Scientometrics, Vol. 26. No. 1 (1993) 37-50

THE SCIENCE STRATEGY INDEX

M. BONITZ,* E. BRUCKNER,** ANDREA SCHARNHORST**

**Central Institute for Nuclear Research O-8051 Dresden, P.O.Box 19 (Germany)*
***Wissenschaftszentrum Berlin für Sozialforschung (WZB), Reichpietschufer 50, D-1000 Berlin 30 (Germany)*

(Received January 6, 1992)

A new indicator, Science Strategy Index, is proposed, which is based on the scattering of a country's science activity over all science fields and related to the world distribution of the science fields. The indicator allows to compare the structure of the publication output of countries as reflected by the used database, irrespective of the size of the countries.

If the science structure of each country is related for comparison to that one of each other country, the indicator converts into a structure measure which enables to cluster countries according to their structural similarity. The cluster map of countries achieved in this way deserves intense discussion upon the different science strategies of countries and their geographic, political, communicative, and socio-cultural background.

Introduction

Science indicators for the assessment and comparison of countries on the basis of publication and citation counts can be followed through several levels. At each level less or more complicated indicators can be derived and less or more far-reaching conclusions for science policy decisions can be drawn.

To begin with the most simple indicators, namely the absolute values of the *publication and citation counts*, respectively, the countries can be compared (by one field or by all fields) after ranking them according to descending publication counts. The analogue procedure can be carried out with the citation counts.

Another way of getting indicators is based on relative *numbers*, i.e. *shares* of the fields within a country's field distribution ("country cake") or within the world's field distribution ("world cake").

Schubert, Glänzel and *Braun* introduced a quotient *"activity"* in the case of publication shares, and *"attractivity"* in the case of citation shares by dividing the country share by the world share[1]. In this way the world share of a field is taken as a kind of standard for that field. Relating this standard to the field's share of any country, one can compare countries according to one field.

Elsevier, Amsterdam – Oxford – New York – Tokyo
Akadémiai Kiadó

It is quit clear that these indicators make sense *for single fields only*. On the other hand 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. Following this need different methods have been proposed, e.g. correlation measures[2], factor analysis[3], chi-square methods[4] etc.

Obviously it is needed to develop *structural indicators*, which simultaneously include *all fields* and thus enable *pattern comparison* of, for instance, science structures or field distributions.

Definition of the indicator Science Strategy Index (SSI)

Among several possibilities we have chosen a simple one:

$$SSI_i = 100 - 1/2 \left(\sum_f \left| p_{if} - p_{wf} \right| \right) \quad (\%)$$

Here, at first for a field f and a country i the country share of the field, p_{if} [%], is subtracted from the world share of the field, p_{wf} (%), taken as a positive deviation, and then it is summed over all fields. As a result of such a pattern comparison (country cake *vs*. world cake) we get SSI-values between zero and unity. Unity means exact matching of the country pattern to the world pattern, zero indicates that there is no matching at all.

The data analysed were taken from the comprehensive report of *Schubert, Glänzel* and *Braun*[1] which covers the world publication and citation output of the years 1981 – 1985 as reflected by the *Science Citation Index*.

The science structure of a country is set up by considering five major fields: life sciences, physical sciences, chemistry, engineering and mathematics. Those countries have been taken into consideration only having more than 50 publications in each field in the period.

As an example let us compare the patterns of the United States (USA) and the Soviet Union (SUN) with the world pattern (see also Fig. 1). The SSI value for USA is high, 93.6%, whereas the SUN value is rather low, 68.8%. This is apparently due to the big SUN deviations in the life sciences (lower than world value), as well as in chemistry and physical sciences (much higher than world values).

Full data for all countries are given in Table 1, which is the basis for the presentation as a rank distribution (Fig. 2).

Now, let us shortly discuss the results achieved so far. Firstly, the form of the rank

