Scientometrics, Vol. 31, No. 1 (1994) 59-84

THE GENERATION OF AGGREGATED JOURNAL-JOURNAL CITATION MAPS ON THE BASIS OF THE CD-ROM VERSION OF THE SCIENCE CITATION INDEX

L. LEYDESDORFF

Department of Science and Technology Dynamics,
Nieuwe Achtergracht 166, 1018 WV Amsterdam (The Netherlands)

(Received December 13, 1993)

A method is described for the generation of journal-journal citation maps on the basis of the CD-ROM version of the *Science Citation Index*. Various sources of potential errors in using this data are discussed, and strategies are suggested to counteract these errors. A number of scientometric journal mappings are analyzed in relation to mappings from previous studies which have used tape data and/or data from ISI's *Journal Citation Reports*. The quality of these mappings is compared with the quality of those for previous years in order to demonstrate usefulness of such mappings as indicators for dynamic developments in the sciences.

Introduction

Journal-journal citations – like the ones listed in the *Journal Citation Reports* of the *Science Citation Index* – have long been used as instruments in scientometric mapping [e.g., *Carpenter* and *Narin* (1973)[1]; *Narin* (1976)[10]; *Doreian* and *Fararo* (1985)[3]; *Leydesdorff* (1986)[4]; *Tijssen* et al. (1987)[11]]. More recently, *Leydesdorff* and *Cozzens* (1993)[7] have suggested ways of using these journal mappings to indicate change in science. Aggregated journal-journal citation data is readily available in print or on microfiche, but not in machine-readable format. Although the Institute of Scientific Information produces the *Journal Citation Reports* from tapes, few research centres are able to purchase these tapes for processing on a regular basis.

Does the nowadays widely available CD-ROM version of the *Science Citation Index* provide a means for obtaining the data underlying scientometric mappings in machine-readable format? What are the problems and pitfalls associated with using this version to construct an equivalent to the *Journal Citation Reports*? In this study, I report on the use of data for this purpose from the CD-ROM for the first quarter of 1993. I shall discuss the sources of errors like misspellings, and suggest ways of handling these problems. Additionally, I am able to draw some conclusions on the

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

basis of comparisons over time, since I analyzed the tape data for 1988 in the context of another project.*

In principle, the on-line version of the *Science Citation Index* as installed on host-computers like DIALOG and DIMDI (*SciSearch*) provides us with another source of electronic citation data. Since its update in 1988, *SciSearch* enables the researcher to search on-line for, among other things, journal titles within the cited references [cf. *Moed* (1988)[9]]. However, information retrieval tools are document-oriented, and not citation-oriented: the same document may cite different articles from a specific journal, but a search on the *cited* journal title abbreviation would count this multiple occurrence as only one relation. For example, in 1984, the aggregate of articles in the *Journal of the American Chemical Society (JACS)* contained 2240 references to articles in the *Journal of Chemical Physics*, while the number of hits on DIMDI for this relation was only 655. A reconstruction of the *Journal Citation Reports* on the basis of the on-line version but in terms of unique citation relations would require that all the references be downloaded, and that the data be processed in a way similar to the one described here for the CD-ROM version of the *Science Citation Index*.

The CD-ROM version of the *SCI* enables the researcher to reconstruct its own relational database without time constraints on the downloading, and most of the operations can be done unattended. However, this data is not organized for the purpose of scientometric mapping using aggregated journal-journal citations. The references are displayed on an article-by-article basis, and therefore, the data has to be thoroughly reorganized: the journal-journal network has to be represented as a relational database containing pointers from the cited side to the citing side, and vice versa, at the aggregate level.

This is not a *sine cure*: first, the journal title abbreviations are not standardized on the cited side; second, the abbreviations on the cited and the citing sides do not always match. For example, *The New England Journal of Medicine* is abbreviated and standardized on the citing side as *N Engl J Med*, while it is cited almost exclusively as *New Engl J Med*. Table 1 lists the variations in the spellings of this journal abbreviation in the cited references. The various problems reflect the lack of standardization in journal title abbreviations within citations on the original tapes.**

* The 1988 data were obtained with the support of grant nr. 8810197 from the National Science Foundation of the US. See also: *Leydesdorff & Cozzens* 1993.

** *T. Braun*, personal communication. See also: E. GARFIELD, *Citation Indexing*, New York, Wiley, 1979.

Table 1
Variations and misspellings of the *New England Journal of Medicine* in the *cited* journal abbreviations, for SCI-data Jan – March 1993

Cited journal abbreviation	Number of occurrences	
N-E-J-MED	1	
N-EGL-J-MED	1	
N-EGN-J-MED	1	
N-ENGL-J	1	
N-ENGL-J-ED	1	
N-ENGL-J-MD	1	
N-ENGL-J-ME	1	
N-ENGL-J-ME-D	1	
N-ENGL-J-MED	11	(citing journal abbr.)
N-ENGL-J-MED-1	2	
N-ENGL-J-MED-2	2	
N-ENGL-J-MED-S	1	
N-ENGL-JMED	4	
N-ENGL-M-MED	1	
NEW-ENGGL-J-MED	3	
NEW-ENGJL-J-MED	1	
NEW-ENGL-J-MED	22684	(> 99%)
Total	22720	

Methods and materials

In order to limit the amount of processing, I used data only for the first quarter of 1993. This amounts to 167,249 records containing 2,932,666 references. These records were gathered by ISI from 2901 journals. Six journals contained no references at all, and six journals only one. These twelve journals were discarded from further analysis, leaving us with 2889 well-defined journals on the "citing" side. Analogously, singular journal-journal citation on the cited side was discarded in this initial phase of organizing the data. This filter on singular journal-journal citations reduced the number of indicated relations in the journal citation network from 817,366 to 302,227.

All journal abbreviations in references (including misspellings) were saved for later error-correction purposes. The abbreviations of journal titles in the cited references amounted to 110,678 unique journal title abbreviations, of which 68,266 occurred only once. These large numbers are based either on misspellings or on references to journal which are otherwise not covered by the ISI-database.

Processing

The journal title abbreviation can be used as a search field for creating extensive document sets (e.g., with a wild card like A*, B*, etc., and/or with statements containing a Boolean "or"). These sets can be downloaded in custom-defined formats so that the saved file contains only the abbreviated journal titles and the citations. This data is imported into a relational database. Although these processes are time-consuming, most of the processing can be done unattended.

As noted, the major problem is eventually to relate the cited abbreviations to the citing abbreviations. In a number of cases, the title abbreviations on both sides are identical, as when the title consists of only one word which is not abbreviated (e.g., "CELL"). However, the technical challenge is to establish a maximum of correct relations with a minimum of intervention from the keyboard. We experimented a bit with various algorithms, and eventually decided to use the concatenations of the first two letters of each component of an abbreviation – separated by either spaces or hyphens – as identifiers. For example, the *American Review of Respiratory Diseases*, which is abbreviated as *AM-REV-RESPIR-DIS* on the cited side, and as *AMER REV RESP DIS* on the citing side, can be coded as *AMREREDI* on both sides, and this creates a unique relation in the database.

