, the real G and RG are always smaller than the ideal ones we have calculated here.

5.1 *Proposition*

a) For $N > 1$ the functions $G(f,N)$ and $RG(f,N)$ are decreasing in f.

b) For $0 < f < 1$ the function $G(f,N)$ is increasing in N.

c) For $0 < f < 1$ the function $RG(f,N)$ is decreasing in N.

Proof.
That G is decreasing in f follows immediately from (14) (for $N > 1$). Similarly, it follows from (15) that RG is decreasing in N $(0 < f < 1)$. Further, rewriting G and RG as follows :

$$G = \frac{1}{f}\left(1 - \frac{1 - f}{Nf + (1 - f)}\right) \quad (16)$$

$$RG = \frac{1}{N - 1}\left(\frac{N}{1 + (N - 1)f} - 1\right) \quad (17)$$

shows that for fixed f, $0 < f < 1$, G is increasing in N, and that for fixed $N > 1$, RG is decreasing in f.

5.2 *Some special values*

Considering the case f = 0.5, we obtain the following table :

N	G
1	1
2	4/3
3	3/2
4	8/5
5	5/3
∞	2

The function G (0. 5, n) is illustrated in Fig.2

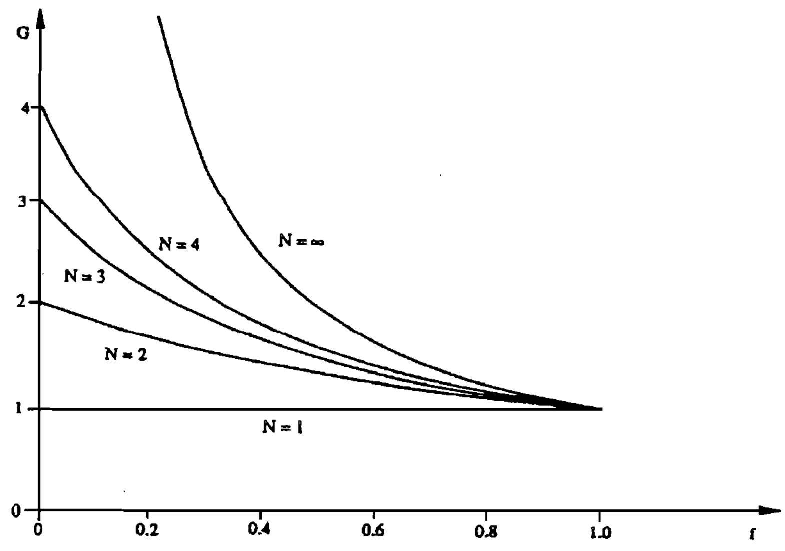


Fig. 2. The Gain Function G (0.5,N) for Different Values of N.

5.3 *Comments*

For long projects or very urgent ones, even a small gain is important (working for two years or for one year and a half on a project makes a serious difference, although the gain is here only 2/1.5 = 4/3 (c.f the case f = 0.5 and N = 2)). On the other hand, for shorter, not so urgent projects, it makes little difference whether one collaborates or not (from the point of view of the time spent to finish the work), especially when f is large and taking into account that we have calculated an ideal G!

6. Amdahl's Law and the Quality of Research

In this somewhat speculative section we will apply the same ideas as in the previous sections but with another interpretation. Here we will study the important issue of finding a formula for the gain in quality of a multi-authored paper, with respect to a single-author paper. To keep things simple we will only consider the case that team members either collaborate or not. Only when they collaborate there is a gain in quality.

Again the project is subdivided into S steps of which a fraction f is done in collaboration (serially) and a fraction $(1 - f)$ is done in parallel (no collaboration). We denote by b the average basic quality of a step if done by one scientist on his/her own. If scientists actually collaborate, e.g. by discussions, the quality of this step increases and will be given as

$$bq^{F(N)} \tag{18}$$

where $q > 1$ and $F(N)$ is a function of N which we will assume to be linear, namely $N - 1$, for small N (say, $N \leq 6$) but which will go to $-\infty$ for $N \to +\infty$. This last requirement is added for the obvious reason that, from a certain number on, more people lead to more and more inefficiency. Note that the function (18) should be considered as nothing more than a reasonable proposal. We are fully aware of the fact that other functions can be proposed with equal justification. This is the reason why we have stated that this section is somewhat speculative. Fig. 3 shows an example of a function of the form $b\, q^{F(N)}$, with $b = 1$ and $q = 1.1$. We challenge our colleague scientists to find other functions and to prove that they describe reality better than our tentative proposal.

The quality gain by collaboration, denoted by Q, is then defined as the ratio of the quality of the paper if done by N collaborators over the quality of the paper if done by one scientist. Again we assume that all scientists can do the project equally well.

$$Q = \frac{Sfbq^{F(N)} + S(1 - f)b}{Sb} = fq^{F(N)} + (1 - f) \tag{19}$$

which becomes for small N :

$$Q = fq^{N-1} + (1 - f) \tag{20}$$

Note again that this is, even within the assumptions we have used, an ideal Q. Indeed, it is not so that when it is impossible to work in parallel, co-authors actually collaborate. E.g., when one of the authors is typing the article the other authors can not discuss the typing work.

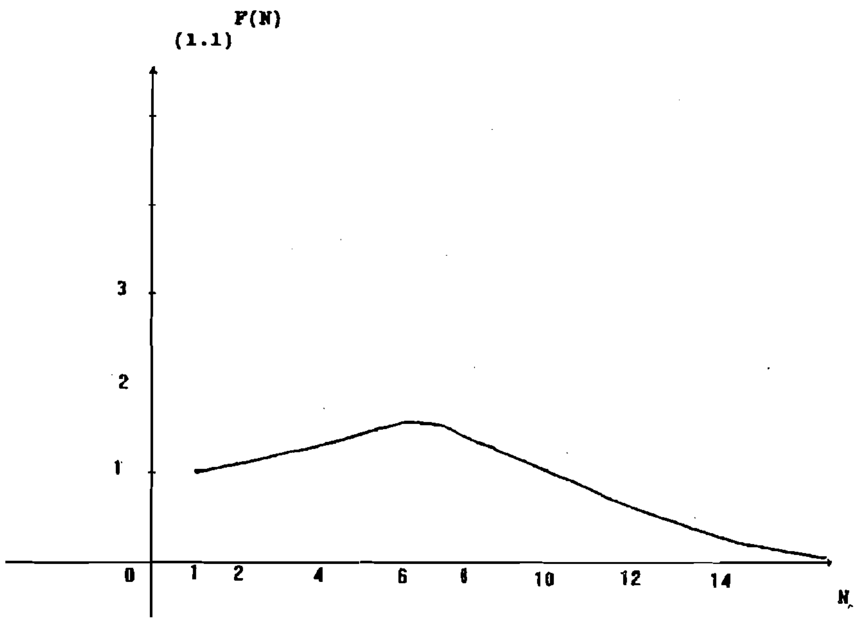


Fig. 3. A Quality Function of the Form $bq^{F(N)}$

6.1 Proposition

1°) If $f = 0$ (everything can be done in parallel) $Q = 1$ (i.e. no gain in quality).

2°) If $f = 1$ (no step can be done in parallel) $Q = bq^{F(N)}$.

3°) If $N = 1$, $Q = 1$ (as it should be)

4°) $\lim_{N \to \infty} Q(N) = (1 - f)$; hence a quality decreases when $f < 1$.

5°) For $N > 1$ the function Q is an increasing linear function of f, and, for $f > 0$, Q, as a function of N, $N < 6$, is also increasing (but not linear).

Proof.

The statement in 1°) to 4°) are obvious. We will prove only 5°). To show that for fixed $N > 1$, Q is a linear increasing function, we rewrite Q as :

$$Q = (q^{F(N)} - 1)f + 1.$$

As the slope, $q^{F(N)} - 1$, is positive this linear function is increasing.

Finally, for fixed f and small N, we have

$$\frac{\partial Q(f, N)}{\partial N} = f \ln(q)\, q^{N-1}$$

which is always positive (as $q > 1$).

6.2 Notes

The notion of *quality of a scientific article* is difficult to define and even harder to measure. It is well known that the number of citations and the quality of an article are related, yet we do not think that the relation is as simple as sometimes thought [14]. On the other hand, one could equally well as a *quality gain*, define the citation gain by collaboration as the ratio of citations garnered by a multi-authored paper to the number of citations garnered by a single-authored paper. In this situation the gain function, analogous to the function $bq^{F(N)}$, could be determined empirically [15].