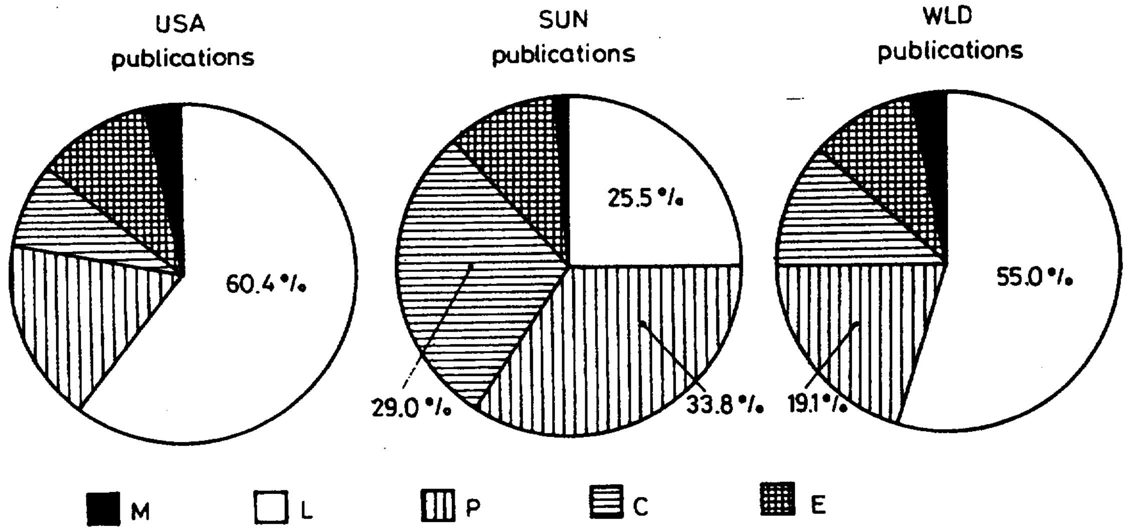


Fig. 1. The science structures of USA, USSR (SUN), and the world (WLD). L – Life Sciences, P – Physical Sciences, C – Chemistry, E -Engineering, M – Mathematics. Data source Ref.1

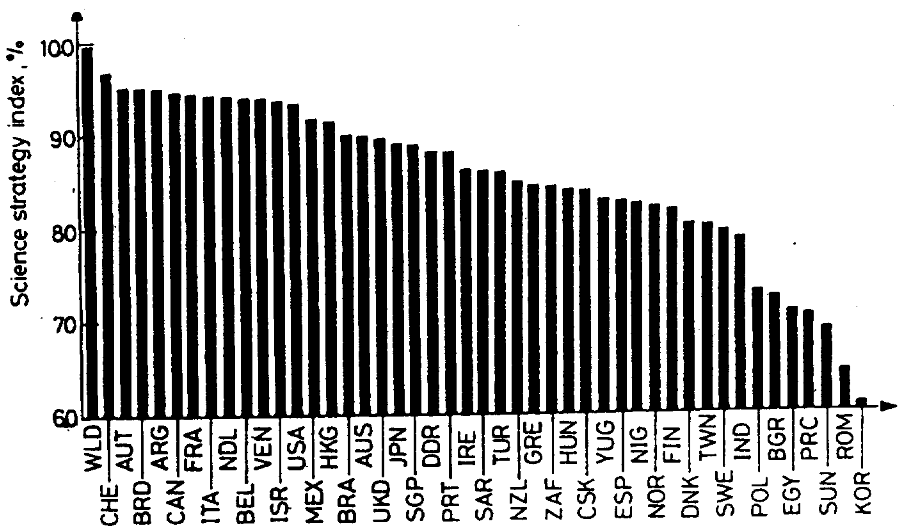


Fig. 2. 45 countries ranked by their SSI value. Data source Ref.1

distribution does not exhibit any Bradfordian behaviour at all. The curve decreases smoothly. This indicates, according to a paper of one of us,[5] that the indicator SSI really reflects a structural phenomenon. Secondly, the distribution shows, that every country has a fair chance to be positioned at low ranks (or to possess a high SSI value), if it only has a science structure near to the world science structure (or when the country cake is very similar to the world cake).

The largest science nations do not necessarily rank first, although they of course highly dominate the world structure. On the other hand a country with a small contribution to the world total not necessarily has a small SSI-value.It can fairly well match the world structure. The first ranks in the distribution are occupied by Switzerland (CHE), Austria (AUT), Federal Republic of Germany (BRD), Argentina (ARG), and Canada (CAN). USA has rank number 12, Japan (JPN) has rank number 18, India (IND) has rank number 38, and SUN has even rank number 43.