This procedure, however, also creates artifactual relations since, for example, the *International Journal of Geophysics* ("Int J Geoph" or "Int J Geophys) and the *International Journal of Geography* ("Int J Geogr") evaluate to the same abbreviation (INJ-GE). (A space and a hyphen are both evaluated as a hyphen.) Analogously, the abbreviation for *Physiological Review* (PHRE) equates with the one for *Pharmacological Review*, and this abbreviation can also interact with abbreviations for *Physical Review* whenever the specific parts of this journals (*Physical Review A*, etc.) are not properly distinguished in the citation.

Using this procedure, relations between the cited and the citing journal title abbreviations could automatically be generated in 2411 of the 2889 *citing* journals (83.5%). However, two sources of errors remain: the relations are not necessarily unique, and the evaluation may not succeed in the case of misspellings or deviant abbreviations. If the relation cannot be made, this harms the coverage of the database, while misspellings and erroneous relations affect the quality of the data. Both types of errors may seriously affect the results of the statistical analysis underlying the envisaged mapping efforts.

Error correction I: extension of the coverage

A coverage of 83.5% of the (uniquely abbreviated) *citing* journal titles means a lack of coverage in 16.5% of the cases. This seems unacceptably and unnecessarily high. The 478 missing relations can, of course, be established manually on the basis of visual inspection of the data; but let us first raise the question of whether significant improvement of automatic processing might be achieved through further refinement of the search algorithm.

As a test for such improvements we used the abbreviation "TRANS" for *Transactions* on the citing side, which was evaluated into "TR" by our strategy of using two characters, while only the "T" is used in the abbreviations within the cited references. Analogously, "BULL" is often used as an abbreviation of *Bulletin* on the citing side, while it is usually abbreviated to one character ("B") on the cited side. (Note that the cited abbreviation will not necessarily be shorter than the citing one.)

By correcting for these two abbreviations, we were able to improve the result with an additional 31 relations, and thereby to raise the total number of relations to 2442, or 84.5% of the domain of 2889 citing journals – an improvement of only one percentage point. Specifying additional rules for establishing relations among abbreviations produces decreasing marginal returns. Therefore, I accepted this level as final for automatic processing.

The remaining 447 missing values were inspected visually. An on-screen interface was constructed that allows the user to relate or disrelate cited and citing abbreviations by pressing functions keys. In this way, citing journal abbreviations could be related to cited journal abbreviations in all but 39 cases (1.3%). These 39 journals have been excluded from further analysis.

Error correction II: misspellings, erroneous relations, etc.

A similar interface can be used for correcting erroneous relations, misspellings, etc. But a corrective review of the whole database would still be a major operation, given the more than 100,000 unique journal title abbreviations listed on the cited side [cf. *Narin* (1976)[10]].

Methodological reflection, however, can help us to limit this daunting task. Since scientometric distribution are well known to be heavily skewed, the latent structure in the data is often extremely pronounced. In other words, the journal-journal citation matrix is known to be extremely sparse, and the corresponding multi-dimensional

space is only thinly populated with clusters of journals. If the misspellings occurred at random, their effects on the results of a statistical analysis of structure in this highly structured data might be much less significant than in the case of normally distributed data.*

But do the errors in our case occur stochastically? The coupling procedure which has been described above generates artificial identifications of a systematic nature like the ones between "geophysics" and "geography," or between "physics" and "physiology". How would one expect a mistakenly included physiology journal to affect the citation relations among a set of physics journals? The artifactual links are not expected to be sensitive to specific differences among sub-fields of physics, nor is a physiology journal expected to have a larger propensity to cluster with, for example, *Physical Review A* or *Physical Review B*. Erroneous citation relations among completely different research fields can be introduced in processing, but they are not expected to play a major role in the structural properties of a local citation network like physics, given the relatively local containment of a citation cluster in the multi-dimensional space.

In summary, the scientometric properties of the data allow us to delimit a specific local area using the file containing all the misspellings, etc., in an explorative round, and then to clean up this local area until a stable solution in terms of relevant journals is achieved. Thus, one would expect a certain disturbance of, for example, *Physiological Review* when analyzing a physics field of, say, forty physics journals; but once these forty-one abbreviations have been cleaned up case by case, the journal mapping for this area of physics should be correct.

In other words, one does not have to clean up the whole file in one big operation. The specific structure of the *Science Citation Index* allows us to zoom in to specific regions and densities, and to create a reliable mapping for these areas as needed. Note that by improving the data locally, one gradually invests in improving the database as a whole, and thus local cleaning is an effort with increasing returns. Using the noted interface, a local cleaning procedure takes only fifteen minutes or so.**

* For example, when we used *SciSearch* on-line in a previous research project for generating a journal-journal citation map on the basis of document-to-document relations (using the method described in the introduction), the resulting factor analysis and mapping were *significantly* similar to the ones based on unique citation relations using *JCR* data, despite the noted differences. Thus, the (eigen-)structure of the network could be robustly retrieved despite differences up to an order of magnitude in the data.

** After such a local clean up, one has to reconstruct the network of relations among the databases, but as noted this can be done unattendedly, and over night.

In order to provide the reader with an idea of the effects of this cleaning on the quality of the data, some examples will be given below. The first two examples are methodological. Among other things, they show the effects of error corrections. In two further examples, these methods are applied to generate the 1993 journal maps for two areas of specific policy relevance.

Case I: *The Journal of Chemical Physics*

The *Journal of Chemical Physics* is a leading journal on the interface between physics and chemistry. We have discussed the journal map, taking this journal as a point of entrance, in a study of macro-indicators, since the existence of a cluster of "physical chemistry" and "chemical physics" journals defies the administrative division of the natural sciences into disciplines like physics and chemistry [cf. *Cozzens* and *Leydesdorff* (1993)[2]].

Figure 1A exhibits this journal map on the basis of the citing patterns using the 1988 tape data; Figure 1B is based on the cited patterns.* Figures 2A and 2B provide the comparable mappings for 1993 on the basis of the CD-ROM version. (In both years, the relevant journals are selected on the basis of the criteria described in our previous study; the areas in the mappings are indicated on the basis of a factor analysis of the citing and cited patterns of journal-journal citations, respectively.**)

In comparing the mappings for 1988 and 1993, some turn-over in terms of journals can be observed within the clusters, but the correspondences between the maps are obvious. (Since the choice of the axes is not meaningful in the case of multi-dimensional scaling, the reader is welcome to rotate the pictures for the interpretation.) First, one can observe in all pictures a central "chemical physics"/"physical chemistry" cluster. In the citing dimension, nine of the ten journals with their highest loading on this first factor in 1993, were also part of this configuration in 1988. The one new journal title is the *International Review in Physical Chemistry. The Annual Review of Physical Chemistry* and *Advances in Chemical Physics* were no longer covered by the database in 1993.