7. Conclusion

We have applied some of the ideas underlying Amdahl's law, deriving results on the effectiveness of scientific collaboration. We proposed a simplified model and concentrated on two aspects of scientific collaboration, namely speed and quality. Formulae are given describing the gain and the relative gain in time when collaborating on a scientific project. Similarly we proposed a formula expressing the gain in quality when a scientific project is done in collaboration. We conclude this investigation by pointing out once again, that the gain parameters G and Q play a dual role. The more work can be done in parallel the sooner a job is done, yet, the more time is spent in close collaboration (serially) the higher the quality gain. Note finally, that we have not yet attempted a proposal to mix these two aspects into one function. Metaphorically speaking this would mean to find a compromise between *To many cooks spoil the broth* and *Many hands make light work.* We leave this problem as a suggestion for further research. We hope that this paper may encourage others to consider more sophisticated and realistic models.

Acknowledgements

We thank an anonymous referee for spotting a calculation error in an earlier version. We also thank him/her for a number of suggestions which considerably improved the presentation of this paper.

References

1. Patel, N. (1972). Collaboration in the professional growth of American sociology. *Social Science Information* **12(6)**, 77-92.
2. Harsanyi, M. (1993). Multiple authors, multiple problems - bibliometrics and the study of scholarly collaboration : a literature review.*Library and Information Science Research* **15**, 325-354.
3. Rousseau, R. (1992). Breakdown of the robustness property of Lotka's law : the case of adjusted counts for multi-authorship attribution. *Journal of the American Society for Information Science* **43**, 645-647.
4. Egghe, L. (1991). Theory of collaboration and collaboration measures. *Information Processing*

and Management **27**, 177-202.

5. Amdahl, G. M. (1967). Validity of the single processor approach to achieving large scale computing capabilities. In : *AFIPS Conference Proceedings*, **30**, (AFIPS Press, Reston (VA) 483-485.
6. Gustafson, J. L. (1988). Reevaluating Amdahl's law. *Communications of the ACM* **31**, 532-533.
7. Lauwereins,R.,M.Engels,J.Peperstraete., E.Steegmans and J.Van Ginderdeuren (1990). GRAPE : a CASE tool for digital signal parallel processing. *IEEE ASSP Magazine*, April 1990, 32-43.
8. Metzler, D. P. (1990). Connectionist and symbolic information processing : a critical analysis and suggested research agenda for connectionism from thesymbolic perspective.*Proceedings of the American Society for Information Science* **27**, 261-278.
9. Grandi, F. and P. Tiberio and P. Zezula (1992). Frame-sliced partitioned parallel signature `files (In Belkin, Ingwersen, Pejtersen, eds. SIGIR'92, *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval)*, 286-297.
10. Sejnowski, T. J. and P. S. Churchland (1992). Silicon brains. *Byte* **17(10)**, 137-146.
11. Westland, J. C. (1990) Collaboration and productivity in information systems research.*The Information Society* 7, 33-51.
12. Kleinrock, L. and J. H. Huang (1992). On parallel processing systems : Amdahl's law generalized and some results on optimal design.*IEEE Transactions On Software Engineering* **18(5)**, 434-447.
13. Egghe, L. and R. Rousseau (1990) *Introduction to informetrics* (Elsevier, Amsterdam).
14. Rousseau, R. (1992). Why am I not cited or Why are multi-authored papers more cited than others? *Journal of Documentation* **48**, 79-80.
15. Rao, A. R. and M. N.Vahia (1986). Multiple citations of multi-author publications. *Publications of the Astronomical Society of the Pacific* **98**, 511-512.

Bradford Law versus Physical and Economic Laws

Hajime Eto
The University of Tsukuba
Institute of Socio-Economic Planning, Tsukuba, Ibaraki 305 Japan
E-mail : eto@shako.sk.tsukuba.ac.jp
Home address (private):
Nakano 3-3-17-305, Nakano-ku, Tokyo 164 Japan
Homephone and Fax : (country code 81) 3 3384 2791

A variety of interpretations are assessed for the Bradford's bibliometric law in relation with natural and social phenomena. A cumulative advantage-concentration-overflow-diffusion interpretation is assed empirically against physical phenomena of overflow-diffusion of viscous material and theoretically against the concept of spillover in a prevalent economic theory on research and development (R&D). The bibliometric diffusion is interpreted in relation with the learning or the imitation in strategies of authors or editors rather than the material flow.

Keywords : Information flow, Material flow, Diffusion, Conduction, Spillover, Cumulative advantage, Self-catalytic process,

1. Introduction

Bradford empirically discovered a bibliometric law (the Bradford law) that the number of scientific articles on a research topic is distributed among journals in a peculiar manner. A number of attempts have been made to explain this peculiar distribution. It has usually been interpreted as resulting from the cumulative advantage effect [e.g., 3] and has mathematicallly been given a number of formulae [e.g., 21] and deductive explanations [e.g., 13]. Bibliometric and other distributions were quantitatively compared to demonstrate an essential difference between them [e.g., 15].

This paper compares theories of cumulative advantage effect in physical, biological, sociological and economic phenomena and thereby conceptually clarifies the process yielding the Bradford law.

2. Concentration-Overflow-Diffusion Process

The Bradford law may be stated as follows (Fig. 1). Let the vertical axis denote the accumulative number of articles on a topic and let journals publishing the articles be arranged by the rank order in the number of articles on the topic along the horizontal axis of logarithmic scale. In this transposed rank semi-log coordinate system,

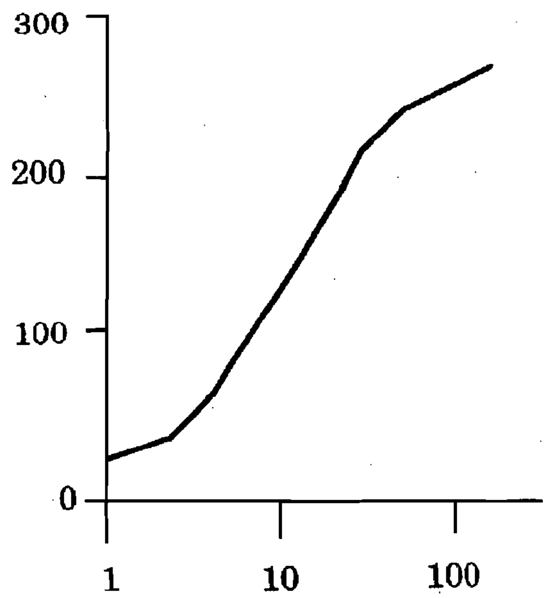


Fig. 1. Typical Bradford graph

the graph is linear for the most part (referred to as the middle semi-log linear section) except the left extreme tail (referred to as the core or nucleus) and possibly the right extreme tail (referred to as the peripheral section). The peripheral section is often concave (referred to as the droop) while it can be convex (referred to as the rising tail). The concavity is interpreted as the incomplete diffusion of the topic to peripheral journals [10] and the convexity is interpreted as the explosive diffusion to peripheral journals [7].

The Bradford law is often interpreted in terms of behaviour of article authors in the following bibliometric concentration-overflow process [e.g., 3].

(B1) The author of the first article on a topic sends the article to a journal by chance.

(B2) The author of the second article on this topic tends to send the article to this journal. The same behaviour is seen for the third and so forth (the cumulative advantage effect).

(B3) This journal is regarded to be specialising in this topic, and hence (nearly) all articles on the topic are sent to this journal (the concentration stage).

(B4) Articles on the topic takes so many pages up to the limit of this journal (the overflow stage).

(B5) Then, authors of articles on the topic begin to send the articles to other journals (the diffusion stage).

Another interpretation in terms of competitition policy-relevant diffusion process may be obtained by replacing (B4) and (B5) above with (B4*) and (B5*) below in regard to policy of journals.

(B4*) The concentration of the topic on the journal begins to give an impact on the editorial policies of other journals (the impact stage).

(B5*) Such journals take an imitation strategy to collect articles on this topic (the spillover stage), and hence articles on the topic begin to appear in these journals (the diffusion stage).

The cumulative advantage, the overflow and the diffusion are not mutually contradictory but are deemed as different stages of the same process.

3. Reproducibility by macro-micro isomorphism

The discussion above assumed as if only the first journal takes an initiative and the others merely follow the first one. However, it may be expected that, among these "followers", relatively active one takesd set of journals reproduces the same behaviours as the original entire set ; that is, the micro-set is "isomorphic" to the macro-set. This discussion leads to a hypothesis: Removing the first journal, the rest also follows the Bradford law ; further, removing the second one, the rest again follows the Bradford law. Moreover, removing all the core journals, the rest again follows the Bradford law. That is, a smaller-sized replica of the Bradford graph is obtained.

