Scientometrics, Vol. 26. No. 1 (1993) 135-156

THE DELINEATION OF SPECIALTIES IN TERMS OF JOURNALS USING THE DYNAMIC JOURNAL SET OF THE SCI

L. LEYDESDORFF,* SUSAN E. COZZENS**

**Department of Science Dynamics, Nieuwe Achtergracht 166, 1018 WV Amsterdam (The Netherlands)*
***Department of Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, NY 12180–3590 (USA)*

(Received January 6, 1992)

In order to attribute journals to specialties in a dynamic journal set by using aggregated journal-journal citations derived from the *Science Citation Index*, it is necessary to complement the multi-variate analysis of this data with a time-series perspective. This calls for a more analytical approach to the problem of choice among the many possible parameters for clustering. Changes in the disciplinary structure of science are tracked by using the *differences* among the multi-variate analyses for the various years. It is impossible to attribute change systematically to structure, noise, or deviance if these uncertainties are not clearly defined *ex ante*. The study discusses the various choices which have to be made, in both conceptual and methodological terms. In addition to hierarchies among journals, one has to assume heterarchy among journal groups (and their centroids). For comprehensive mapping, a concept of "macro-journals" as representations of a density of points in the multi-dimensional space is defined. Empirical results indicate the feasibility of dynamic journal-journal mapping by using these methods.

The science policy problem

In the 1950s, when the National Science Foundation was first setting up its statistical systems for tracking scientific and technical manpower and other resources of the US, a category scheme for scientific disciplines was developed. It used seven broad disciplinary categories: physical sciences, life sciences, environmental sciences, mathematics, engineering, psychology, and social sciences. Within the broad categories, several different sets of "detailed fields of science" are used. "Detailed fields" are occasionally added, but almost never dropped. Major structural shifts in the sciences are introduced into this scheme only with difficulty; take for example the emergence of the computer sciences, a shift which was reflected in some statistical series in the mid–1960s and in other in the mid–1970s.

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

Has the disciplinary structure of the sciences been as static over the last three decades as the stability of this set of categories would suggest? We suspect not. While differentiation within fields may be the major form of structural change in the sciences, there are other common forms as well, such as, merging among existing areas and the emergence of new fields between the boundaries of existing disciplines. Furthermore, the interdependence of the broad disciplinary categories themselves may be changing. Each of these types of change may be the *result* of policy actions (for instance, when needs identified through the policy process pose practical problems which bring the sciences together in new ways). In turn, these changes may suggest *policy responses*: the reorganization of funding programs or the establishment of coordinating mechanisms, for example. For these reasons, it would be of value to depict them.

Journals as indicators of disciplinary organization

In relation to the scientific literature, disciplines are currently operationalized in terms of journal sets. For example, the number of papers in physics is approximated by the number that appear in journals which have been classified as physics journals, plus an estimate based on the proportion of physics articles in some multidisciplinary journals (e.g. *Science*). The classification of journals was originally done in the early 1970s through a combination of subjective assignment and the examination of cross-citation patterns among journals.[1] The journal sets, at aggregate and disciplinary levels, are held constant over the years to facilitate calculation of all of the literature-based indicator series.*

The contrast between the dynamic structure of science and these constant journal sets presents specific problems for bibliometric indicators of national participation in the various fields of science. It has been shown for science in the aggregate that over time, the shares of literature produced by certain major scientific countries decline when the journal set is held constant, but increases when the full dynamic journal set of the *Science Citation Index* is used.[2-5, 6-9] (The full *SCI* shows a turnover rate in journals of about 7 percent annually.) One possible cause of this discrepancy is that the older journals which are maintained in the constant journal set may gradually

*In 1981, the 1973-journal set was expanded, and the new journals were categorized according to the existing scheme; but there was no major overhaul of the placement of journals into categories nor of the categories themselves. Many relative indicators (e.g. citation ratios) have about the same value in the 1973 and 1981 journal sets, although publication counts are obviously much larger in the 1981 set.

internationalize, while the new journals which the *SCI* adds each year may show higher levels of participation from the world's leading scientific nations. If this is true at the aggregate level, it is likely to be true at the level of disciplines, subfields, and specialties as well.

Thus for two purposes – the general one of describing science accurately and the specific one of measuring national performance – it is of interest to track changes in the disciplinary structure of science.

The data

The data we report here are drawn from the databases of the Institute for Scientific Information (ISI). When we acquired the data (1989), ISI processed approximately 11,000 journal titles,* of which about 3,100 appeared in the *Science Citation Index* and 1,200 in the *Social Science Citation Index* and in the *Arts and Humanities Citation Index*. Additionally, citation data were stored for the *Compumath Citation Index*, which contains 400 of the world's leading computer science, mathematics, statistics, and operations research journals. Journals which were processed only for *Current Contents* (*CC*) or the *Automatic Subject Citation Alert* (*ASCA*) were excluded from our analysis, since they were not systematically processed for citations.

Our data consist of listings of all journals processed in 1984, 1986, and 1988, and the corresponding total journal-journal citations contained in the journal-citation file. During processing, we introduced only a threshold of 0.5% contribution to either a journal's total citing *or* cited environment. On the one hand, the 0.5% cut-off level limits a journal's direct citation environment necessarily to a maximum of 200 journals. However, because of the well-known skewedness of the underlying distributions, one achieves usually with this low-level threshold, the handier size of approximately 50 journals, and thus a 50 times 50 matrix.** On the other hand, discarding links which do not surmount a minimum cut-off level removes a lot of noise. Note that the 0.5% threshold is applied equally to larger and smaller journals, and therefore smaller journals which are processed by ISI, are not removed by this threshold.

---

*Approximately 11,000 journal titles are actually processed on the tape, but only 7,000 are systematically complete.