* See *Tijssen* et al. (1987) for a procedure which combines this information into a single map using quasi-correspondence analysis.

** One minor change in relation to the previous study is that we had previously set the citation threshold at 0.5% of the *minimum* of the total number of citations either cited or citing. In this study, the two sides are evaluated separately, and the citation threshold for drawing a journal into the analysis is set at the level of 0.5% of either total cited or total citing, respectively.

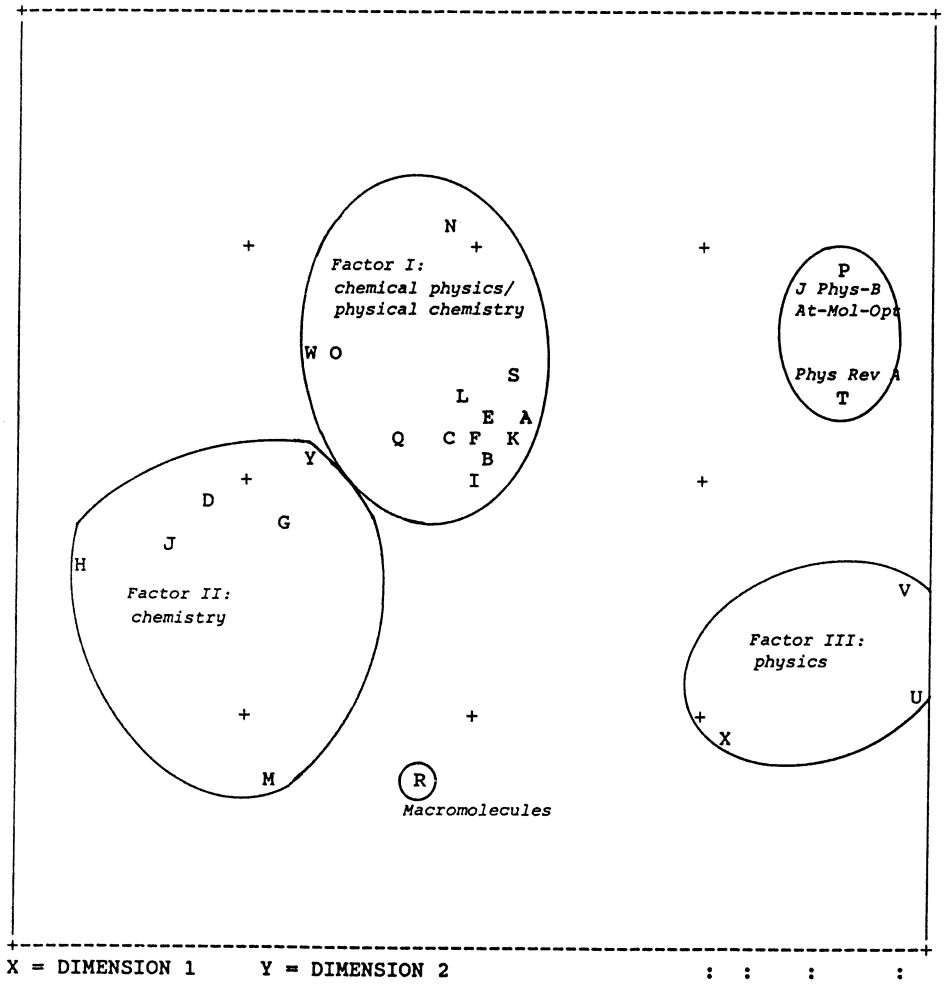

Fig. 1a. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1988, citing patterns. Journal name abbreviations:

A. ADV CHEM PHYS
B. ANNU REV PHYS CHEM
C. BER BUNSEN PHYS CHEM
D. CAN J CHEM
E. CHEM PHYS
F. CHEM PHYS LETT
G. CHEM REV
H. INORG CHEM
I. INT J QUANTUM CHEM
J. J AM CHEM SOC
K. J CHEM PHYS
L. J CHEM SOC FARAD T 2
M. J MAGN RESON
N. J MOL SPECTROSC
O. J MOL STRUCT
P. J PHYS B-AT MOL OPT
Q. J PHYS CHEM-US
R. MACROMOLECULES
S. MOL PHYS
T. PHYS REV A
U. PHYS REV B
V. PHYS REV LETT
W. SPECTROCHIM ACTA A
X. SURF SCI
Y. THEOCHEM-J MOL STRUCT

The groupings are based on the VARIMAX factor solutions

On the cited side (i.e., comparing Figs 1b and 2b) more change can be observed, although the major impression remains one of stability. In 1988, two journals – the *Journal of Molecular Spectroscopy* and *Spectrochimica Acta A* – had a significantly different pattern of *being cited*, while they belonged to the core set of "chemical physics" on the citing side. By 1993, these two journals had disappeared from this citation environment at 0.5% threshold level. However, the interface with the chemistry cluster exhibits a new differentiation in this year. Thus, the structure of the archive develops gradually in terms of being cited, while the pictures on the citing side have remained almost identical (except for the permitted rotation).