Table 1
SSI values and publication counts for 45 countries and for the World

Country	SSI	Publications	Country	SSI	Publications
WORLD	100%	1937470	SAR	86.0%	1632
CHE	97.1%	23672	TUR	85.9%	1489
AUT	95.5%	10670	NZL	84.9%	9510
BRD	95.4%	116138	GRE	84.5%	4573
ARG	95.3%	5495	ZAF	84.4%	10253
CAN	95.0%	82567	HUN	84.1%	9406
FRA	94.7%	88201	CSK	83.9%	15159
ITA	94.5%	45191	YUG	83.0%	4541
NDL	94.4%	33717	ESP	82.8%	16551
BEL	94.3%	16943	NIG	82.5%	3871
VEN	94.2%	1421	NOR	82.3%	10189
ISR	93.9%	21130	FIN	81.9%	12443
USA	93.6%	718334	DNK	80.3%	16219
MEX	92.0%	3413	TWN	80.1%	2594
HKG	91.6%	1788	SWE	79.6%	32313
BRA	90.1%	6885	IND	78.8%	50015
AUS	90.1%	43589	POL	73.0%	17927
UKD	89.6%	172924	BGR	72.4%	3026
JPN	89.2%	138722	EGY	70.8%	5088
SGP	89.0%	1057	PRC	70.4%	7596
DDR	88.3%	17143	SUN	68.8%	126613
PRT	88.2%	1210	ROM	64.4%	2969
IRE	86.2%	3548	KOR	60.9%	1590

It follows from these considerations, that by no means a conclusion like 'CHE ist best, South-Korea (KOR) is worst' can be permitted. The 'first five', CHE, AUT, BRD, ARG, and CAN do *match best* to the world structure. Their SSI values are lying between 95 and 100%.

The proposed indicator is a measure for the similarity of a country's science structure to the world structure exclusively. In principle, the science structure of any other country could be taken for comparison. Thus we arrive at a more general definition of the Science Strategy Index:

$$SSI_{ij} = 100 - 1/2 \left(\sum_f \left| p_{if} - p_{jf} \right| \right) \quad (\%)$$

which compares the structure of the country i with the structure of the country j, and measures the degree of similarity of the two countries. Let us now imaging an abstract five-dimensional space, where the axes are related to the proportion of a given field with respect to the total publication output. Than one may represent each country by a point in this space. The second term of the right hand side of the above equation is a metric, a distance measure in this five-dimensional space. Of course, in this space also other distance measure may be introduced, e.g. an Euclidian Metric. In a first step we choosed a very simple possibility.

The SSI indicator reflects the 'neighbourhood' in a structural sense of the countries i and j in the considered space. Now to each country as many different values of SSI can prescribed as this country finds partners. These values can be inserted into a COUNTRIES-COUNTRIES matrix as matrixelements (Table 2). The diagonal elements are, of course, equal unity, the matrix is symmetrical, i.e. we get one SSI value for each pair of countries. In Table 2 the sequence of countries corresponds to their absolute number of publication counts (from USA to CAN). The world (WLD) is added like a any country.

From Table 2 can be seen, that neighbourhood with respect to publication output has generally nothing to do with structural neighbourhood. Countries with very similar publication output can differ significantly in their distribution of the scientific fields.

What we can hope to find, however, is that there exist groups or clusters of countries with similar science patterns. Or we can hope to meet groups of countries belonging to different structural types. But we must also be ready to observe a multiplicity of countries structurally isolated from each other or mutually connected only by chance.

Table 2
COUNTRIES-COUNTRIES matrix with the SSI_{ij} values (in %) as matrix elements

	WLD	USA	UKD	JPN	SUN	BRD	FRA	CAN	...
WLD	100.0	93.6	89.6	89.2	68.8	95.4	94.7	95.0	...
USA		100.0	93.5	83.7	62.4	89.8	88.6	97.8	...
UKD			100.0	79.6	58.9	84.9	86.0	93.7	...
JPN				100.0	76.3	92.6	87.4	84.8	...
SUN					100.0	72.6	70.2	63.7	...
BRD						100.0	94.9	90.9	...
FRA							100.0	90.2	...
CAN								100.0	...

The co-structure clustering process

Now, the COUNTRIES-COUNTRIES matrix may be taken as a basis for a simple co-structure clustering process in order to determine groups of countries with similar structure. A usual procedure in clustering is thresholding which is performed to get rid of the noise. Our 46 countries (WLD included) form $46 \times 45 = 2070$ partnerships, most of which are weak and of no interest. Our threshold was set at SSI = 95%, yielding 32 countries (WLD included) which form 49 partnerships indicated by bold lines (dashed lines for WLD as partner) on Fig. 3a. One has to bare in mind that the lines in this figure just denote, that neighbouring countries have similar science patterns, they are nearest neighbours in a higherdimensional space. No quantitative conclusions can derived from these figures which are a simple twodimensional representation of a fivedimensional phenomenon.