**Originally, we experimented with a threshold of one percent in order to keep the maximum number of journals drawn into each analysis under hundred, but for the noted reason this can be reached with the lower-level threshold of 0.5% as well.

Processing was done under the regime of SPSS-X under VM/CMS at an IBM-3090 Mainframe of the University of Amsterdam. Subsequently, we downloaded the data to PCs for further analysis. Table 1 gives an overview of the number of journals, citations, etc., involved, and the effects of processing.

Table 1
Statistics on data and processing

Processing of journals	1984	1986	1988
number of journals	7570	7616	7926[1]
eliminated from the analysis	2227	2265	2632[2]
not processed citing	2085	2106	2480
not cited	590	586	671
both of the above	448	427	519
included journals	5343	5351	5294
total cited	6,682,421	7,464,197	8,045,463
total citing	6,958,803	7,724,647	8,442,514
Processing of records	**1984**	**1986**	**1988**
number of cross-journal records	805,465	851,339	886,373
threshold 0.5%	317,567	322,894	323,341
mismatch in journal abbreviation	10,340	6,607	1,082[3]
valid cases	307,227	316,287	322,259
total citations	5,745,615	6,387,604	6,964,824

[1]ISI listed 7721 journals in 1984, 7616 journals in 1986, and 8247 in 1988. However, in the journal list file some of the journals were included twice, and some of the abbreviations in the citation file were not included in the journal list file. Whenever in doubt, we used a conservative strategy and discarded data.

[2]Cases in which either the total "cited" or the total "citing" is equal to zero, are eliminated here.

[3]Cases are eliminated from the analysis if the abbreviation given for the citing journal does not refer to a full journal name unambiguously and automatically.

While for information retrieval purposes the division of the database into the various printed indexes may be pragmatic, for science policy analysis purposes we which not to make *ex ante* decisions on "how to cut the cake into pieces," since one is often interested in developments which take place between various disciplines. Since we wished to work as much as possible on the basis of the information contained in the data, we did not make any *ex ante* distinctions among journals (e.g., review journals), nor did we align high values on the diagonal to expected values.** In these respects, the data are similar to those in the first data column of the printed editions of the *Journal Citation Reports*.

Methodologies

Citations are a network indicator. By citing another article an author becomes "linked" to the archived literature. Therefore, we can use methods developed in formal network analysis to study citation patterns among sets of articles in various statistical terms.

A variety of techniques for analyzing journal-journal citation relationships have been reported in the literature, including influence mapping, journal clustering, and block modelling.[12-19] Many of these techniques have achieved results which are satisfactory in the sense that they produce depictions of structures which are congruent with expert knowledge of the fields analyzed. None of these techniques, however, has emerged as dominant.

In our opinion, an explanation of the differences in the results of the various forms of multi-variate analysis and mapping requires an analytical approach. If one wishes to extend the analysis to a comparison of the results of multi-variate analysis in different years, a more analytical approach to the problem of choice among the many possible parameters becomes unavoidable.

Firstly, assessment in terms of usefulness in science policy or in terms of recognizability by practicing scientists ("validation") does not give sufficient guidance in the choice of similarity criteria and clustering algorithms, since (1) this involves not a simple choice among a few alternatives, but a parameter space of, for example, 40 similarity criteria and 20 clustering algorithms; and (2) maps may be useful or recognizable by practicing scientists without being necessarily a correct or best representation of the structure in the data.*

Secondly, the choice of parameters for the multi-variate analysis necessarily implies a focus on certain parts of the variance, and the definition of other parts of

---

**Normalization of deviant diagonal values in terms of the distributional characteristics would have improved upon our results, since irregularities in the raw data tend to depress the correlation among citation patterns. See also Refs 10 and 11.

*For example, ISI's *World Atlas of Science* is assessed as very useful for bibliographic and policy purposes, although the clustering is based on single linkage clustering of Euclidean distances among cocitation data. We will argue below for the use of other parameters. Therefore, our results will *a priori* not be comparable to those from cocitation clustering. See also Refs 17 and 20.

the variance as noise or unique variance. If one subsequently subtracts the multivariate solution for one year from that for another year, it becomes impossible to attribute change systematically to structure, noise, or deviance if these uncertainties are not declared *ex ante*. Thus, the dynamic mapping of multi-variate structures requires further reflection on the appropriateness of the available methodologies.

Various parameters for multi-variate analysis

Available algorithms for distinguishing subgroups in networks can be segregated into three classes: factor analysis, or eigenvector, algorithms; graph theoretic algorithms; and spatial algorithms.[21]

Factor analysis determines the number and nature of the underlying dimensions of variation among many variables. It can be used in an exploratory way or for hypothesis testing. For example, if one assumes that some journals share patterns of subject relations with other journals, this assumption can be tested, using citations to operationalize the subject relationships. In a factor analytic solution, one can specify the extent to which a variable (in this case, the citation patterns of a journal) represents an underlying dimension (the so-called "communality" of the variable for the factor), and the extent to which its variance is unique. If we are able to show that communalities among variables can indeed be reduced to underlying structures, variables which load heavily on only one factor can be taken as indicators for that dimension.

In contrast to factor analysis, in graph theoretic algorithms, the focus is on dyadic bonds, and the algorithms seek to localize strong or weak components in a network.[22] This type of analysis is particularly effective in distinguishing subgroups or cliques in a network, even if this network is rather large, since the focus is not on the whole pattern of relations.

Spatial algorithms seek to represent observed relations among actors in a relational space: intense relations create close proximities, and vice versa. As *Burt* has noted,[22] the advantage of these algorithms over the graph theoretical ones, namely that they require no absolute criterion for identifying cohesive bonds, is also a drawback, since they will produce clusters no matter how intense the relations are. However, graph theoretic algorithms and spatial algorithms have in common that they analyze *relations* among the components of a network, while the factor analytic approach exhibits *positions* of units with reference to latent dimensions.