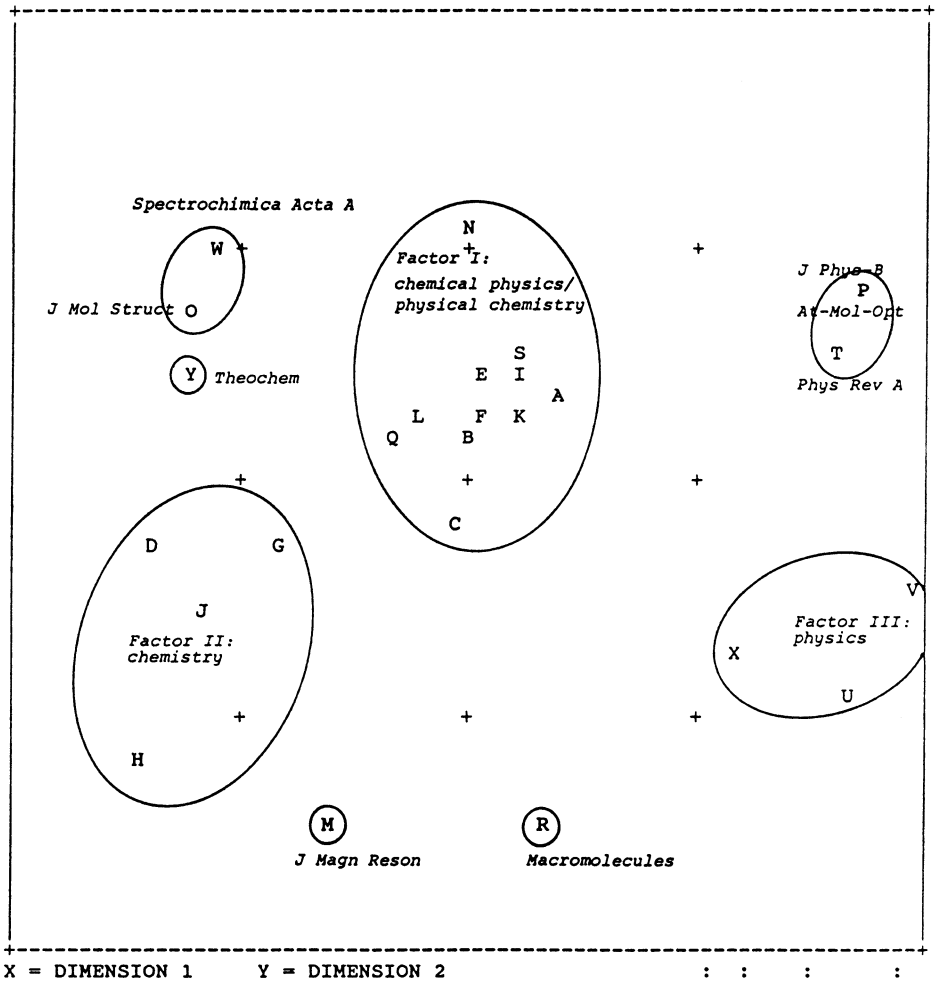

Fig. 1b. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1988, cited patterns. Journal name abbreviations as in Fig. 1a

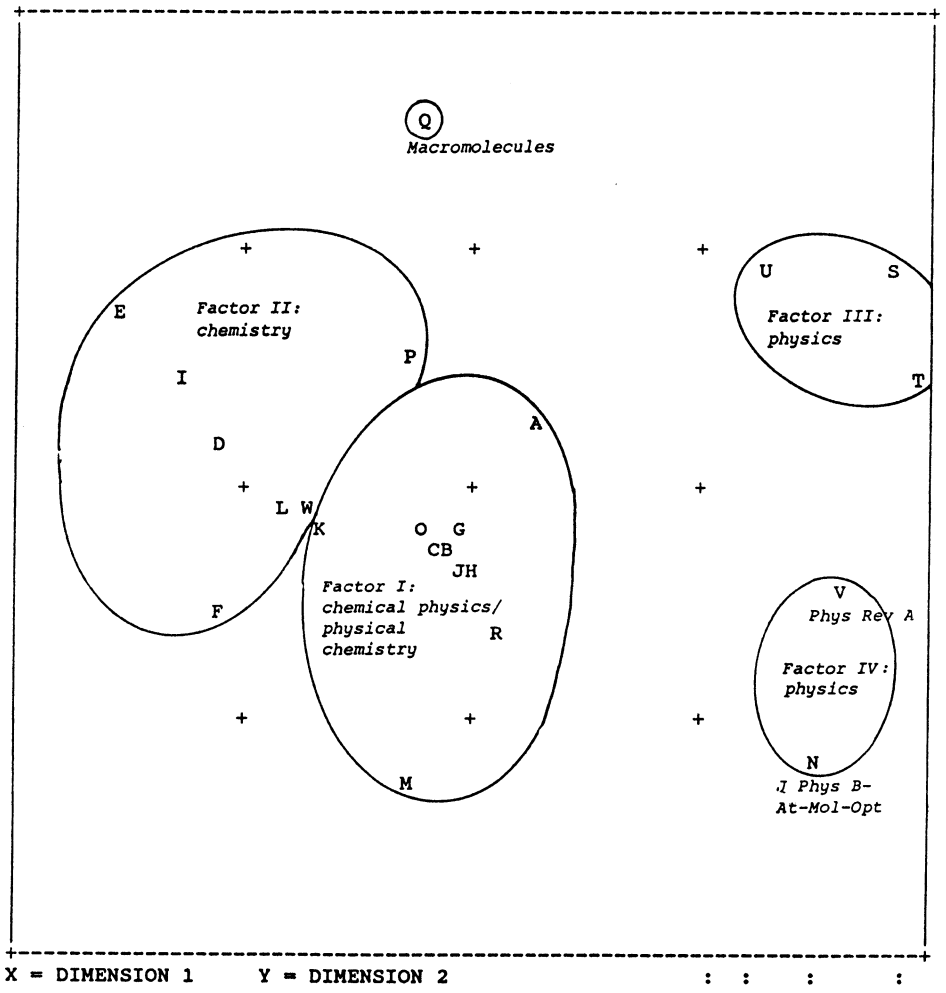

Fig. 2a. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1993, citing patterns. Journal name abbreviations:

A. BER BUNSEN GES PHYS CHEM
B. CHEM PHYS LETT
C. CHEM PHYS
D. CHEM REV
E. INORG CHEM
F. INT J MASS SPECTROM ION PROC
G. INT J QUANTUM CHEM
H. INT REV PHYS CHEM
I. J AMER CHEM SOC
J. J CHEM PHYS
K. J CHEM SOC FARADAY TRANS
L. J COMPUT CHEM
M. J MOL SPECTROSC
N. J PHYS-B-AT MOL OPT PHYS
O. J PHYS CHEM
P. LANGMUIR
Q. MACROMOLECULES
R. MOL PHYS
S. PHYS REV B-CONDENSED MATTER
T. PHYS REV LETT
U. SURFACE SCI
V. PHYS REV A
W. THEOCHEM-J MOL STRUCT
Z. PHYSIOL REV

Grouping are based on the VARIMAX factor solutions

Finally, one may wish to note that in terms of both dimensions and years, the journal *Macromolecules* occupied a separate position between the cluster of "general and inorganic chemistry" and "condensed matter & surface science", and more distanced from the cluster containing journals specialized in atomic and molecular physics.