Fourteen countries are missing in Fig. 3a, among them big scientific nations like Soviet Union, India (IND) and the Peoples Republic of China (PRC). They appear at the following threshold values for SSI_{ij} and in the neighbourhood of the following countries: Turkey (TUR) at 94.6% near GRE and PRT, IND at 93.7% near BGR, Yugoslavia (YUG) at 93.7% near IND, Mexico (MEX) at 93.7% near NDL, Hungary (HUN) at 93.2% near ESP and CSK, Brazil (BRA) at 93.1% near FRA, SUN at 92.7% near BGR, Singapore (SGP) at 91.7% near PRT, Saudi Arabia (SAR) at 91.0% near PRT, Romania (ROM) at 90.9% near SUN, Taiwan (TWN) at 89.8% near TUR, PRC at 89.3% near BGR, Egypt (EGY) at 86.4% near IND, and finally South Korea at 85.1% near ROM.

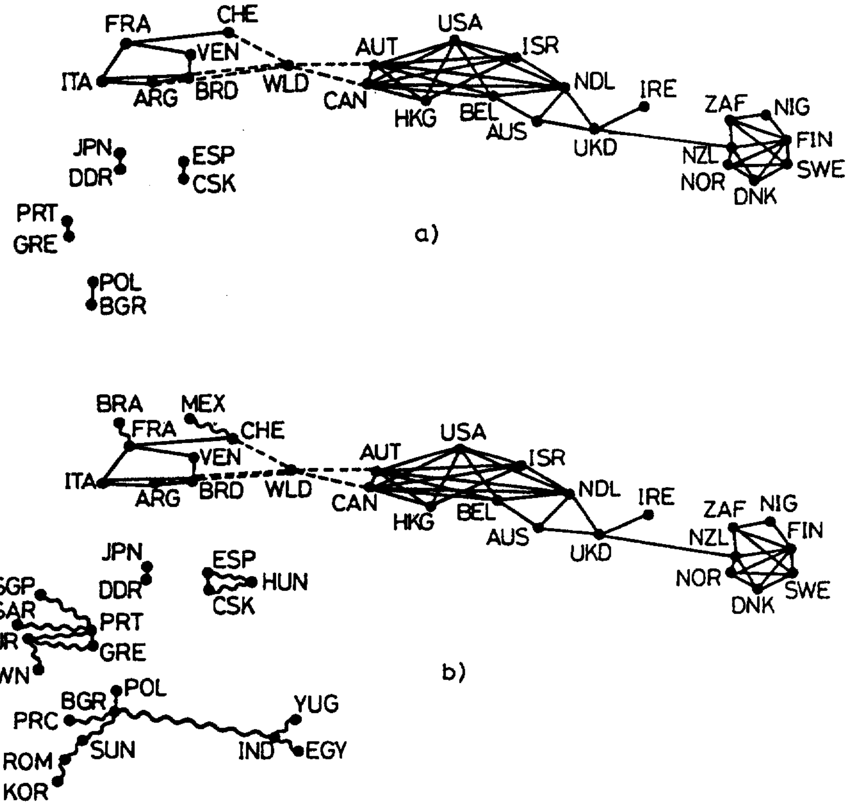


Fig. 3. a) Cluster map for couples of countries with SSI_{ij} values of 95.0% and higher; b) cluster map after lowering the threshold and including the rest of the countries locating them near by the countries with the highest structural similarity

It should be emphasized, that in Fig. 3b these fourteen countries are not shown in their multiplicity of partnerships, but as single circles linked with wavy lines to their most similar partner. Thus every country is present in Fig. 3b and nonetheless the presentation is not overloaded.

Discussion of the cluster maps

Already at first glance, the cluster map of Fig. 3a exhibits several remarkable regularities. Let us begin with the 'country' world (WLD), which we positioned into the center of the map. There are, as mentioned before, five countries nearest to WLD: CHE, AUT, CAN, ARG, and BRD. Among them only ARG and BRD are inter se connected.

The 'world' functions like a bridge between neighbouring clusters. There are other countries obviously fulfilling a bridge mission between greater clusters, namely The Netherlands (NDL), and United Kingdom (UKD). Like WLD 5 structurally similar neighbours possess Isreal (ISR), Belgium (BEL), and New Zealand (NZL). 6 neighbours have USA, Austria (AUT), Canada (CAN), and Finland (FIN). The Netherlands (NDL) are even closely linked to 7 other countries.