Note that relations are attributes of units, while eigenvectors are attributes of the network. In a first order cybernetics lower level units generate the architecture of the network, while the eigenvectors are units of analysis in a second order cybernetics. In this study, we use the aggregated citation relations among journals to indicate these latent units of the network since we assume these to be organized along cognitive lines.

Various parameters for time-series analysis

The time axis poses a special problem in analyzing complex data. The data exhibits change and stability, both phenomenologically and in terms of underlying multi-variate structures. On the one hand, that part of the data which exhibits stability causes auto-correlation in the data. In terms of underlying structures this auto-correlation corresponds with the reproduction of structure in different years, i.e., with the self-referentiality of the structure. On the other hand, change can be a consequence of (non-stationary) trends or drift. In scientometrics, of course, we are particularly interested in explaining changes caused by science policy efforts at a structural level. Here again, a set of dynamic mechanisms can be hypothesised, e.g., "emergence," "differentiation," and "goal-referentiality" with respect to priorities.

There are unresolved methodological problems in the integration of multi-variate analysis and time-series analysis (e.g., Ref. 23, p. 727). Time series analysis has been developed for application to a single variable, and the more complex information produced by multi-variate analysis of a point in time is easily lost when the time dimension is built in. For our data, one of us has proposed using measures drawn from information theory in order to address this problem.*

However, graph analytical techniques make it possible to distinguish strong groups in large datasets, while factor analysis is particularly apt to reduce larger numbers of variables into a smaller set of underlying common dimensions. These two tasks are major ones which we face for each year separately. The objective of pursuing a dynamic analysis gives guidance with respect to the selection of parameters in the static analysis: the primary question is whether the multi-variate analysis can provide us with a baseline in order to assess change systematically. What "fixed points" can be compared, *ceteris paribus?*

---

*Information theoretical measures are non-parametrical and therefore are not dependent on mathematical idealizations. This allows us to use them on short time series of data. See also Refs. 24 and 25.

These "fixed points" do not necessarily have to be adequate descriptions of the relevant units of analysis for case studies, since, for example, the case may cover more than one such point. The delineation of units of analysis in the static analysis and the problems of aggregation raise different questions.

Choice of parameters

Let us reflect for a while on the nature of the multi-dimensional space which is represented by the grand matrix of 5000 or so journals which cite one another. This matrix contains far more than 10^7 cells, of which far fewer than 10^6 have a value larger than zero. The distribution is known to be heavily skewed so that we may expect that most of the non-zero cells will contain ones.

If we apply the noted threshold for the citations of each journal at the level of the minimum of either 0.5% of its total citing or 0.5% of its total cited, the number of cells with a non-missing value drops to about 300,000. This is less than one percent of the cells in the grand matrix; the grand matrix is extremely sparse, and consequently the multi-dimensional space which it represents is virtually empty.

However, the distributions are not only heavily skewed, they are also very specific. The points are not scattered all over the multi-dimensional space, but occur in clouds. These clouds are not necessarily more dense towards the centre, and therefore we will use the concept centroid for the mathematical definition of the centre of each density. (Incidentally, we find that journals of an interdisciplinary nature sometimes link otherwise discrete clouds in another dimension. However, these journals may also manifest themselves in groups. We may, therefore, find journals with specific intermediate functions between two or more clouds in an otherwise empty landscape.)

With respect to measuring change dynamically, we can now reformulate the conceptual questions in the previous section in the following methodological terms:

1. Which formal criteria should we use to delineate clouds in each year?
2. Can one operationalize the centroids of the clouds so that they can be used as reference points for change from year to year?
3. Can we account for changes in the data using these delineated clouds and this operational definition of their centroids as systems of reference?

The first question corresponds to the question of how disciplines, subdisciplines and specialties develop; the second with the question of how to construct a multi-variate baseline for the time-series analysis; and the third addresses the question of interdisciplinary changes.

Alternating the relational and positional approach

The attribution of centroids to the clouds requires a factorial design, or, in terms of network analysis, a positional approach. A latent factor is hypothesized which will then exhibit a (partial) correlation with individual cases. The cases then have a *position* in relation to this factor.

Thus, the definition of the centroid of the cloud is operationalized in terms of a factor.* Factors, however, can only be defined for a specific factor domain. Therefore, we have to find a more formal way to delineate the *relevant* environment (in our case, this is the citation environment). Note that this delineation of the domain is not a positional, but a relational problem: the weaker the relationships between points in one particular cloud and those in others, the less their influence on the position of that cloud's centroid. However, since the space is multi-dimensional, a large distance in some dimensions does not preclude vicinity in others.

One relevant definition of the environment is the overall database, which is by nature a selection of the universe of possible citations among journals. However, as noted above, there are good arguments for further restricting the relevant environment, since the multi-dimensional space of the total set is virtually empty. (For example, one does not expect a journal specializing in solid state physics to have a major citation relation with a journal of organizational sociology.) However, the definition of the relevant environment by means of the relational approach is also the definition of the domain in the positional approach, and therefore the two solutions are not independent.

We propose to solve this problem by the method of successive approximation. If one first delineates the relevant environment for a particular area (i.e., a group of journals) or for a specific journal, how then after extraction and rotation of factors is this domain to be described in terms of underlying dimensions? Would the results of the analysis lead to a redefinition of the criteria for selecting the domain?