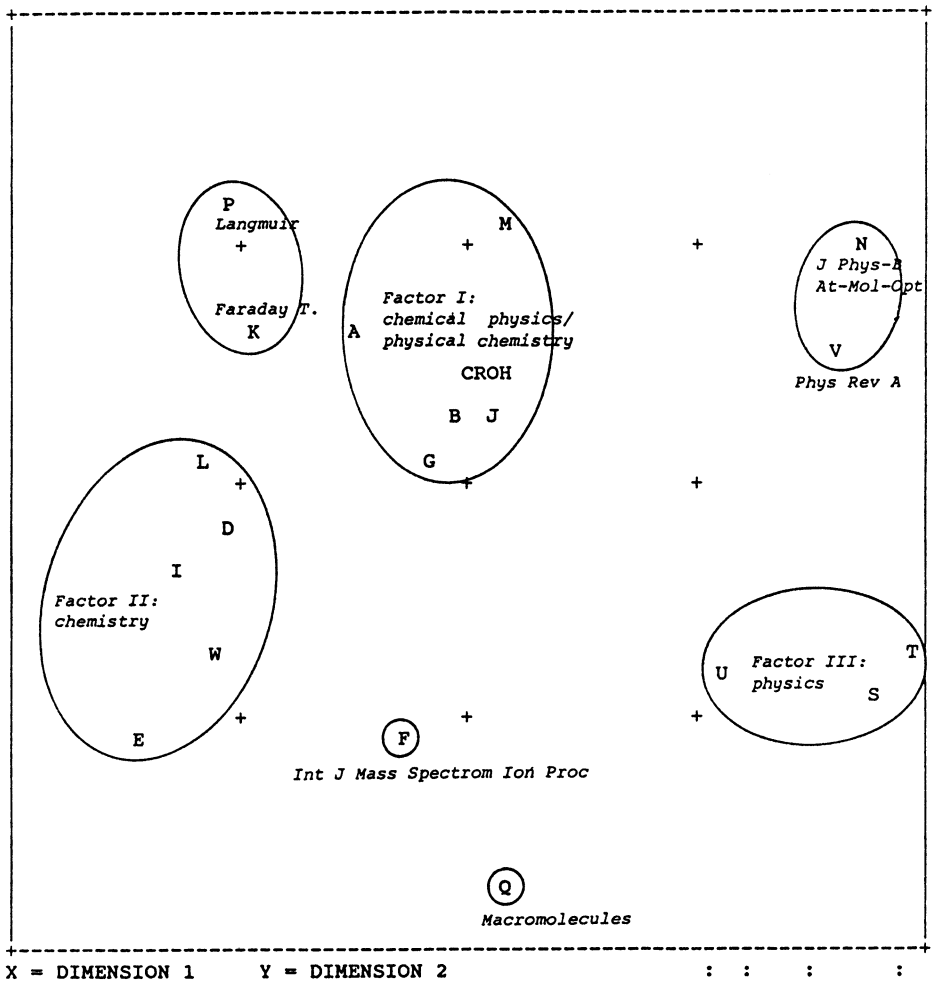

Fig. 2b. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1993, cited patterns. Journal name abbreviations as in Fig. 2a

In summary, we conclude that in accordance with previous analyses, the journal-journal citation map of an established scientific discipline exhibits change in terms of the relative positions of journals, and in relation to a certain turn-over, but the developments exhibited are gradual. Shifts occur in the order among the factor loadings of the journals, but stability prevails in the order among dimensions.

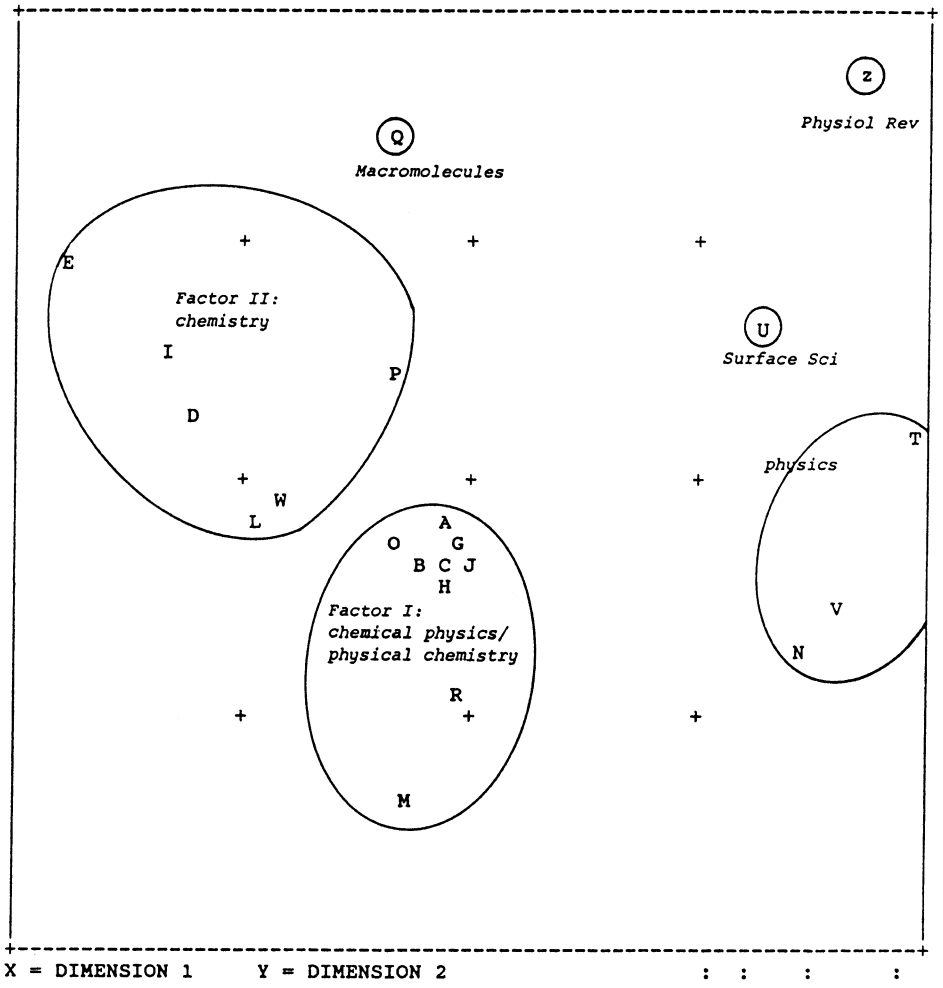

Fig. 3. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1993, citing patterns, *before any corrections on misspellings or coverage*. Journal name abbreviations as in Fig. 2a

Let us now take a closer look at the effects of misspellings and missing linkages in the case of the citing patterns. (The next example will consider the effects on the cited patterns.) Figure 3 exhibits the mapping results on the 1993 data before either of the two types of error correction discussed above had been performed. In this stage, *Physical Review B-Condensed Matter* and the *Journal of the Chemical Society Faraday Transactions* have not been covered because of mismatches in the abbreviations, and as noted above, *Physiological Review* has been drawn into the analysis erroneously. The net consequence is that the physics factors are incorrectly represented in Fig. 3.