There are two five-member clusters: USA-AUT-CAN-HKG-ISR and USA-AUT-CAN-NDL-ISR. Belgium is involved in two four-member clusters: BEL-CAN-AUT-NDL and BEL-AUT-USA-NDL, and the scandinavian countries form a four-member cluster too.

24 countries are contained in a great complex cluster, the members of which are interlinked in one or the other way: they are at least not isolated, although they may have quite different science structures.

There are four isolated pairs of countries without any linkage to other countries: Poland (POL) and Bulgaria (BGR), Spain (ESP) and Czechoslovakia (CSK), Japan (JPN) and German Democratic Republic (DDR), Portugal (PRT) and Greece (GRE).

Not contained in the cluster map of Fig. 3a are 14 countries, among them IND, SUN, and PRC. These countries, as already mentioned, are joining to the picture, when the conditions are weakened. Eventually we have the 46-countries presentation of Fig. 3b.

Now we start with the task to interpret these maps, keeping in mind, that they are built up by *co-structure clustering*. Everybody will certainly raise a first question: where is my own country located? Then follows the question: why here? Then comes the ultimate question: what kind of conclusions, if any, could be drawn from the picture?

We observe features in the map which seem to be not unexpected. There is, firstly, the cluster of the scandinavian countries: NOR, FIN, DNK, SWE; we conclude, that this is due to their geographical neighbourhood, their similar languages, and that these countries function as if it were *one* country.

There is, secondly, the "bridge" United Kingdom (UKD), which has several links: to Australia (AUS), Ireland (IRE), New Zealand (NZL), Nigeria (NIG), and South Africa (ZAF) – has that to do with the worldwide Commonwealth?

One more link goes from UKD to The Netherlands (NDL), which bridge via Belgium (BEL) to another cluster complex containing the big science nations USA and Canada, together with NDL, BEL, ISR, and HKG. This definitely cannot be explained neither by geographical neighbourhood nor by historical roots only.

That the 'world' (WLD) is not member of a cluster, but rather serves as a bridge between the northamerican states and european countries like France (FRA), Federal Republic of Germany (BRD), Switzerland (CHE), and Italy (ITA), deserves explanation. These countries have as close neighbours the southamerican countries Argentina (ARG), Venezuela (VEN), and Brazil (BRA). That should not be a random coincidence.