This hypothesis is verified for the number of scientific articles among nations in several research areas [10]. In other words, the combination of the concentration with the overflow (or spillover) is sequentially repeated in the diffusion because follower nations are not only passive receptors but also active creators. This may be related with the Ortega hypothesis that middle class scientists are also creative and with its extension that middle class nations may be creative [8].

4. Material Flow Processes

The concentration-overflow-diffusion process is seen in the material flow. An example is the stream of lava in volcanic eruption. Once, lava comes out to the surface through a gap (crator), then lava continues to come out through this crator (a version of the "cumulative advantage" effect), because it is easy for lava to come through the same crator. Consequently, lava is mounted high there (the concentration stage). Then the mounted lava overflows there (the overflow stage) and flows down and skirts therearound (the diffusion stage). An interesting question is whether this "diffusion" follows the Bradford law or not. In order to see this, a test is done in the following procedure : A grid is drawn on the map of a konide volcano (a volcano with a conical shape like the Mount Fuji) so that square cells with equal area are generated by the grid (Fig.2). It is easy to know the height of the center point of a cell formed by the grid by reading the contour of the map. Note that, when the area of each cell equals the unity, the height of a point (more exactly, the height times the unity) is equal to the volume of a column associated with the point (Fig. 3). Therefore, when the height of

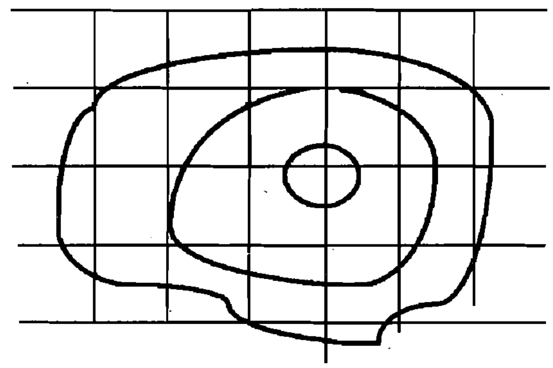


Fig. 2. Contour and grid

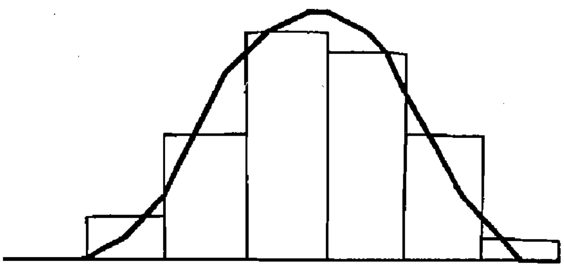


Fig. 3. Hight and volume

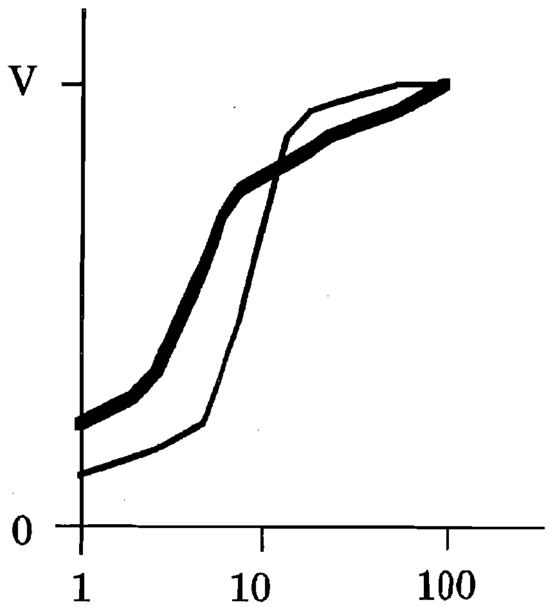


Fig. 4. Bradford graphs of volcanoes
V : The normalised volune of volcanoes

———— : H

▬▬▬ : K

each point is plotted against the vertical axis (the horizontal axis denotes the index number of each cell), the area of the histogram denotes the volume of the volcano, i.e., the amount of erupted lava. Instead of a customary histogram, one plots the accumulated value of height of the points against the vertical axis in the manner of the Bradford graph, where the entire accumulation along the vertical axis denotes the volume of the volcano. For any konide volcano (the Mount Kaimon in Kagoshima Prefecture is more conical than the Mount Fuji), the obtained graph has a short middle semi-log linear section and is not typically of the Bradford graph (K in Fig. 4).

The failure of the Bradford law here may seem consistent with the view that the law is not for physical phenomena [4, 15]. However, let take another volcano Hakone which is a group of volcanoes with several distant crators and hence with seemingly standing alone peaks although the source is common (H in Fig. 5). The application of the same procedure as above to Hakone yields a more typical Bradford graph (H in Fig. 4). This may lead to an interpretation that the Bradford law holds in the case of multi-crator (generically, multi-creator or multi-generator) where the second and the third (and so forth) highest peaks do not only receive lava from the top but also they themselves create independent volcanoes as in the macro-micro isomorphism.

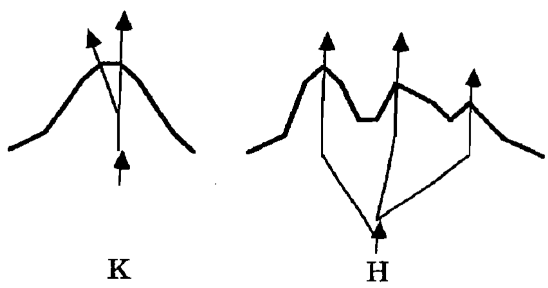


Fig. 5. Konide and multi-crators
The arrows denote eruption and crators

This interpretation leads to a hypothesis that the distribution of scientific articles among nations would exhibit the typical Bradford graph if middle nations were sufficiently creative as almost independent sources in scientific development. Actually, the crator is multiple even for konide volcano (K in Fig. 5), but the crators are not distant enough and are hence merely the passage of the lava stream for the konide (e.g., Kaimon) whereas the crators appear to be creative theselves in the volcanic group (e.g., Hakone).

Further, taking all volcanoes in Japan or in the world (most of the volcanoes are nearly independent but all basically result from the same earth activity) yields much more typical Bradford graphs (here the coordinate system is different from above because only the height of top peak is taken, and the vertical axis therefore does not denote the volume). This finding that the Bradford law holds for the entire volcanoes may be related with the macroscopic property of the Bradford law [2].

5. Population Flow

The concentration-overflow-diffusion process is seen in the population migration. Cities attract people (the cumulative advantage effect). When the population concentrates on cities, it overflows there and is diffused to smaller cities. The population size of cities forms a typical shape of the Bradford graph.

Specific populations such as the population of particular occupation and the criminal population in regions more or less form the Bradford graph. The degree of fitness of these population sizes with the Bradford law depends on the types of populations. The entire population fits the Bradford graph fairly well perhaps because of the macroscopic property of the law [2]. The populations of recently increasing occupations such as fashion designers, scientists, and advertisement professionals have heavier droop sections than those of traditional ones. Similarly, the populations of recently increasing crimes such as drug-related ones have heavier droop sections than those of traditional ones. The last two facts are well explained by interpreting that heavy droops appear in the immature diffusion [10].

The population migration is different from the stream of lava in that medium-sized cities are not merely passages of migration from large cities to small ones like the halfway of volcano but also attract populations by the cumulative advantage. In these respects, the population migration behaviour is similar to the behaviours of article authors' decisions and editorial policies.

6. Concept of Spillover in R&D Economics

In no regard to the above-stated behaviouristic interpretation in terms of overflow, economists of the neoclassic school [5, 23] use the word "spillover" in R&D, which was proposed by Arrow [1]. The spillover means the unintentional overflow in general and the undesirable leak in economics. The spillover in R&D means the impossibility to monopolise the results of R&D activities despite patent laws, or the necessary leak of the results to competitors. As another aspect, R&D activities in a firm contribute to the rise of scientific capabilities in the entire society including its competitors.

These economists are not interested in how to promote technological progresses, but simply recognise that the spillover is necessary in technological progresses. As they assume the economic rationality that each firm tries to maximise its own benefit, they conclude that firms (hence private ownership economy) tend to decline R&D investment. The spillover is an inhibitor of R&D in the market economy. The Arrow's theory was published in 1962 after the USSR surpassed the USA in space technology and Kennedy launched the national space project. This theory nicely explained why the market economy was inferior to the planned economy in R&D and why the USA as a market economy nation needed such a centrally planned national project for R&D.