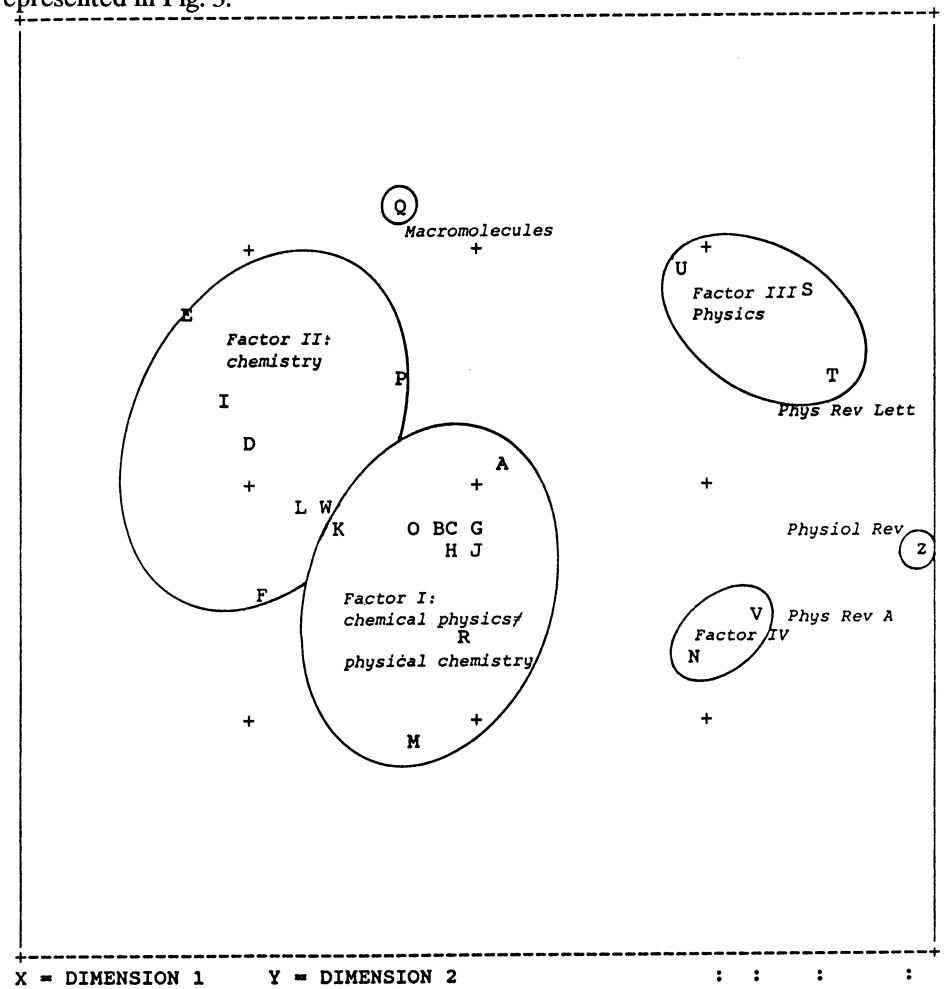

Fig. 4. Multi-dimensional scaling for the citation environment of the *Journal of Chemical Physics* 1993, citing patterns, *after corrections of the first type, but before corrections for misspellings, etc.* Journal name abbreviations as in Fig. 2a

Figure 4 shows the situation after error-correction of the first type. The two missing journals are now recovered, but the erroneous inclusion of *Physiological Review* has not yet been corrected. The resulting picture is almost correct when compared with the fully corrected one in Fig. 2a. The one misplaced journal is set far apart, and this has a spatial effect on the scaling by pushing the other journals closer together in order to capture the projection of the whole set, including the one misplaced and deviant journal. The factor loadings for the physics journals, however, correlate at r = 0.9644; i.e., the factor structure is almost identical. Furthermore, no physics journals are misplaced in this solution. In conclusion, this suggests that the errors related to failure in coverage are more serious than the remaining errors from misspellings, etc.

Case II: *The Lancet*

As noted above, the *New England Journal of Medicine* was abbreviated differently on the citing side and on the cited side in more than 99% of the references (see Table 1). Therefore, one would expect a larger impact from error-correction on misspellings in this case. Since the variation in spelling lies in the cited dimension, the effect of correction should appear greater on the being cited patterns than on the citing patterns.

The choice of the *New England Journal of Medicine* itself as the entrance point to the analysis would obviously lead to a breakdown of the methods, and therefore fail to provide us with a means of assessing the disturbance of the noted variations in spelling on the journal mapping. The map would be predictably distorted. However, if *The Lancet* is chosen as an entrance journal, one can evaluate the effects of the disturbance. *The Lancet* is an important medical journal which one would expect to be closely related to the *New England Journal of Medicine* in its relevant citation environment.

Figures 5a and 5b show the citing and cited mapping for 1993 before error correction on misspellings and variations in journal abbreviations, while Figs 6a and 6b exhibit these mappings after all corrections for misspellings, etc., have been made. Although one would expect a major effect in the cited dimensions, Figs 5b and 6b are not dramatically different. An interpreter might not be able to make a reasoned choice between these pictures on the basis of visual inspection alone.

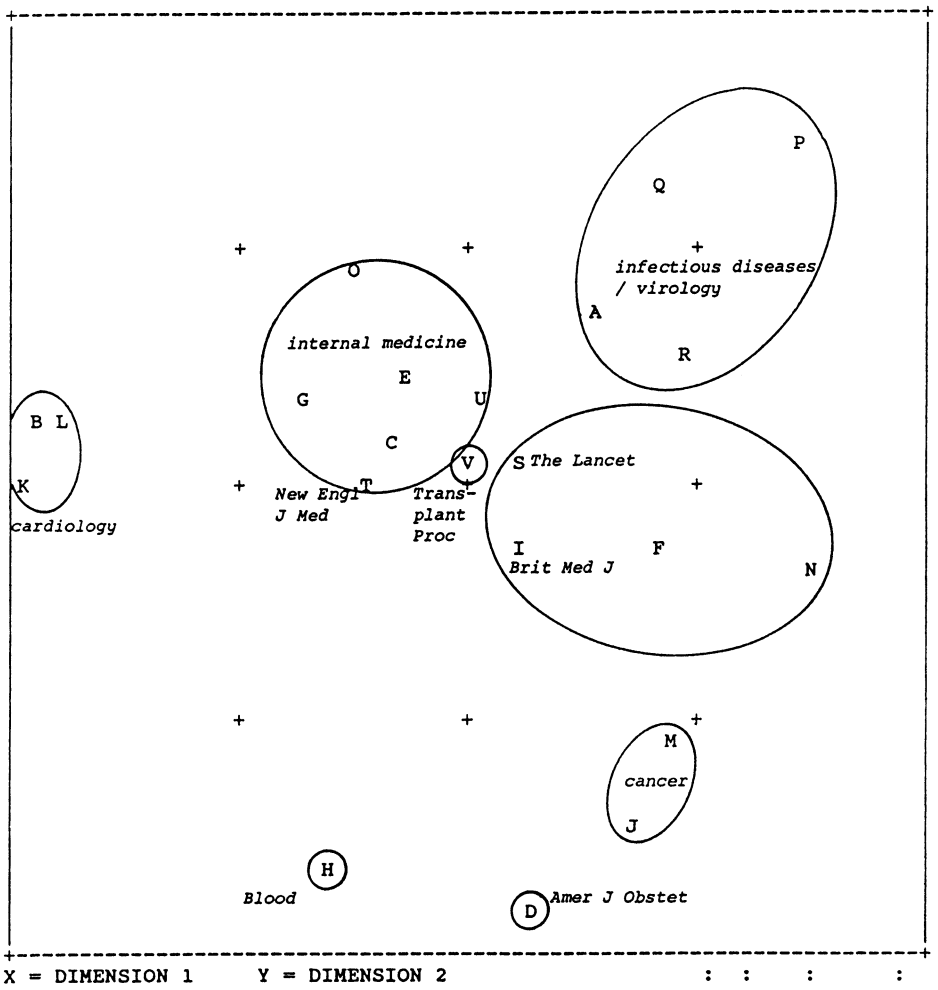

Fig. 5a. Multi-dimensional scaling for the citation environment of *The Lancet* 1993, citing patterns *before corrections on misspellings, spelling variations, etc.* Journal name abbreviations:

A. AIDS	I. BRIT MED J	Q. J INFEC DIS
B. AMER J CARDIOL	J. CANCER	R. J MED VIROL
C. AMER J MED	K. CIRCULATION	S. LANCET
D. AMER J OBSTET GYNECOL	L. EUR HEART J	T. N ENGL J MED
E. ANN INTERN MED	M. EUR J CANCER	U. MED CLIN
F. ARCH DIS CHILD	N. GUT	V. TRANSPLANT PROC
G. ARCH INTERN MED	O. JAMA-J AM MED ASSN	
H. BLOOD	P. J. CLIN MICROBIOL	

Groupings are based on the VARIMAX factor solutions

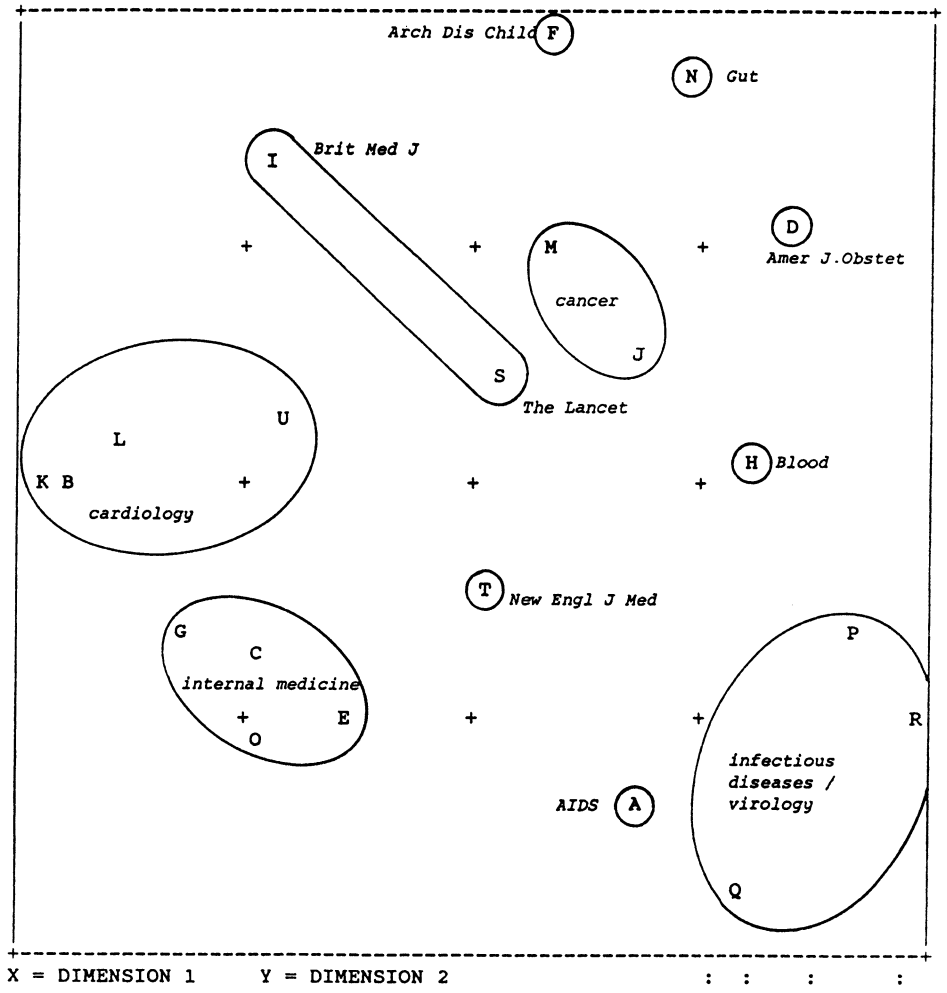

Fig. 5b. Multi-dimensional scaling for the citation environment of *The Lancet* 1993, cited patterns *before corrections on misspellings, spelling variations, etc.* Journal name abbreviations as in Fig. 5a

The major effect of the error correction is not in the spatial arrangement using the MDS program, but in the factor analytical results which are used for legitimating the grouping in the figures. After correction of the cited journal abbreviations, the *New England Journal of Medicine* can be attributed unambiguously to the "internal medicine" cluster, while before this correction it had appeared as an isolate.