In the case of one journal, one can define the relevant citation environment as all journals which maintain either a cited or a citing relation with this journal. (As noted above, we introduced a 0.5% threshold as a minimum in either dimension.) For this

*See, for the statistical model behind factor analysis, e.g. Ref. 26. See also Ref. 27, pp. B40–69.

domain a citation matrix can be constructed which can then be factor analyzed. Since the factor analysis is based on the Pearson correlation matrix, size effects among variables (i.e., citing patterns of journals) are normalized.

The entrance journal necessarily loads differently on various factors, since the distributions are skewed. The highest factor loading indicates the primary group to which the initial entrance journal belongs. This journal or another journal will have the highest loading on this factor.* In the latter case, we repeat the analysis with the other journal as the entrance to the analysis, and so on, until the entrance journal is also the highest loading journal on the factor which indicates the grouping. If this is the case, we call this journal a "central tendency journal".

One can attribute to each central tendency journal a set of journals which point to that journal as its central tendency. Note furthermore that all journals in the database are either central tendency journals or point to another journal which in turn can be either a central tendency journal or yet another indicator, etc. This definition leaves room for loop-like structures: journals which refer to one another as better indicators of the underlying commonality. In this case, we will have to take this *group* of journals as the best representation of the centroid of the cloud.

Since in the case of a loop-like structure, the centroid is indicated by a group of journals, correspondingly more environments are drawn into the analysis. Therefore, one has the option of aggregating the results either with a Boolean AND or with a Boolean OR. In this analysis, we are interested in the core group of a cloud as a system of reference for dynamic change, so we use the AND-operator. Otherwise, journals which are not even part of the relevant environment of one central tendency journal can in principle be included in the core group by using another journal from the same loop. The core group is then no longer strictly definable.

*In principle, there is a small probability that two variables exhibit the same highest factor loading, and again a small possibility that taking either of them in the iterative procedure leads to different central tendency journals. In our study this never occurred. However, the procedure in terms of macro-journals to be described below would systematically indicate two macro-journals in this case.

Table 2

Factor loadings of biochemistry journals in 1984, 1986, and 1988.

Entry journal is *Arch. Biochem. Biophys*[1]; the common set is boldfaced

Journal	1984	1986	1988
Agr Biol Chem Tokyo	(.36453)[2]	.39737	.39495
Anal Biochem	.80158	.71659	.75176
Annu Rev Biochem	(.56347)[3]	.68423	.75413
Arch Biochem Biophys	.94451	.94109	.96594
Biochem Biophys Res Co	.83853	.78067	.82494
Biochem Int	.96344	.95564	.96117
Biochem J	.83382	.87098	.87415
Biochem Pharmacol	.68241	.65137	.69662
Biochem Soc T	(.—)	.76048	(.—)[4]
Biochemistry-US	.81683	.79985	.80487
Biochim Biophys Acta	.84694	.86314	.85851
Comp Biochem Phys B	.80580	.77903	.82605
Eur J Biochem	.88254	.89467	.88543
FASEB J	(.27932)[5]	(.32813)[5]	.72292[6]
FEBS Lett	.86024	.84092	.84951
H-S Z Physiol Chem	.72139	(.*****)	(.*****)[7]
Int J Biochem	.94239	.94424	.96086
J Biochem-Tokyo	.83485	.86535	.87804
J Biol Chem	.86924	.86572	.87576
J Neurochem	(.36811)[8]	.43217	(.42808)[9]
Method Enzymol	.87254	.82018	.95449
Mol Cell Biochem	.83183	(.—)	(.—)[4]
Mol Pharmacol	(.—)[4]	.71032	.74764

[1]In 1984 *Arch Biochem Biophys* formed a loop-like structure with *Biochem Int* and *Int J Biochem*; in 1986, with *Biochem Int* only.

[2]In this year, *Agr Biol Chem Tokyo* has a marginally higher loading on a factor with the *J Am Chem Soc*.

[3]The factor loading on the "molecular biology" cluster is .77860.

[4]The indication (.—) means that in this case the journal is not drawn into the relevant citation environment using the 0.5% threshold.

[5]In 1984 and 1986, *Federation Proceedings* loads highest (> 0.70) on a factor with the *Am J Physiol* as central tendency journal (> 0.85).

[6]In 1984 and 1986, this journal is included under the title *Federation Proceedings* of the Federation of American Societies for Experimental Biology (FASEB).

[7]The name of *H-S Z Physiol Chem* was changed into *Biol Chem H-S* with the net effect that it is no longer processed under the German name under which it, however, remains to be cited. By using the English name it is poorly "cited", and relatively small "citing".

[8]Highest factor loading on a factor which is led by *Cancer Res.*

[9]Highest factor loading on a factor which is led by *Toxicol Appl Pharmacol.*

Table 2 shows the factor loadings for *biochemistry*, and *Arch. Biochem. Biophys.* as entrance journal. *Arch. Biochem. Biophys.* is the single journal which is among the central tendency journals in all three years. Obviously, this is a well-defined set which exhibits stability over the years under study.

Table 3
Different views of the *molecular biology* density area in 1988 given by using different CTJs as entry journals to the analysis. (Communalities are boldfaced)

Journal	a	b	c	d	e
Annu Rev Genet (c)	*	*	*	*	*
Bioessays					*
Bio-Technol			*		
Cancer Surv		*			
Cell (e)	*	*	*	*	*
Curr Genet			*		*
Curr Top Microbiol	*	*			
Dev Biol		*			
DNA-J Molec Cell Bio		*			*
Embo J (d)	*	*	*	*	*
Exp Cell Res		*			
Gene	*	*	*		*
Gene Anal Tech		*			
Gene Dev	*	*	*	*	*
Genetics				*	*
Genetika +			*		
Immunogenetics	*				
J Cell Biol			*		
J Mol Biol	*	*	*	*	*
J Mol Evol			*		
J Virol	*		*		
Mol Biol Evol			*		
Mol Cell Biol	*	*	*	*	*
Nature	*	*	*	*	*
Neuron				*	*
Nucleic Acids Res	*	*	*	*	*
Oncogene (b)	*	*	*	*	*
Oncogene Res		*			
Philos T Roy Soc B	*		*		
Plant Mol Biol			*	*	
PNAS	*	*	*	*	*
Science	*	*	*	*	*
Somat Cell Molec Gen	*	*	*		
Trends Genet (a)	*	*	*	*	*

a = *Trends Genet* is taken as entry journal.
etc.