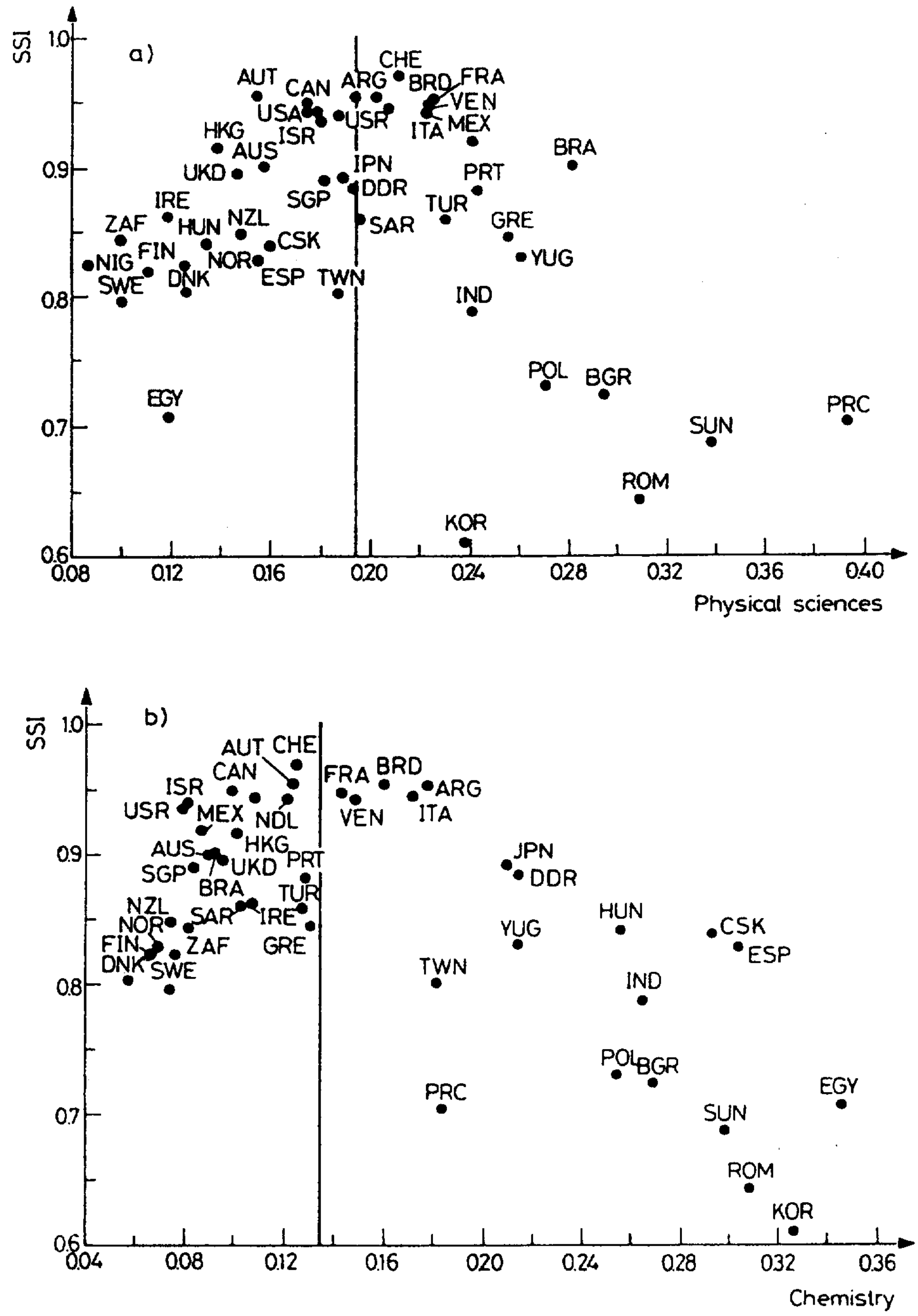
To the isolated countries belong not only all (former) socialist countries, but also Japan (JPN) and India (IND). So, oversimplified explanations surely will not satisfy our curiosity. Especially the close constellation of JPN and German Democratic Republic (DDR) obviously does not allow the explanation neither that DDR is 'as good' as JPN nor that JPN is 'as bad' as DDR: both countries are not in geographical or political neighbourhood. There must be other reasons for their very similar science structure or, as we believe, very similar science strategy. It may be, that a severe neglect of the life sciences, stemming from a restrictive science policy in the case of DDR, and an overweigh of the classical sciences, stemming from deep cultural roots in the case of JPN, bring both countries so close together in our picture.

Now we will try to improve our understanding of the cluster maps by consulting our data concerning three *single* major fields: physical sciences, chemistry, and life sciences.

In Fig. 4a the SSI values of the countries (with respect to WLD) are plotted against the shares of the countries in physical sciences. The straight line is the WLD value of the physical sciences' share: 19.1%. In Fig. 4b the SSI values of the countries (again the WLD value is 100%) are plotted against the shares of the countries in chemistry. The straight line gives the WLD value of chemistry's share: 13.2%.

Nearly all (with few exceptions) of the isolated countries in the cluster maps, which possess relatively low SSI values, have much higher shares in these classical fields than the world (WLD). In other words, their 'activity index' is higher than unity in these fields. On the other hand, all members of clusters in Fig. 3a, in general all countries of the big joint structure, are found together also in the plots for the physical sciences and chemistry.

Even more impressive, obviously due to the dominating world share of the life sciences (55.0%) is the Fig. 4c: SSI values versus life sciences' share. Here all, without exception, isolated countries of the cluster map have life sciences' shares smaller than the WLD value: their 'activity index' is smaller than unity, and the joint cluster's members of Fig. 3a keep their tight neighbourhood also in Fig. 4c.