In the early 20th century, Schumpeter discussed innovation heroes who willingly take risks of first innovation against Arrow's economic rationality. In this regard the two economic theories contradict each other. But Schumpeter also discussed that, as firm sizes grow in economic development, many entrepreneurs tend to avoid risks in a bureaucratic manner. This prediction by Schumpeter held for American industries around the 1960s, on the basis of which Arrow presented his theory [1]. In this respect these two theories are compatible.

It is evidenced that many entrepreneurs with technological background willingly take risks of innovation [22]. Such entrepreneurs little care about the possibility (or necessity) of spillover. Recently, a couple of local governments began to invest in advanced R&D (rather than in the modernization of underdeveloped local industry) by regarding the possibility of benefiting other regions as a sideffect favourable for the entire development [16]. Now it is possible to rediscover the spillover as a promoter (rather than as an inhibitor) of cooperative development.

Combining the economists' concept of spillover with the bibliometric concept of cumulative advantage concentration, the following strategic process may be presented for entrepreneurship.

(E1) First, an entrepreneur (original innovator) begins to invest in R&D.

(E2) This entrepreneur continues the R&D investment by his moral value and thus attracts R&D manpower (the cumulative advantage effect).

(E3) This entrepreneur achieves R&D results and becomes an industrial R&D leader (the concentration stage).

(E4) Some of the R&D results informationally spillover to other entrepreneurs (the spillover stage).

(E5) Such entrepreneurs begin to invest in R&D to develop the results which are spilled over (the diffusion stage).

A similiar process may hold for innovation in general. Analogously to R&D investment, development of high tech industry also follows the Bradford law fairly well. The amount of growth of high technology industry in regions is better explained by the self-reproductive process than by others [10]. Among many industrial terms, technological terms such as the number of patents and the amount of R&D expenditure are more concentrated on core firms than any economic terms such as the amounts of sales and profit [13] or on core nations than any economic terms such as GNP and the income [12].

7. Sectorial behaviour of R&D in competition

The degree of competition affects R&D investment behaviours. As it differs among industrial sectors, the validity of the Bradford law may depend on sectors. Schumpeter claimed that the severe competition (i.e., the low concentration) leads to the decline of innovations. This does not hold for Japanese industries. The steel sector and the automobile sector are more highly concentrated on a few firms and are less R&D intensive than the electrical equipment and electronics sector, the chemicals and synthetic fibre sector and the pharmaceutical sector. The high concentration in the former sectors is related with their low R&D intensiveness because their technology is scientifically saturated and are on the implementation stage for mass production requiring the capital concentration rather than on the R&D stage of diverse trials. It may be argued that Schumpeter's claim does not hold only because a large portion of the industrial R&D expenditure is for development whereas Schumpeter was implicitly thinking of basic research (BR). Indeed, it is theoretically proved that firms under severe competition (i.e., in non-monopolistic sectors) decline BR [17].

As for the R&D expenditure of approximately one thousand firms in the entire industries of Japan, the Bradford law holds quite well. Sorting the firms in sectors (tens of firms in each sector), the validity of the Bradford law is lowered for the sorted sets. A reason is the macroscopic property of the Bradford law [2]. This property does not only mean the problem of sample size. This may rather be related with whether all or only some members (firms or journals) produce and give effects on others. R&D investment of firms in a sector does not directly affect that in other sectors, thus the macro-micro isomorphism is not applicable here. Thus, sectors are not similar to the whole. Among sectors, the degree of (in)validity of the Bradford law depends on characteristics of sectors rather than the size of sectors [6]. The Bradford law is more valid for more competitive or less concentrated sectors (textile ; chemicals and synthetic fibre ; pharmaceuticals) than for less competitive or more concentrated ones (steel ; non-ferrous ; machinery ; electrical equipment and electronics). Among pairs of seemingly similar sectors (textile versus synthetic fibre ; steel versus non-ferrous ; machinery versus electrical equipment), the synthetic fibre, the non-ferrous and the machinery sectors are more competitive and have higher degrees of validity for the Bradford law than the textile, the steel and the electrical equipment and electronic sectors, respectively. This fact may indicate that the validity of the Bradford law is related with the competitiveness. In fact [6], the Bradford law is more valid for R&D-intensive

sectors (electronics ; pharmaceuticals ; chemicals) than for less R&D-intensive ones (textile ; steel ; machinery ; automobile and shipbuilding). The R&D competition more affects the validity of the Bradford law for the R&D expenditure of firms than the commercial competition (measured by concentration of sales and the like). It is an open question whether this finding is consistent or inconsistent with a theory by Scumpeter that the competition discourages innovations.

8. Armament Race Process

The amounts of military resources of nations is found to follow the Bradford law. In particular, the number of personnel of the ground force exhibits a more typical Bradford graph than other military resources (e.g., the numbers of tanks, airplanes or submarines). The latter resources largely depend on specific (geographic or economic) conditions of individual nation (for examples, nations without sea have no need of navy, and other nations without foreign currency can not afford to purchase expensive weapons such as airplanes). As military force does not overflow to enemy nations, its diffusion may rather be explained by competition, that is, by the impact of military policies of some nations on those of other nations (the impact stage as in B4*). Here again, the effect of decision under competition is seen. The impact is not only on the enemy nations but also on the friend nations because the degree of armament is often decided by the formal or informal agreement of allied nations. The impact stage in armament race may be stated as follows.

(M4*) The heavy armament of or the concentration of military force on one nation motivate its enemy nations to have heavy arms and urges its allied nations to cooperate it by the agreement.

9. Biological Self-multiplicative Process

There have been efforts to give mathematical formulae to the Bradford law. However, they were not highly successful in giving a conceptual insight into the structure of the Bradford law. Among the three sections of the Bradford graph (the core, the middle semi-log linear and the peripheral sections), the middle semi-log linear section was proved to be approximated by the Yule distribution [18]. Another mathematical formula was given to both the core and the middle semi-log linear sections[3].

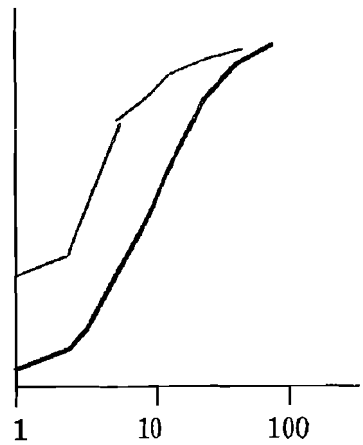


Fig. 6. Armament race
—— : Ground force
—— : Tanks
(The scale of the vertical axis is normalised)

Yet another formula was given to the major part excluding outlyers on the both extreme tails [21]. A unified formula was given to all the three sections [19], but this formula is not flexible enough to fit real data.

Mathematics has been successful in physics but not for the Bradford law. Interestingly, the above-stated formulae represent no physical causality. Further, the heat conduction equation, the conventional diffusion equation or the Fokker-Planck stochastic diffusion equations provide no foundation of the Bradford law. In short, differential equations (both deterministic and stochastic) and most of empirical data for physical phenomena of overflow and diffusion fail to fit the Bradford law composed of the three sections with an exception of the above-stated lava stream and perhaps a few others. This fact may appear to support a view that the Bradford law is particularly for information and is irrelevant to natural science [4, 15].

It may be expected that biological phenomena follow the Bradford law while, as was stated above, physical phenomena do not follow the Bradford law. The Yule distribution, which approximates the middle semi-log linearity section of the Bradford graph, is a stationary distribution of the self-multiplicative stochastic process as in biological population dynamics [6, 9]. The Yule process as the basis of the Yule distribution was conceived as a model of branching of species in the biological evolution process, whereas it is very different from the branching process formulated in a physical stochastic process on the basis of the behaviour of the cosmos ray. These theoretical facts indicate that a major section of the Bradford graph may be relevant to biology in which each entity is creative and self-multiplicative but may be irrelevant to physics in which energy and mass are conserved rather than created.