Table 3 summarizes the results for *molecular biology* in 1988. In this case, various journals refer to one another as central tendency journals in all years. Over the period under study (1984–1988) the group obviously merges with a group of *genetics* journals, as is evident from a comparison of results for 1984 and 1986 (see Table 4). Some *genetics* journals which did not belong to the relevant environment of this group in 1984 have come to be among the best indicators of its central tendency in 1988.

Table 4
Different views of the *molecular biology* density area given by using different CTJs as entry journals to the analysis, for 1984 and 1986, respectively. (Communalities are boldfaced)

Journal	1984 a	b	c	1986 a'	b'	c'
Annu Rev Biochem	*					
Annu Rev Genet (b')	*	*		*	*	
Bioessays				*		
Chem Scripta						*
Cell (a, a')	*	*	*	*	*	*
Curr Genet					*	
DNA			*			
DNA-J Molec Biol						*
Embo J (b)	*	*	*	*	*	*
Gene	*	*		*	*	
Genetics				*		
J Cell Biol						*
J Mol Biol	*	*	*	*	*	
J Mol Evol			*			*
Mol Cell Biol	*	*	*	*	*	*
Mol Gen Genet				*	*	
Nature	*	*	*	*	*	*
Nucleic Acids Res (c, c')	*	*	*	*	*	*
P Natl Acad Sci USA	*	*	*	*	*	*
Philos T Roy Soc B	*	*				
Plant Mol Biol		*		*	*	
Progr Nucleic Acid Re (d)	*		*			
Science	*		*	*	*	*
Somat Cell Molec Gen	*					
Trends Genet				*	*	

Note that the above attribution of journals to groups, which are defined in terms of their central tendency, is strictly analytical with the primary objective to provide us with reference points for indicating change from year to year. An alternative strategy which would be particularly apt for a case study approach that primarily looks for the best representation of the core group (and not for the best representation of the centroid) involves taking the whole group of journals which results from the first analysis with the highest loading on the relevant factor, and define a relevant environment for this *group*. Then, orienting oneself with the help of the entrance journals in iterative rounds, one may be able to define a group of journals which, when used as an entrance group, also gives itself as a single reference group. Alternatively, entrance groups may refer to each other in loop-like structures. This is exhibited in Table 5 for the case of *Cell* as the original entrance journal in the 1988 dataset.*

Table 5
Iterations with sets of journals (1988; *Cell* is entry journal)

Journal	*Cell*[1]	1	2	3[2]	4	5	6	7
Annu Rev Genet	*	*	*	*	*	*	*	*
Bioessays	*							
Cell	*	*	#	*	*	*	*	*
Curr Genet	*	*	*		*		*	
Development				*				
Dev Biol			#			*		*
DNA-J Molec Cell Biol	*	*	*					
Embo J	*	*	*	*	*	*	*	*
Gene	*	*	*	*	*	*	*	*
Gene Dev	*	*	#	*	*	*	*	*
Genetics	*	*	*	*	*	*	*	*
J Cell Biol			#			*		*
J Mol Biol	*	*	*	*	*	*	*	*
J Mol Evol			*					
Mol Gen Genet		*	*		*		*	
Mol Cell Biol	*	*	*	*	*	*	*	*
Nature	*	*	#	*	*	*	*	*
Neuron	*							
Nucleic Acids Res	*	*	*	*	*	*	*	*
Oncogene	*	*	#	*	*	*	*	*
P Natl Acad Sci USA	*	*	*	*	*	*	*	*
Plant mol Biol		*	*		*		*	
Science	*	*	#	*	*	*	*	*
Trends Genet	*	*	#	*	*	*	*	*

[1]Based on *Cell* as an entry journal. Boldfacing indicates journals which were previously marked as "core journals" of "molecular biology" in 1988.
[2]"*" indicates Factor II, "#" Factor III. (Factor I is always *biochemistry*.) The iteration is continued with Factor III, since *Cell* has highest loading on this factor.

*These results are produced by a programme which was written by *Wamelink*.[28]

The problem of aggregation

At the level of aggregated citations among individual journals, this approach to the data provides us with a detailed picture of clouds with centroids. We can indicate the centroids by journals or groups of journals. However, for many purposes, one is interested in a higher level of aggregation, especially since the clouds do not correspond necessarily to common definitions of disciplinary groupings. How can we arrive at a more comprehensive mapping of the structure, and dynamics, of science?

Obviously, one may wonder whether a top-down approach would not be more fruitful to arrive at a comprehensive mapping than the bottom-up approach which we pursued in this project. However in the top journals the disciplinary delineations are more blurred than in the lower ranks. By using a hierarchical clustering algorithm, one can not define positions.[21, 22, 17] In other words: journal systems for various disciplines and specialties coexist alongside each other, and the relational patterns between, for example, major journals do not necessarily indicate relational patterns between these latent structural units.**

The heterarchical units of science which we described in the previous section are relatively small when the operationalization is in terms of journal-journal citations. Can we use the groupings of the clouds for further aggregation to achieve a more comprehensive mapping?