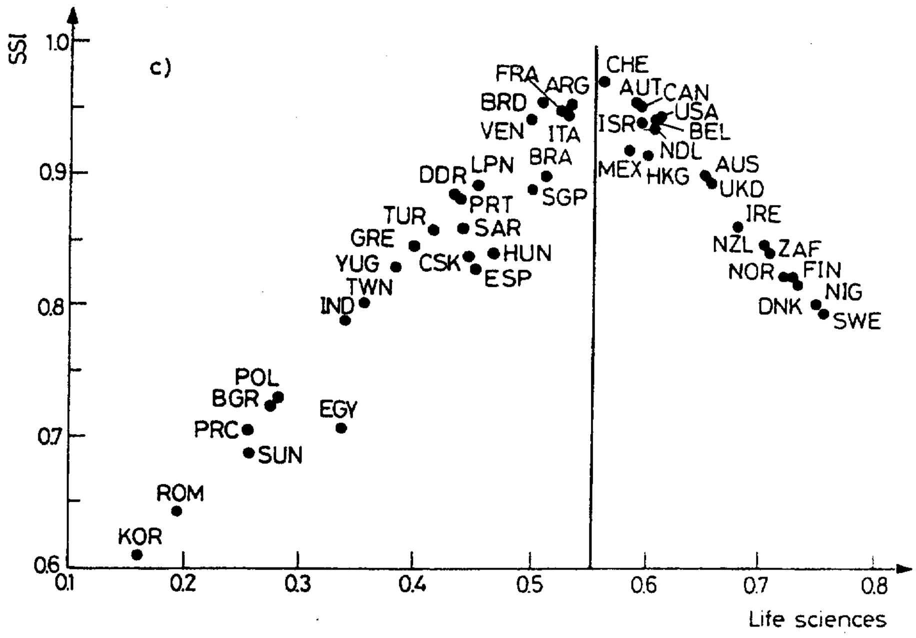


Fig. 4. Science Strategy Index (with respect to WLD) for 45 countries versus the countries' publication shares in main scientific fields. a) Physical Sciences b) Chemistry c) Life Sciences. Data source Ref.1

So, the onedimensional, ore one-field, representation of Fig. 4c may well give an impression of what is actually happening in the fivedimensional space, or five-fields structural world of science, arbitrarily projected into the twodimensional representation of the cluster maps in Fig. 3.

The obvious dominating role of the life sciences' shares for the layout of the co-structure cluster maps inspired us to return to Fig. 3 once more, and to try another representation: in Fig. 5 the country circles were replaced by the circle diagrams of the countries. The white sectors are the shares of the life sciences, the black sectors the shares of all other fields. Although the shares of all fields decide upon close or loose neighbourhood of two countries, it can be clearly seen from the structure of the "*atoms in the world science molecule*", that from KOR (15.9%) via BRD (50.9%) to SWE (75.9%) the life sciences' share is one of the most important parameters for the location of the countries in the map.

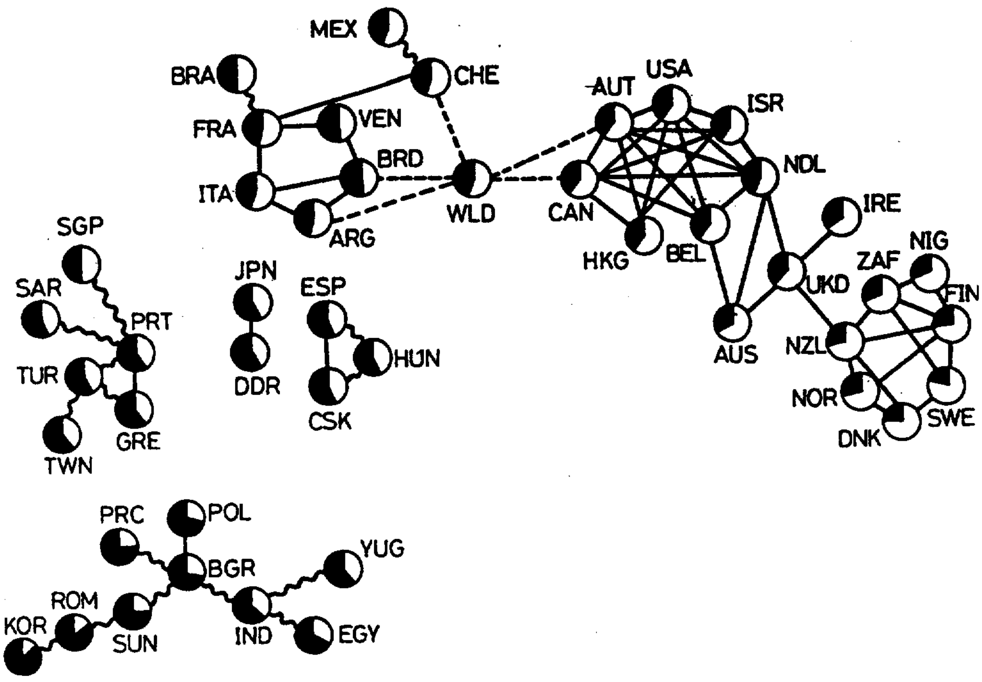


Fig. 5. *Country atoms* in a *world science molecule* underline exceptional role of the life sciences' shares

Main conclusion and discussion

Every country doing science has a science structure (country cake, circle diagram). Due to the choice of our indicator the science structures of the countries are comparable irrespective of the size of a country's publication output. This may be a disadvantage as well as an advantage. However, it seems to be encouraging that the obtained cluster map shows more regularities than common scientific sense would have predicted.