10. Closed Versus Open Systems and Bivalent Entropy

Mathematical approach to biological population dynamics is based on a variety of differential equations (including stochastic differential equations) of the type to represent the self-multiplication or auto-catalytic process, for example,

$$dx/dt = a\,x + b \qquad a > 0$$

where a and b denote coefficients and x denotes the population size [9, 12, 13]. There has been a severe criticism against the reality of this approach in the recent decades. In fact, this approach quantitatively fails to fit biological data [24]. This failure does not seem to be solved by sophisticating the equations by increasing the mathematical complexity (e.g., by increasing the number of variables or equations in simultaneous systems, and the order of functions). Ironically, the increase of the mathematical complexity leads to the decrease of the reality of the approaches. More specifically, this differential equation approach even qualitatively gives conceptually unrealistic conclusions about the relationship between the stability and the complexity of population (or ecological) systems [20]. That is, the population is theoretically proved to be stable for several cases of simple systems, but the sophistication of these systems by increasing the complexity is proved to destroy the stability. Contradictorily, the stability is empirically observed in complex systems, and the simplification of these systems (i.e., the decrease of the number of species by the extinction) is observed to destroy the stability. It may be argued that the differential equation is valid only for closed systems with the conservation of mass and energy and with the increase of entropy, whereas biological population systems are open without the conservation and with the decreasing (or negative) entropy. The extension of locally valid differential equation to a global domain (that is, the integration) is based on the determinism assumption that no new force but only the inherent force affects a system. Even in inorganic or simple organic systems in chemistry (which have the increasing entropy), however, the kinetics using differential equation approaches sometimes fails to fit the reality [25]. These facts might lead to a hypothesis that the differential equation type approach may not be valid for processes relevant to the Bradford law. The validity of the Bradford law may depend on the closedness or openness of systems, the conservation or the non-conservation of elements and forces, and the increase or decrease of entropy.

Usually, the entropy is said to increase in the diffusion process. In fact,

$$E_P = - \hat{A}_i \; p_i \log p_i$$

increases in the diffusion and decreases in the concentration, where E_P denotes the (physical or positive) entropy, p_i denotes the amount of scientific resource of country (or firm) i. Indeed, E_P attains the maximum when every country (or firm) shares the same amount of resource. On the other hand,

$$E_N = - \hat{A}_j \; q_j \log q_j$$

decreases in the diffusion and increases in the concentration, where E_N denotes another (negative) entropy, q_j denotes the frequency (i.e., the number) of countries or firms with the discretised amount j of scientific resource. Indeed, E_N attains the maximum when only one country (or firm) monopolises the entire resources.

This double character of entropy reflects not only that of the process relevant to the Bradford law with the contraditory character of concentration and diffusion but also the contradiction in concentration or in diffusion itself. The perfect diffusion that every country shares the same amount of resource means that all the countries concentrate on the same status (the same column in a histogram), while the perfect concentration that only one country monopolises the entire resources means that countries are disjointly separated between the have and the have-not (different columns in a histogram).

11. Concluding Discussion

The cumulative advantage effect, that is deeply related with the Bradford law as has been claimed by many investigators, is deeply related with the self-multiplicative process. This process is rather biological or non-physical in that mass or energy is not conserved and entropy is not increasing and in that even followers are not only receptors but also convert to generators. In this regard, the Bradford law is related with the effect of decision because even followers make decisions in social systems with multi-generator.

Extending an idea of the information-relevancy of the Bradford law, this paper claims that the decision-relevancy is a determining factor for the validity of the Bradford law.

A remaining problem is to properly place the problem of science alliance [e.g., 13] in the context of intentional and planned "spillover" and diffusion across firms or nations. Another problem is the catchup of peripheral followers with core leaders [7, 8, 11] in the diffusion process as opposed to the concentration on a few core leaders. The science alliance is related with the catchup in that the former might unintentionally lead to the latter. These problems will be discussed elsewhere.

References

1. Arrow, K. J. 1962. The economic learning by doing. *Review of Economic Studies*. **29**. pp. 155 - 173.
2. Bonitz, M. 1980. Evidence for the invalidity of the Bradford law for the single scientist. *Scientometrics*. **2** pp. 203 - 214.
3. Brookes, B.C. 1969. Bradford's law and bibliography of science. *Nature*. **224**. pp. 953 - 956.
4. Brookes, B.C. 1979. The Bradford law : A new calculus for social sciences ? *Journal of the American Society for Information Science*. **30**. pp. 233 - 234.
5. Bernstein, J. I., and Nadri, M.I. 1989. Research and development and intra-industry spillovers : An empirical application of dynamic duality. *Review of Economic Studies*. **56**. pp. 249 - 269.
6. Eto, H. 1984. Bradford law for R&D expending of firms and R&D concentration. *Scientometrics*. **6**. pp. 183 - 185.
7. Eto, H. 1988. Rising tail in Bradford distribution : Its interpretation and application. *Scientometrics*. **13** pp. 263 - 279.
8. Eto, H. 1991. Science revolution and Ortega hypothesis in developing countries. *Scientometrics*. **20**. pp. 281 - 293.
9. Eto, H. 1992. Stochastic approach to scientific development and innovations (Monograph in English). *Tokyo Daigaku Seisan Gijutsu Kenkyujo Hokoku* (Report of the Institute of Industrial Science, The University of Tokyo), **37**(3). (Serial No. 237). Tokyo.
10. Eto, H. , and Candelaria, P.M. 1987. Applicability of the Bradford distribution to international science and technology indicators. *Scientometrics*. **11**. pp. 27 - 42.
11. Eto, H. , and Fujita, M. 1989. Regularities in the growth of high technology industries in regions. *Research Policy*. **18**. pp. 135 - 153.
12. Eto, H. , and Makino, K. 1983. Theoretical and empirical analysis of differentiation process in technology gap between developed and developing nations. Chapter 9. pp. 149 - 159. In (W. Isard and Y. Nagao eds) *International and Regional Conflict : Analytical Approaches*. Ballinger, Cambridge, Mass.
13. Eto, H. , and Makino, K. 1983. Stochastic model for innovation and resulting skew distribution for technological concentration with verification in Japanese industry. *Scientometrics*. **5**. pp. 219 - 243.
14. Farr, C.M. , and Fischer, W.A. 1992. Managing international high technology cooperative projects. *R&D Management*. **22**. pp. 55 - 67.
15. Haitun, S.D. 1982. Stationary scientometric distribution, Part II. Non-Gaussian nature of scientific activities. *Scientometrics*. **4**. pp. 89 - 104.
16. Iwasawa, M. , Fujita, M. , and Eto, H. 1993. R&D Strategic Policy of Local Government. Chapter 3. pp. 31 - 48. In (H. Eto ed) *R&D*

Strategies in Japan : The National, Regional and Corporate Approach. Elsevier. Amsterdam.

17. Joglekar, P. , and Hamburg, M. 1989. Industry concentration and allocation of resources to basic research. *Management Science*. **35**. pp. 208 - 225.

18. Kendall, M.G. 1960. The bibliography of operational research. *Operational Research Quarterly*. **11**. pp. 31 - 36.

19. Maia, M.J.F. , and Maia, M.D. 1984. On the unity of Bradford's law. *Journal of Documentation*. **40**. pp. 206 - 216.

20. May, R.M. 1973. *Stability and Complexity in Model Ecosystem*. Princeton University Press. Princeton.

21. Morse, P.M. , and Leimkuhler, F.F. 1979. Exact solution for the Bradford distribution and its use in modeling informational data. *Operations Research*. **27**. pp. 187 - 198.

22. Roberts, E.B. 1989. The Personality and motivations of technological entrepreneurs. *Journal of Engineering and Technology Management*. **6**. pp. 5 - 23.

23. Spence, M. 1976. Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies*. 43. pp. 217 - 235.

24. van der Vaart, H.R. 1976. Some examples of mathematical models for the dynamics of several-species ecosystems. Chapter 4. pp. 78-160. In (H. Marcus-Roberts and M. Thompson eds) *Life Science Models*. Springer. Berlin.

25. Wayne, R.P. 1969. The theory of the kinetics of elementary gas phase reactions. Chapter 3. pp. 189 - 301. In (C.H. Bamford and C.F.H. Tipper eds) *The Theory of Kinetics*. Elsevier. Amsterdam.

LETTER TO THE EDITOR

The Three Faces of the Citation-Janus

Analogy of two faced Roman god , Janus is used to depict the Citation Indexes (CI's). Janus proper has only two faces, looking backwards and forwards connected by the same head. The *Citation Janus* looks at the cited items, while opening a perspective for future publications. But for CI's a third face appears to have emerged.The first face, the handsome face of *Citation Janus* (h-face) is the very substance and the reason for which the CI was created. It will continue to have its definite advantages over other structures of bibliographic recordings because it stemmed from an operational mechanism inherent to the process of science . The second face, the mirror-face (m-face) emerged when CI's were made to convolute on themselves in order to rank the very journals nourishing the CI's. The main object of *Citation Indexes (CI)* is retrieval of scientific literature, which is done by h-Face and m-Face. But the third face of the *Citation Janus* is really repelling. While the m-Face impinges only indirectly on to the social process of science, via journals editing, Face-3 does it most directly.The author suggests that ISI® should stop including peripheral journals from the third world countries. Otherwise the scientific endeavour of developed countries are likely to survive the ill effects of m-Face or Face-3 of citation Janus but not that of the developing countries.