Here, *Small's* notion of "clustering the clusters" can be helpful.[29] However, if one wishes to cluster in terms of the latent network-structure, the clustering should proceed with reference to the factors which were revealed by the previous round of the analysis, and not with reference to individual journals and their citation-relations. The factors are the higher-order "eigenvectors" of the network-system, while the individual variables are the lower-level vectors which went into the construction of this network. Therefore, the best aggregation is that we can "zoom out," i.e., take each cloud as one point, and then repeat the procedure. The creation of "macro-journals" is thus the aggregation of rows and columns for the groupings found in the previous round. Each macro-journal can conveniently be labeled with the name of the central tendency journal which indicated the center of it.

This aggregation provides us with a matrix which is about one-tenth of the size in both dimensions (cited/citing) as the original matrix representation for the overall

---

**Additionally, there may be processes of de-differentiation at higher levels (e.g., *Nature* and *Science*) and at the lower end, for example, in technological applications and at the clinical bed-side.

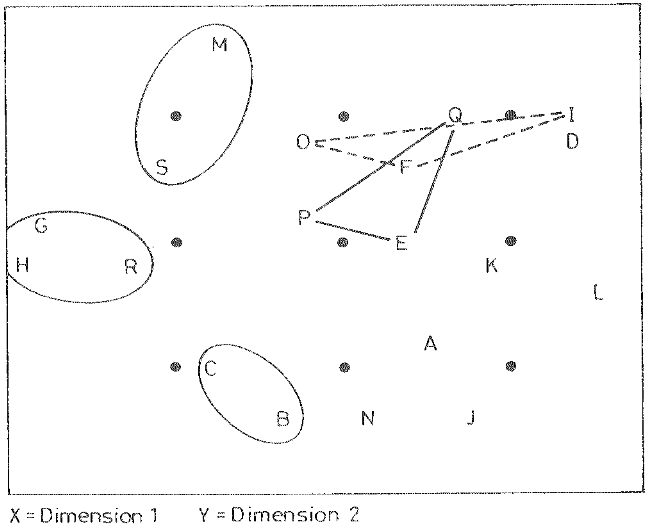

Fig. 1. Factor analysis and MD-SCAL for "macro-journals"; (1988; citing patterns).

Central tendency journal of macro journal	Factor number	Loading
R. Z Phys C Part Fields	I	.88528
H. Gen Relat Gravit		.82388
G. Few-Body Syst		.66770
Q. Surf Sci	II	.76559
P. Solid State Phys		.75313
E. Appl Phys Lett		.65743
F. Chem Phys	III	.80524
I. J Colloid Interf Sci		.65181
O. Phys Rev A		.64643
M. Kvantovaya Elektron +	IV	.85219
S. ZH Eksp Teor Fiz +		.73219
D. Appl Catal	V	-.83268
K. J Phys E Sci Instrum	VI	.82710
C. Annu Rev Astron Astr	VII	-.72679
B. Ann Geophys		-.52810
J. J Nucl Mater	(VII)	.48242
L. J Rheol	VIII	-.90819
A. Acta Metall	IX	-.82334
N. Phys Fluids	(IX)	.60459

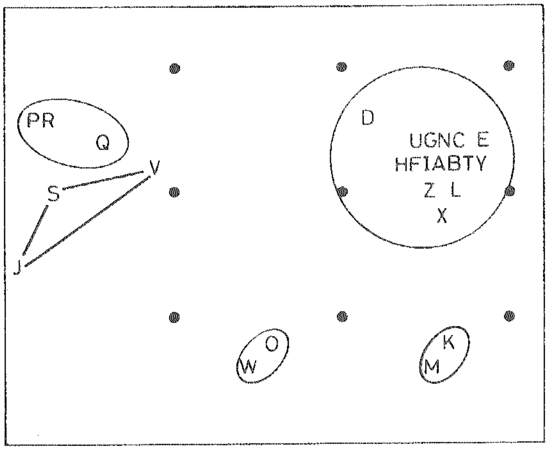

Fig. 2. Factor-analysis and MD-SCAL for ANNU REV ASTRON ASTR, citing patterns (1988; threshold = 0.50%)

Journal name abbreviation	Factor number	Loading
A. Annu Rev Astron Astr	I	.98591
F. Astrophys J		.98378
G. Astrophys J Suppl S		.97662
H. Astrophys Lett Comm		.97656
T. Publ Astron Soc JPN		.96358
U. Publ Astron Soc Pac		.96217
Z. Space Sci Rev		.95334
I. Astrophys Space Sci		.95129
B. Astron Astrophys		.93812
C. Astron J		.93616
N. Mon Not R Astron Soc		.93400
E. Astron ZH+		.90781
Y. Sov Astron Lett+		.90381
L. IAU Symp		.86547
X. Sol Phys		.66044
D. Astron Nachr		.61950
P. Phys Lett B	II	.92639
R. Phys Rev D		.91925
Q. Phys Rep		.86616
S. Phys Rev Lett	III	.84144
J. Aust J Phys		.74742
V. Rep Prog Phys		.74072
O. Nature	IV	.95217
W. Science		.92668
M. Icarus	V	.95867
K. Earth Mooin Planets		.85860

database. In this matrix, we can pursue the same analysis as in the previous round at a higher level of aggregation, while maintaining the positional perspective from the previous round. "Organic chemistry" and "inorganic chemistry," both separate self-referential units if analyzed in terms of journals, for example, now come together in a more comprehensive mapping of chemistry. Figure 1 gives such a mapping for nineteen macro-journals in physics in 1988. Note the presence of a Soviet (physics) cluster, which illustrates that our initial assumption that journal-journal citation relations indicate cognitive delineations, is not always true.* Figure 2 zooms into the fine-structure at journal C in Fig. 1, i.e., *Annu Rev Astron Astr.*

Conclusion and discussion

In this study, we have made some methodological distinctions of relevance to the dynamic mapping of science using aggregated journal-journal citations.