The observation that the science structure of countries is characterized by certain patterns (clusters) is well in accordance with earlier investigations performed by *Frame* and others[3] who used correlation coefficients and methods of factor analysis for investigating national science production (SCI data base 1973).

Although the application of another method bars us from directly comparing results it turns out to be interesting, that some of these regularities obviously exist over a greater time scale. In particular, this concerns the similarity between Northern American and Western European countries with respect to the dominance of life science, the interrelation of countries of the Commonwealth and also the correlation between Japan and Eastern European countries.

More interesting, in principle a recent study of *Eto*[2] using different correlation measures to investigate the similarity between the science structure of each two countries in the same databasis as we relied on, finds structural similarities between countries, which following our method are identified as "neighbours" in a cluster in the five-dimensional space. This concerns e.g. the interesting but non-obvious correlation between Japan and German Democratic Republic.

These results appear to confirm our main conclusion, that these structures are not merely artifacts of the method applied but a real phenomenon.

However in order to validate the results obtained further a detailed investigation of other metrics (e.g. Euclidian metric), other similarity respectively correlation measures up to factor analysis and other cluster methods should be carried out. Also should be further checked whether the structural similarities found are stable against changes in the data base (including error estimation) and against changes in the classification.

But the most interesting question is to explore *how these* structures have been developed. To what degree they can be understood as a result of a dynamical self-organization process in the science system?[6] To investigate this problem it is necessary to consider temporal changes of these stuctures. The implementation of a metric measure for clustering the science structure of countries represented in a five-dimensional space also allows to investigate the evolution in time of these clusters. The dynamics of this evolution is represented by a movement of country points on a simplex in the space and can be modelled correspondingly.

Possible questions for further investigations might be:

Is there a movement of countries within the cluster map? Which countries are involved, and which are not? Is there a trend for certain countries to join to certain types of clusters – may be to one common type for all countries? Can different types of countries be identified dynamically competing in science as *Eto*[2] proposed?

Another problem concerns the consequences from the enormous differences in the science structures exhibited by great science nations like, for instance, USA, SUN, JPN, IND, BRD, and PRC. Will our supposition turn out to be correct, that the absolute amount of resources spent for science by a country is one thing, the *right distribution* of these resources over the main science fields quite another one?

Hopefully, the observed regularities will be found interesting and a challenge not only for scientometricians, but also for the scientist of science, and may be for the decision maker in science policy.

The world citation data should be exploited as well. An indicator constructed in analogy to SSI (SRI: *Science Response Index*) could be used to derive cluster maps based on the field structures of citations received by the countries. The interrelation of SSI und SRI and the correlation of these indicators with socio-economic indicators could be of interest.

*

We are grateful to *Wolfgang Glänzel*, *Peter Vinkler* and *András Schubert* for helpful discussions in the starting phase of this paper and to *Remi Barre* for important comments. We wish to thank *Peter Weingart* for his hospitality during a stay at the *Center for Interdisciplinary Research*, in Bielefeld October 1990, where the first ideas on the new indicator sprang up.

References

1. A. Schubert, W. Glänzel, T. Braun, Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields, 1981-1985, *Scientometrics*, 16 (1989) 3-478.
2. H. Eto, Science revolution and Ortega Hypothesis in developing countries, *Scientometrics* 20 (1991) 283-295.
3. J. D. Frame, F. Narrin, M. P. Carpenter, The distribution of World science, *Social Studies of Science*, 7 (1977) 501-516.
4. M. Pianta, D. Archibugi, Specialization and size of scientific activities: a bibliometric analysis of advanced countries, *Scientometrics*, 22 (1991) 341-358.
5. M. Bonitz, Journal ranking by different parameters. Part II. Individual or collective: Which parameters are best suited for journal ranking? *Scientometrics*, 18 (1990) 75-93.
6. E. Bruckner, W. Ebeling, A. Scharnhorst, The application of evolutionary models in scientometrics, *Scientometrics*, 18 (1990) 21-41.