For Convenience, I make use here of the name of a Roman god - Janus - to depict the Citation Indexes (CI's, published by the Institute for Scientific Information - ISI - Philadelphia, USA). True enough, Janus proper has only two faces, looking in opposite directions. Janus was worshipped at the beginnings and endings of enterprises. One of his faces, the old one, was meant to look towards past events, the other, the young one, towards the future, both of them interconnected by the same head. To stretch the analogy to CI-s, I presume the "Citation-Janus" looks at the cited items, while opening a perspective for future publications. Within a span of 10 years I expressed my criticicm of the CI's two faces twice, in 1979 and in 1989 [5, 2, 3]. There is now the third one uncovered fully (to me) in a recent article in *Scientific American* [1].

The handsome face of our Citation-Janus (h-Face) is due to the very substance and the reason for which the CI was created by Eugene Garfield : the retrieval of scientific literature. It will continue to have its definite advantages over other structures of bibliographic recordings because it stemmed from an operational mechanism inherent to the science process. It is still my contention [3] that the h-face could shine even more if ISI dropped out the tail of journal titles surrounding the Bradford core representative of the research front, onto which CI's were based. Indeed in [1] there is data indicating that the Third World journals participation in CI's has dropped from some 2.5 percent in 1980-1 to about 1.5 percent in 1992-3. In absolute numbers I guess it amounts to some tens of journal titles. However, the tail I am pointing at contains several hundreds of journal titles. It thus means that the *"outer-shell tail"* has not been curtailed sufficiently. Just how much, and whether appropriately for other regions of the world, remains to be analyzed.

The mirror-face (m-Face) emerged when CI's were made (by ISI) to convolute on themselves in order to rank the very journals nourishing the CI's. This "added-value" m-Face took the real shape as *"Journals Citation Reports"*®. It could not have retained an initial indifference of academic sterility because, obviously, it was creatd to show the CI's "evaluative" potentials and, probably also, to re-approve the selection of journals dealt with by ISI.

As thence editor-in-chief of two journals from the sciene periphery I saw very concretely the ugly lines of the m-Face [5]. The ugliness is there because its *"evaluative potential"* was encroaching agains the natural equilibrium of the science process. The more ambitious editors of journals from the science periphery have been under domestic pressures, in consequence of the ISI's *citation policy* plus the parochialism and local vanities, to bring their journals within the "ISI's Club" (for a very illustrative and recent example from Mexico [see 1]. I exposed the fallacy of such a strategy on two accounts.

Firstly, while doing a great service to the scientific community by the CI's unsurpassed retrieval potential, ISI causes great trouble with its journals' tail from the science periphery in principle, and particularly because of its very dubious selection policy [3, 5].

Secondly, we showed for two chemical journals from science periphery that their incorporation into the *Science Citation Index®* journals selection did not increase their "visibility' in the world science as judged by the SCI's data itself [6]. It appeared that whatever citing there is to the literature from the science periphery, the CI's were not decisive in bringing those results to the attention of the authors from the research front.

My suggestion [3] to ISI was to discard the *outer circle* or the *tailing* of their journals selection. In his editorials in Current Contents, while reflecting on [2 and 3] Eugene Garfield missed (?) this main point on both occasions. The Third Face of the "Citation-Janus" is really repelling. While the m-Face impinges only indirectly onto the social process of science, via journals editing, Face-3 does it most directly. The introduction of citation counts, the crudest of all "evaluation parameters", and counting as much more valid publicaions in journals of the ISI's Club, frequently affects inadvertently the promotion procedures of *individual scientists.* Several illustrative examples of this sort can be found in Gibb's article [1]. (Recently, I was invited to take part in a round-table discussion about the validity of citation data in academic promotions. The encounter was organised by the *Croatian Mathematical Society*, dissatisfied with the decisions taken or announced by the science administrators. It was sad to hear a colleague of mine, a very good scietist at that, defending the administration's "citation case").

My approach here is, of course, from a peripheral *scientist's standpoint.* The huge scientific endeavours in the United States of America, in Western Europe or in Japan, are likely to "survive" the ill-effects of Citation Janus's m-Face and Face-3 within their own social fabrics. But what about the fates that await science periphery, including the h-Face, the good one for retrieval purposes?

Let me tackle first the latter aspect, the h-Face. I believe that, notwithstanding the western-science biases well exposed in [1], the good science papers from science periphery will find their ways by the capillary science information structures into the citation indexes through their being cited by *authors* in the journals of the ISI's lot. However, their "offer on the market" could certainly be improved.I am not an advocate of a "peripheral science citation index". A good sign for improving the h-Face with regard to the science periphery representation is that INFORMANIA's CD-ROMS (for medicine, agriculture and science with technology)[1] might bring more effectively the Third World science articles into focus of the mainstream science. It is as if a new arterial pathway is appearing within the capillary information network. The more of the "peripheral" papers will get into the *references* lists of the ISI's journals lot, the more irrelevant will become th ISI's journals tail. And less important it will eventually appear to the Third World editors to squeeze their journals into the ISI's journal selection. We may thus expect a healthier social fabric of the peripheral science endeavours.

Another positive trend with respect to the m(irror)-Face is *indigenous* evaluation of journals (i.e. *without* absolute resort to citation analysis) like it has been done on an official basis in Brazil and Mexico [1], where the National Science Councils selected a very limited number (a few tens - less than 20%) of domestic journals as being outstanding.

References

1. Gibbs, W. W. (1995) Lost science in the Third World. *Scientific American* (August) 76-83.
2. Maričić, S. (1980) Scientific journals : selection criteria for information services. *4S (Society for Social Studies of Science) Newsletter* (winter), 5, 6-8. (Note : this is an extended summary of the full paper from ref.[6] here).
3. Maričić, S. (1989) Impact factors of peripheral journals. *Journnal of Documentation*, **45**(4), 330-331.
4. Maričić, S. (1993) Evaluating periodicals at the Scientific periphery. *Iaslic Bulletin*, **38**(1), 1-16.
5. Spaventi J., Tudor-Šilović, N., Maričić, S. and Labus, M. (1979) Bibliometrijska analiza znanstvenih časopisa iz Jugoslavije (with English abstract : Bibliometric analysis of scientific journals from Yugoslavia) *Informatologia Yugoslavica*, **11**, 11-23.
6. Zmaić, Lj., Maričić, S. and Simeon, V. (1989) Visibility of pripheral journals through the Science Citation Index. *Information processing & Management*, **25**(6), 713-719.

Siniša Maričić
C/o Science Studies Unit
The Central Medical Library
School of Medicine University of Zagreb
Zagreb Croatia

FORM 1V (See Rule 8)

STATEMENT OF OWNERSHIP & OTHER PARTICULARS REGARDING

JISSI : the international journal of scientometrics & informetrics.

1. Place of Publication	:	New Delhi.
2. Period of its Publication	:	Quarterly.
3. Printer's name	:	Prabhat Pandey (Whether citizen of India - Yes)
Address	:	112, Humayun Pur (near NCC office) Safdarjung Enclave New Delhi -110 029.
4. Publisher's name	:	Prabhat Pandey
Address	:	112, Humayun Pur (near NCC office) Safdarjung Enclave, New Delhi -110 029
5. Editor's name	:	Subir K Sen
Address	:	85(old) Devi Nivas Road, PO-Motijhil, Calcutta -700 074
6. Names and addresses of individuals who own the Newspaper and partners or share holders holding more than one percent of total capital	:	Brzark Information Systems (P) Ltd 112 Humayunpur, (Near NCC Office) Safdarjung Enclave New Delhi - 110 029

I, Shri Prabhat Pandey hereby declare that the particulars given above are true to the best of my knowledge and belief

Dated : 28 February 1996

Sd/- P Pandey
Signature of Publisher

Report On The 4th International Conference On Science and Technology Indicators Performance at the National, Regional and Institutional Level Antwerp, 5-7 October 1995

Mari Davis,
300 Queen Street Melbourne 3000 Australia

The conference, spread over three days - Thursday, Friday and Saturday, 5-7 October, 1995 at Antwerp, Belgium - was attended by over 200 people from 25 different countries and included over 35 oral presentations and 35 poster sessions. It was a predominantly European affair. The North Americans were conspicuously absent (2% of attendees). The Dutch and the Belgians (47%), hosts of the conference, were spectacularly present giving many of the best papers at the conference. France, Germany, Spain, Italy, the United Kingdom, the Scandinavians and former Soviet nations were all well represented (between 5 and 9%). The Asian countries (4%) were represented by Taiwan, Japan and a lone scholar from China.