First, we had to revise our initial assumption, which was based on the literature (e.g., Ref. 12) that the journals must be ordered in a hierarchy. Of course, it is always possible to order journals in a hierarchy, for example, in terms of their impact factors, or in two dimension, by using influence mapping.[1] However, in terms of journal-journal citation relations journal sets can also be displayed as discrete clouds of points in an otherwise empty multi-dimensional space. This non-relatedness embodies another, non-hierarchical structure. From this perspective, those journals which are more general are not "overarching," but rather points on the outer surfaces of respective clouds. However, in most cases they also belong distinctly to one or a few of these clouds. Like the interdisciplinary journals, they function as "gateways" to specific clouds for specific citation environments. However, they do so like cathedrals which are visible from a distance in an otherwise empty landscape.

Pursuing the metaphor of spatial organization in the three dimensions of a landscape, one may say that the cathedral also serves the surrounding neighbourhoods, and is therefore not the most pronounced representation of the one in which it stand. The centre of town may lie on the next square, as was common in a mediaeval township. However, from a distance one may either limit one's vision of

*The "cited" dimension, i.e., after transposition of the matrix, is much more sensitive to these linguistically and geographically delineated factors, since local journals are more exclusively cited by authors who write in the national language. However, authors who publish locally often cite the international literature.

the city to its skyline, or gain access to the city over the road through the various hamlets and the hostels of interdisciplinarity along it. Additionally, the domes and cathedrals are organized in a hierarchical structure of bishoprics and archbishoprics.*

*However, the eigenvectors of the network are heterarchical and not hierarchical.*** Therefore, in order to map the disciplinary organization, it does not matter so much at which point one enters the network. One may wish to enter it at a discrete point in order to pursue a case-study, or one may wish to enter it at one cathedral or another in order to map a whole "discipline" of science. However, also with the latter objective, as soon as one enters the network the local density of specialty journals becomes central to any further analysis, and one gets qualitatively the same picture which one gets of the surrounding villages and cities as the one obtained (albeit with more sophistication and greater precision) when one enters from the central tendency journal immediately.

As we have indicated, one may construct macro-journals and their relations from these finer grained nuclei in the network, in order to grasp the overall structure. However, conversely, there is no separate level at which "major journals" hold more specific information about disciplinary delineations than their respective, more specialized immediate neighbours.

Among the discrete densities there are several forms of transition. In landscape terms, there may be the "suburban" transition and the "rural/urban" one. However, the "suburban" arrangement where one has to dwell on the periphery, and does not know precisely when one crosses boundaries, is the exception. In this case, one witnesses mainly a *lack* of structure. This lack of structure may be temporary. However, when there is no structure to be found, there is also no means of measuring change or other dynamic developments. Fortunately, the "rural/urban" format is more common, and more interesting for the purposes of science policy: here there is a nucleus, however weak; will it grow or disappear?

---

*The origins of the words "hierarchy" and "heterarchy" refer to the organization of the Catholic and Protestant religion, respectively. Since the Pope is considered to be the Vicar of Christ, His Holiness is on top of a hierarchy which covers the whole world ("kat' holèn gèn"). In Protestantism, each individual is equal before God, and therefore the world is contained as a latent structure in a network of relations. At various places, Weber, Parsons, Merton, and Luhmann noted the importance of the Protestant ethos for the emergence of modernity in general, and the differentiation of science in particular.

**Of course, eigenvectors can be rank ordered.

Even in emerging "hot" areas of change, i.e., where the system is under pressure from new scientific, technological or social processes of change, the new journals situate themselves in terms of their citing patterns with respect to one or the other of these dense groups, either immediately or after only a short period of time. (In their "cited" patterns they may have been noted by other neighbouring densities also, and yet to such an extent that they formed an independent, deviant pattern of citations in this respect. We will elaborate in another study[30, 31] on the indicators of "emergence" which may be derived from this.)

Within each dense region of the multidimensional space there is a predominant pattern of citations by articles from the same journal of from journals belonging to the same group. This group is often small. The world of science in terms of its journal *literature* is heavily differentiated. The normal process is one of further differentiation: through a potential period of transition, a new core set is established, and the stable pattern of "self-referentiality" both within this core and in its surrounding groupings resumes.[32]

It is important to signal that the organization of scientific journal literature in these terms is even more fine-grained than we are accustomed to assume when we think of the scientific enterprise in terms of missions and institutions. While the latter are sometimes "interdisciplinary" in a more organizational sense, the articles published in the journals do not necessarily have to follow the same scheme. For example, when we pursued the analysis for "environmental chemistry," we found a discrete journal set for "environmental toxicology," for research related to issues of "air pollution," and for issues related to "water research." While we may imagine members of this community publishing in all these journal sets, as journal sets they constitute each a central core which exhibits an aggregated citation pattern with significant specificity. Similarly, when we studied our own field of "science studies," we found a much more pronounced distinction between *Scientometrics* and *Research Policy* on the one side, and journals like *Social Studies of Science* on the other, than we had expected, given the terms in which we wish to understand our own enterprise. Obviously, the issue of the self-understanding and perception of the relevant horizons of scientific development by practicing researchers does not necessarily correspond with the citation patterns which indicate the crystallization of these practices in the scientific literature.

Does this conclusion defeat the purpose of this project? We think not. The self-understanding of practicing scientists is by its very nature contingent upon local conditions, and therefore very much at variance across institutional and national settings. In order to create a baseline against which to evaluate change in the various dimensions, we need dimensions which can be measured more objectively, and which can precisely be described in terms of (iterative) procedures.

*

The authors acknowledge the contributions of *Frank Wamelink* and *Suzanne Bakker* to their research; and the support of the National Science Foundation (Science Indicators Unit) of the USA (grant nr. 8810197). We also thank *Henry Small* and *Margaret Sydleck* of the Institute of Scientific Information for help with the data.

References

1. F. NARIN, *Evaluative Bibliometrics*, Computer Horizons Inc., Cherry Hill, 1976, 190–203.
2. L. LEYDESDORFF, Problem with the 'measurement' of national scientific performance, *Science & Public Policy*, 15 (1988) 149–152.
3. J. ANDERSON, P.M.D. COLLINS, J. IRVINE, P.A. ISARD, B.R. MARTIN, F. NARIN, K. STEVENS, Online approaches to measuring national scientific output: a cautionary tale, *Science & Public Policy*, 15 (1988) 153–161.
4. L. LEYDESDORFF, The Science Citation Index and the measurement of national performance in terms of numbers of scientific publications, *Scientometrics*, 17 (1989) 111–20.
5. T. BRAUN, W. GLÄNZEL, A. SCHUBERT, Assessing assessments of British science: some facts and figures to accept or to decline, *Scientometrics*, 15 (1989) 165–70.
6. B.R. MARTIN, The bibliometric assessments of UK scientific performance. A reply to Braun, Glänzel and Schubert, *Scientometrics*, 20 (1991) 333–58.
7. T. BRAUN, W. GLÄNZEL, A. SCHUBERT, The bibliometric assessment of UK scientific performance. Some comments of Martin's reply, *Scientometrics*, 20 (1991) 359–62.
8. L. LEYDESDORFF, On the "Scientometric Decline" of British science. One additional graph in response to Ben Martin, *Scientometrics*, 20 (1991) 363–8.
9. T. KEYLEY, Government-funded academic science is a consumer good, not a producer good: A comparative reassessment of Britain's scientific and technological achievements since 1794 and a comment on the bibliometry of B. Martin and J. Irvine, *Scientometrics*, 20 (1990) 369–94.
10. D. DE SOLLA PRICE, The analysis of square matrices of scientometric transactions, *Scientometrics*, 3 (1981) 55–63.
11. E. NOMA, An improved method for analyzing square scientometric transaction matrices, *Scientometrics*, 4 (1982) 297–316.
12. M.P. CARPENTER, F. NARIN, Clustering of scientific journals, *Journal of the American Society of Information Science*, 24 (1973) 425–35.
13. F. NARIN, *Evaluative Bibliometrics*, Computer Horizons Inc., Cherry Hill, NJ, 1976, pp. 185 ff.
14. P. DOREIAN, T.J. FARARO, Structural equivalence in a journal network, *Journal of the American Society of Information Science*, 35 (1985) 28–37.
15. L. LEYDESDORFF, The development of frames of references, *Scientometrics*, 9 (1986) 103–25.
16. P. DOREIAN, A revised measure of standing of journals in stratified networks, *Scientometrics*, 11 (1987) 63–72.
17. L. LEYDESDORFF, Various methods for the mapping of science, *Scientometrics*, 11 (1987) 291–320.
18. R. TIJSSEN, J. DE LEEUW, A.F.J. VAN RAAN, Quasi-correspondence analysis on square scientometric transaction matrices, *Scientometrics*, 11 (1987) 347–61.
19. R. TODOROV, W. GLÄNZEL, Journal citation measures: a concise review, *Journal of Information Science*, 4 (1988) 47–56.
20. L. LEYDESDORFF, Some methodological guidelines for the interpretation of scientometric mappings, *R & D Evaluation Newsletter*, (1989) Nr. 2, 4–7.

21. R. BURT, Cohesion versus structural equivalence as a basis for network subgroups, In: R. BURT, M.J. MINOR, *Applied Network Analysis*, Beverly Hills, Sage, 1983, 276f.
22. R. BURT, *Toward a Structural Theory of Action*, Academic Press, London, 1982.
23. R.J.W. TIJSSEN, J. DE LEEUW, MVA methods in bibliometric studies, In: A.F.J. VAN RAAN (Ed.), *Handbook of Quantitative Science and Technology Studies*, Elsevier, Amsterdam.
24. L. LEYDESDORFF, The prediction of science indicators using information theory, *Scientometrics*, 19 (1990) 297–324.
25. L. LEYDESDORFF, The static and dynamic analysis of network data using information theory, *Social Networks*, 13 (1991) 301–45.
26. J.-O. KIM, C.W. MUELLER, *Factor Analysis, Statistical Methods and Practical Issues*, Sage, Beverly Hills/London, 1978.
27. M.J. NORUSIS, *SPSS/PC+ Advanced Statistics*, SPSS Inc., Chicago, 1986.
28. F. WAMELINK, A quantitative strategy for the delineation of scientific fields: a case study of the emergence of biotechnology (in preparation).
29. H. SMALL, E. SWEENEY, E. GREENLEE, Clustering the Science Citation Index using co-citations II. Mapping science, *Scientometrics*, 8 (1985) 321–40.
30. S.E. COZZENS, L. LEYDESDORFF, Journal systems as macro-indicators of structural change in the sciences, In: A.F.J. VAN RAAN, (Ed.) *Proceedings of the Joint EC/Leiden Conference on Science and Technology Indicators*, DSWO, Leiden (forthcoming).
31. L.LEYDESDORFF, S.E. COZZENS, P. VAN DEN BESSELAAR, Tracking areas of strategic importance by using scientometric journal mappings, Paper presented at the *4S/EASST Joint Conference*, August 1992.
32. L. LEYDESDORFF, Irreversibilities in science and technology networks: An empirical and analytical approach, *Scientometrics*, 24 (1992) 321-357.