The programme was divided into a number of sessions based on themes with a total of 38 oral presentations. The sessions were titled :

1. Communication in Science and Technology (3 papers)
2. Science and Technology Interface and Innovation Studies (6 papers)
3. Assessment of National S&T Performance and National S&T Indicators Systems (6 papers)
4. Regional (inter and Intra-national) S&T Performance and Role of Institutes (3 papers)
5. Collaboration and Networks in Research and Development (5 papers)
6. R&D Monitoring, Evaluation and Role of Policy Makers (5 papers)
7. Science and Technology Fields in Regional and International Perspective (5 papers)
8. Science and Technology Policy (3 papers)
9. Anallysis of Emergent Fields : Special Development in S&T Studies (3 papers)

There were also more than 30 poster sessions which were exhibited in several rooms at the conference hotel.

What follows is an idiosyncratic report on some of the presented papers from the perspective of the lone Australian at the Conference. That some papers are not reported on does not signify that they were thought to be of less importance than others; it merely signifies my own personal interests and attention span in what was a long and packed programme of papers, posters and social events.

The first session set the scene. **Hans Roosendaal, Elsevier,** outlined changes to the communication and publication process triggered by technological developments. In responding to change, he said, there are pressures also to make the research process more effective and to serve wider social goals. One of the principal dilemmas for publishers is that *authors want to publish more and readers want to read less*. Roosendaal's view is that publishers are in the centre of this dilemma knowing little about the *use of information* but quite a lot about authors and the publication of new scientific knowledge. He spoke about the nature and role of journals - their authors, their scientific impact, and their form. Journals as intellectual packages are comprised of two elements - functions and identity, which reflect the intellectual and administrative parts of a journal. The functions are certification of the scientific content through peer review, registration of priority in scientific discovery and theory, awareness through making scientific findings public and archival. Under identity are such elements as title, scope, scientific relevance and quality, editorial board and peer review systems. Several areas require research more than others. Bibliometrics may assist in tracking the implications of change by developing new indicators and other tools. He indicates that technology offers the promise of bringing the author and the reader closer together, of bringing about integration and synthesis of the literature which is currently fragmented into separate disciplinary bands, and the possibility of changing the dynamics of the publishing chain. The potential exists for obtaining real-time data on usage patterns that new technological systems of journal will provide. Roosendaal suggests that to shed new light on the communication process, research needs to focus on *Researchers, Intermediaries/Librarians, Information Technology*, and *Access to Information.* First, Roosendaal suggests more research is needed on understanding how information is used because, he said, we know very little about the use of information. The second area for development is in relation to the impact and influence of journals. He pointed to the need for a more dynamic impact factor for journals. The current rather static ISI impact factor which favours short-lived impacts of papers and journals needs to be replaced with a more *useful* one that better reflects the research process.

J. Van Steen, Netherlands Ministry of Education, Culture and Science, spoke on the use of science and technology indicators in the Dutch government. He said that S&T indicators have different functions in the context of science policy depending on the context. In the Netherlands, they are not used directly in government budget allocation as is done in the United Kingdom. The functions include signalling and monitoring to give insight into developments and trends, a symbolic function necessary for national identity purposes, a corrective function for setting aside incorrect images of the functioning of the S&T system, and an accountability and evaluation function where the amounts spent are justified and the performance is measured against the mission and goals of the nation and its various research institutions. Steen referred to the critical question : What use are S&T indicators? He points out that there is a certain tension in the use of indicators. The process is not mechanistic, but a combination of both rational and irrational arguments. Indicators, despite their limitations, are taken into account in the political processes of allocation and resourcing by governments who take S&T seriously. Thus, he concluded, indicators are becoming more political.

Francis Narin, Computer Horizons Inc, spoke about the trend for patents from the US and the UK is to build 'off the scientific frontiers of science'.

Thus, there is a complementary rise of citations to the scientific literature in patent applications. The example he used in his paper was from patents in the human genetics field. He found that similar to citation in the journal literature most citations were to papers by authors within the same country of origin as the patent application (3-5 times the number of citations to papers from other national sources). He also pointed out the strong national component in collaboration seen in patent activity as well as in the scientific literature.

Werner Grunewald, EURO-STAT, presented information on the methodology being used for measuring qualified manpower, HRST (Human Resources devoted to Science and Technology). The HRST model that he presented is based on a stocks and flows model. He described the European Commission's project of building a database of information from member states relating to HRST in their countries. In the near future, the database will provide science policy analysts with a resource with which to answer to questions about the role of human resources in S&T development.

Ed Noyons and **Tony Van Raan, CWTS, Leiden University**, underlined the problems of defining fields and described a new way of obtaining *a structure of science.* They criticize previous methods used to allocate fields to documents or to parts of documents as too rigid and rooted in cognitive structures of science of the past. Present structures of science , they argue, are changing very rapidly and might better reflect the dynamic and emergent nature of many fields. The method they adopt is novel because it allows a retrospective look at the past structure through the lens of the present. The method traces where important present developments found their oigins which may well have been quite outside the field as it was known in the past. To do this, the present structure of a field is analysed and then they reconstruct the present from the past by analysing in the reverse direction. They suggest that the method they have adopted lets the data generate their own structure and that the nature of this is not interfered with by *experts*. However, they use *content descriptors* assigned by the database producer for the cluster analysis of the bibliographic corpus under investigation, in the case described at the conference, neural networks. These *content descriptors* are controlled (index) terms or classification codes. It seems to me that they have fallen into a contradiction-controlled terms (assigned by indexers as *experts*) are based rather more on past rather than present notions about the cognitive structure of science. Index terms are retained until a sufficiently large body of new work forces indexers and batabase producers to recognize that modification or changes to terms are required. However, the authors are aware of the limitations and dangers of using database classification codes.

Giorgio Sirilli, Institute for Studies on Science Research and Documentation, Italy, spoke about the role of innovation in services and how evaluation of its impacts might be achieved. The paper argued that the increasing importance of innovation in the service sector requires a form of measurement different from that applied in the measurement of progress in science. Sirilli referred to the problems of the definition of innovation which he suggested should be strictly related to the use of technology in service delivery or products. He showed a chain-linked model of innovation which moves away from seeing the process in a linear fashion.

Loet Leydesdorff and **Elaine Gauthier, Department of Science and Technology Dynamics, The Netherlands**, pointed out the difference between *field dynamics* and *actor dynamics* in their paper on national performance in Canada and the Netherlands. They focus on the selective stimulation of the science base in these two countries using the emergent fields of biotechnology and advanced materials. They argue that as everything changes over time, an ex-post perspective is needed. Similarly to Noyons and Van Raan, their analysis is reconstructed from a current perspective.

Goran Melin and **Olle Persson, Inforsk, Umea University, Sweden,** reported a study of collaboration patterns at European universities. They looked at differences between universities in terms of collaboration as measured in the output of publications. The differences they investigated were between small and big partners and between small and big countries focussing on the structure of the centre/periphery model. They found no differences between big and small institutions in collaboration showing that, at least for the universities examined, distance and size seem not to matter. They conclude that universities collaborate rather compete and that the principle of scientific self-organization is upheld.

Sylvan Katz and **Diana Hicks, SPRU, Sussex University, UK,** dealt with the question of whether the increasing collaboration between researchers is a direct result of policy initiatives or reflects a process intrinsic to the global scientific community at large. Their paper argues that an extrapolation of data back to 1963 shows that trends in the growth of international collaboration are constant. Their paper suggests that this growth is part of a larger global trend not confined to science. They presented data on international travel and telecommunications systems which show similar growth patterns. Hicks then turned the whole question on its head by suggesting that the presence of a policy might well be considered an indicator of science activity because he claimed that policies follow rather than lead what goes on in science or society.

Remi Barre, Observatoire des Sciences et des Techniques and **Michel Zitt, Lereco, France,** spoke about the transitions being made by nations in translating their scientific publication effort from a domestic model of publication (national language and nationally edited) to international English language publications (transnational model). Their paper points a global transition which is in train in which publication in nationally edited journals is being dropped in favour of publication in English in ISI-source journals. Zitt and Barre suggest that a strategy of publishing in English-language international journals covered by ISI citation indexes will provide greater visibility for small or non English speaking nations. This trend may result in a loss of national publishers of scientific journals. Other transition effects of the *transnational model* of publication may be that papers of less quality will be diverted to national language journals.

Paul Wouters, University of Amsterdam, presented a paper titled the *post-modern citation cycle* in which he presented a dynamic model for science communication which includes both publication, citation and evaluation of science processes. He maintains that scientometrics is redefining the system in subtle ways that relationships in the system are intensifying. In order to describe his extended model, Wouters brought to bear on earlier models of the scientific process by other scholars such as *Rip, Gibbons, Latour and Woogar and Price*. The model shows a reflexive system.

Publication

The Conference Chair, Tony Van Raan, outlined the method of publication for the papers presented during this conference. A group of between 20 and 25 papers will be selected in a peer review process for publication in two journals, **Research Evaluation** and **Journal of the Amercan Society for Information Science.** The proceedings of the conference will include all papers and posters except that those papers accepted for journal publication will be represented by extended abstracts only. The deadline for final copy for presentation is early December. Full publication can be expected in the early part of 1996. Requests for information on publication of the proceedings should be directed to the Editors of the two journals or to the Centre for Science and Technology Studies, University of Leiden, The Netherlands. If you want to be advised when and where the next conference in this series in two years' time is to be held, then write to Professor Van Raan at the Centre for Science and Technology Studies at the Leiden address above.

CALL FOR PAPERS

SIXTH INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS

To be Held at

THE ISRAEL ACADEMY OF SCIENCES AND HUMANITIES

JERUSALEM

JUNE 16-19, 1997

SPONSORED BY
THE INTERNATIONAL SOCIETY FOR
SCIENTOMETRICS AND INFORMETRICS

This is the sixth in a series of increasingly successful biennial conferences. The previous conferences were held at the ***Limburgs Universitair Centrum, Diepenbeek, Belgium;*** at the ***University of Western Ontario, Canada;*** at the ***Indian Statistical Institute, Bangalore, India;*** at the ***Association for Science Studies, Berlin, Germany,*** and ***Rosary College, River Forest, Illinois.***

SCOPE

The scope can be broadly defined as those topics which treat in quantitative fashion the creation, flow, dissemination, and use of scholarly or substantive information.

Representative, but by no means inclusive, topics are : informetric "laws" and distributions; mathematical models of communication; citation analyses; theory of document, text and information retrieval; information and productivity; the quantitative sociology of science and of other substantive information-based activities; study of aging and dispersion; application of informetrics to file design, data compression, etc.; informetric applications to policy analysis, R&D management, etc.; ...

LOCATION

The conference will be hosted by the School of Library, Archive and Information Studies of the Hebrew University of Jerusalem. It will be held at The Israel Academy of Sciences and Humanities.

Jerusalem is warm but pleasant in June. The conference venue is situated in the most attractive part of town, next door to the President's Residence and the Jerusalem Theatre. The Ben-Gurion International Airport is a 35 minutes drive from Jerusalem.

We look forward to seeing you all in Jerusalem in June 1997.

Please submit general information requests to the conference organizer :

Professor Bluma C. Peritz,
Director,
School of Library, Archive and Information Studies,
Hebrew University, Givat Ram Campus,
P.O.Box 503 Jerusalem 91904, ISRAEL.
Fax: +972-2-6585707
Phone : + 972 - 2 - 6585044; 6585045

Timeline : Submission of Abstracts 30 August, 1996
Submission of Papers 30 November, 1996
(Note, the intent to submit a paper is welcomed at any time)

Please submit abstracts and papers or questions about papers to the appropriate committee

Regional Technical Program Committee Chairs:

For the Americas, Africa, and Otherwise Unclassified :
Prof. Abraham Bookstein
Regenstein Library,
1100 East 57 Street
Chicago, Illinois 60637, U.S.A.
Phone : +1-312 -702-8268
Fax : +1-312 - 702-0775
E-mail - Internet : a-bookstein@uchicago.edu
Bitnet: bkst % tira. uchimvsl

For Europe :
Prof. Leo Egghe
Limburgs Universitair Centrum,
Universitaire Campus,
Gebouw,
B-3590 Diepenbeek, Belgium.
Phone : +32 - 11 - 268121
Fax : + 32 - 11 - 268126
E-mail : legghe@luc.ac.be

For Asia :
Prof. I. K. Ravichandra Rao
DRTC, Indian Statistical Institute,
8th Mile, Mysore Road,
R. V. College P.O.
Bangalore 560 059, India.
Phone : +91 - 080 - 8600648/8600485/8600616
Fax : +91 - 080 - 8430265.
E-Mail : drtc@isibang.ernet.in

JISSI Data Sheets

Title of the journal : JISSI : the international journal of scientometrics and informetrics

ISSN : 0971 - 6696

CODEN :

Established : March 1995

Periodicity : Quarterly (issued March, June, September, December)

Subjects covered : Scientometrics, Informetrics, Bibliometrics, Technometrics, Librametrics, Library statistics, Citation & co-citation analysis, Co-word analysis, Science indicating & forecasting, Psychometrics and sociometrics of inention, discovery, authorship and informational and communicational activities, Econometrics of information, Historiometrics of knowledge

Broad subjects : Science of science; History of science, technology, social science, arts and humanities; Sociology of knowledge and information; Library and information science; Information management; Economics of information; Psychology and socio-psychology of information and knowledge; Applied statiic; applied mathematics;

Place of publication : New Delhi, India

Publishers : Brzark Information Systems(P)Ltd

Publishers, address :112 Humayun Pur, Safdarjung Enclave, New Delhi 110 029, India,
Telephone +91-11-688-2366 Telefax+91-11-687-4472

Editor : Subir K Sen (Honorary Executive Editor)

Editorial address : 85 (Old) Devi Nivas Road, P.O. Motijhil, Calcutta 700 074, India,
Telefax +91-33-551-2180

Collaboration : In association with the International Society for Scientometrics and Informetrics (ISSI)

Attributes : JISSI is an international journal representing a challenging new field of scientometrics and informetrics (or rather sciento-infor-metrics). It provides a forum for dissemination and exchange of ideas and scholarship in this field

The subject field : Sciento-infor-metrics is a fast growing multidisciplinary as well as interdisciplinary field of research and study. The field may very well represent quantitative (and qualitative) assessments of man's cultural progress (sociometry and historiometry) and advent into information age. University departments and academic institutions dealing with history and sociology of science, science of science, library and information science etc are imparting regular and irregular teaching courses and running doctoral programmes in different areas of this field

Target readership : Research workers, teachers, students, practising information scientists, librarians, information managers, science and technology policy makers, historians of science, social scientists, statisticians, etc.

Coverage : JISSI is a refereed journal devoted to the quantitative aspects (including mathematical, statistical, computational analysis and modelling) of - science of science & technology (including technometrics) -studies on social sciences, arts and humanities (particularly based on bibliographic data) - all types of information processes e.g., generation, growth, diffusion, dynamics, use etc of information systems and services (including libraries, archives, databases, information centres or any other type of document collections) which may go by the name of librametrics & any study that may fall under the purview of bibliometrics

Features : Research papers, reviews, short communications, letters to the editor, book reviews, reprint of classical and very important but apparently obscure papers relevant to the field, profiles of important personalties (these should be serious essays dealing with the work and just not life sketches), glossaries of scientometric and informetric terms, editorials and guest editorials (solicited and unsolicited), issues and unsolved problems, ISSI Newsletters (as and when provided for by the ISSI board), selected and refereed papers from a relevant Conference, Seminar, Symposium etc.

Size : A4 (27.5cm X 20.5cm approx)

Print area : 24cm X 16cm

Advertisements : JISSI publishes selected ads

Price (Single copy) : US $ 55=00/Rs. 750-00

Detailed Subscription Information

Annual subscription :	INSTITUTIONS	INDIVIDUALS	ISSI MEMBERS
Developed countries :	US$ 200=00	US$ 100=00	US$ 75=00
Developing countries :	US$ 125=00	US$ 50=00	US$ 30=00
Indian subcontinent :	Rs.2500=00	Rs.1000=00	Rs.750=00

**Special offer for advance subscription for three years : Double the amount of Annual Subscription

International Society for Scientometrics and Informetrics (ISSI)

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

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

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

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

2. Mathematical, statistical and computational modelling and analysis of information processes.

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

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

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

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

Meanwhile, plans for the other International Conferences took shape.

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

ISSI Membership Information

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

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

JISSI is issued quarterly. Correspondence regarding advertishment and subscription should be made to the publisher. Payments may be sent by Cheque/Bank Draft/Money order.

For cheques outside Delhi but from within India a sum of Rs 15/- should be added. For cheques out side India a sum of U.S. dollar 2.00 or equivalent should be added.

Claims for missing number(s) should be made within four months of the date of issue of such number to the Publisher.

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

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

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

Subscription Rates (All rates include airmail surcharges)

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

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

Special Offer for Other (Developing) Countries:

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

Rates for Indian Subcontinent:

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