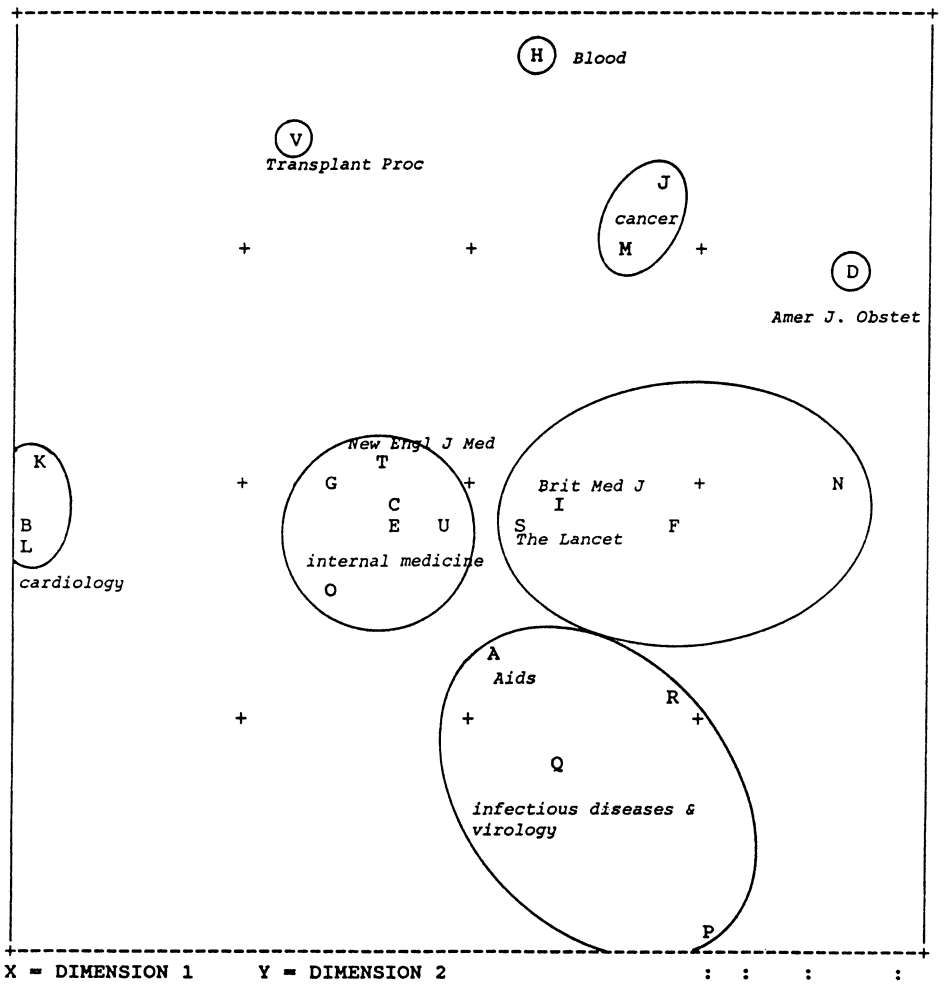

Fig. 6a. Multi-dimensional scaling for the citation environment of *The Lancet* 1993, citing patterns, *after corrections for variations in spelling*. Journal name abbreviations as in Fig. 5a

When one compares these results with the corresponding ones for 1988, two major developments can be noted. First, one group of cardiology journals has gained a prominent place in this citation environment, although these journals failed to pass the 0.5% citation threshold when *The Lancet* was used as the entrance journal to the network in 1988. Second, the journal *AIDS* has unambiguously joined the group of core medical journals in this period, while it previously held a more specialist position.

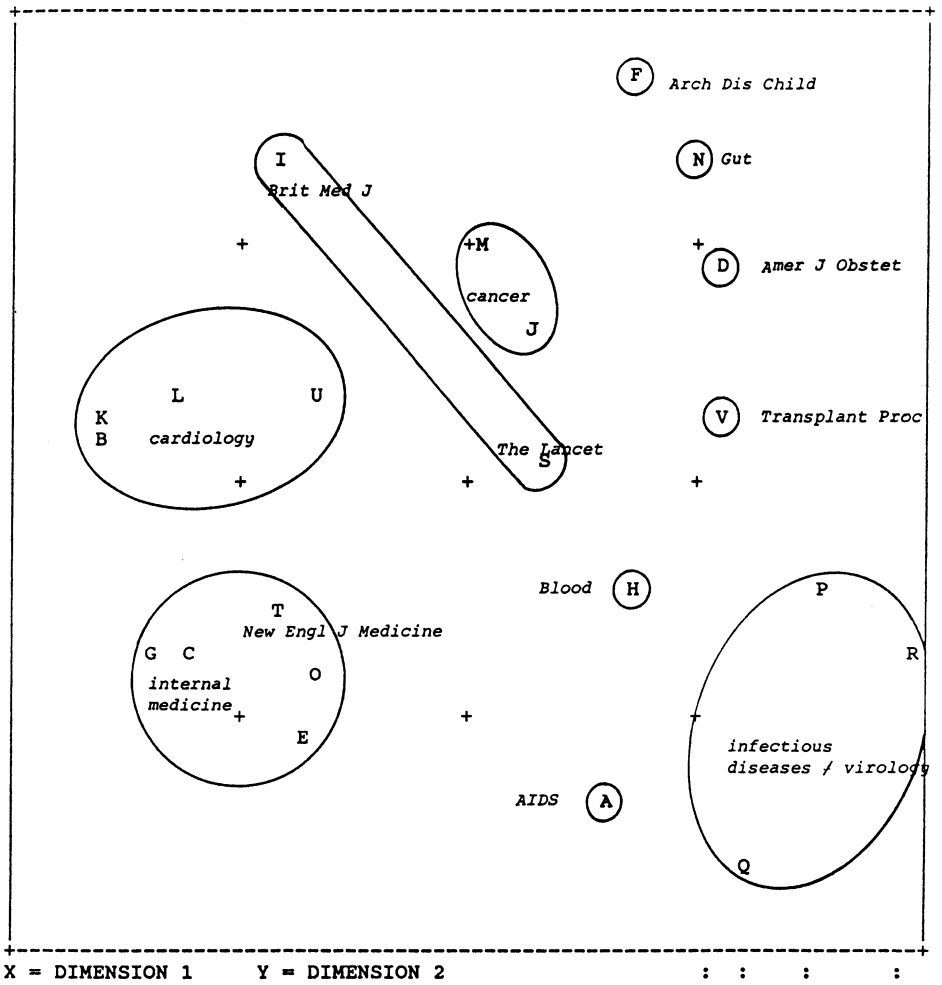

Fig. 6b. Multi-dimensional scaling for the citation environment of *The Lancet* 1993, citing patterns, *after corrections for variations in spelling*. Journal name abbreviations as in Fig. 5a

Case III: *AIDS*

In a previous study[8] we discussed the journal *AIDS* as an indicator of an emerging development over the 1980s. What has happened to this specialty area in the meantime? In our study, we suggested more generally, as a criterion for identifying significantly new developments that journals which are newly added to the database, should additionally exhibit a noticeably different pattern of *being cited* among the

longer established journals in its citation environment. The reasoning behind this idea is that *citing* can be considered as an intentional act on the part of authors, while *being cited* is a network effect. Since we are interested in network effects and not so much in the behaviour of authors, newly added journals should be distinguished in terms of whether or not they have had an impact on the citation network structure, i.e., in terms of their cited patterns, and independently of the pattern of the citations used by the collective of authors in these journals (the citing patterns).

The journal *AIDS* was newly added to the *Science Citation Index* in our data for 1986, and then exhibited this additional property of indicating a specific dimension when the citation matrix was analyzed in terms of its "cited" structure. This specific pattern was also found in 1988. By 1993, however, the cited pattern of *AIDS* had become second in terms of loading on a specific factor, while the noted indicator function had been taken over by the newly added *Journal of Acquired Immune Deficiency Syndrome*. This pattern is not dependent on whether one takes either of these two journals as a point of entrance for creating the relevant citation environment. Furthermore, both analyses attribute the journal *AIDS* to a cluster of virology journals on the *citing* side, while the *Journal of Acquired Immune Deficiency Syndrome* is attributed to a cluster of internal medicine journals. Figures 7a and 7b provide the reader with the respective journal mappings using the journal *AIDS* as point of entrance to this citation environment.

In the terminology of the previous study, we may conclude that the AIDS journals in 1993 still exhibit a specific pattern of being cited, and thereby manifest an impact on the relevant citation network different from neighbouring disciplines like internal medicine, virology, and immunology.

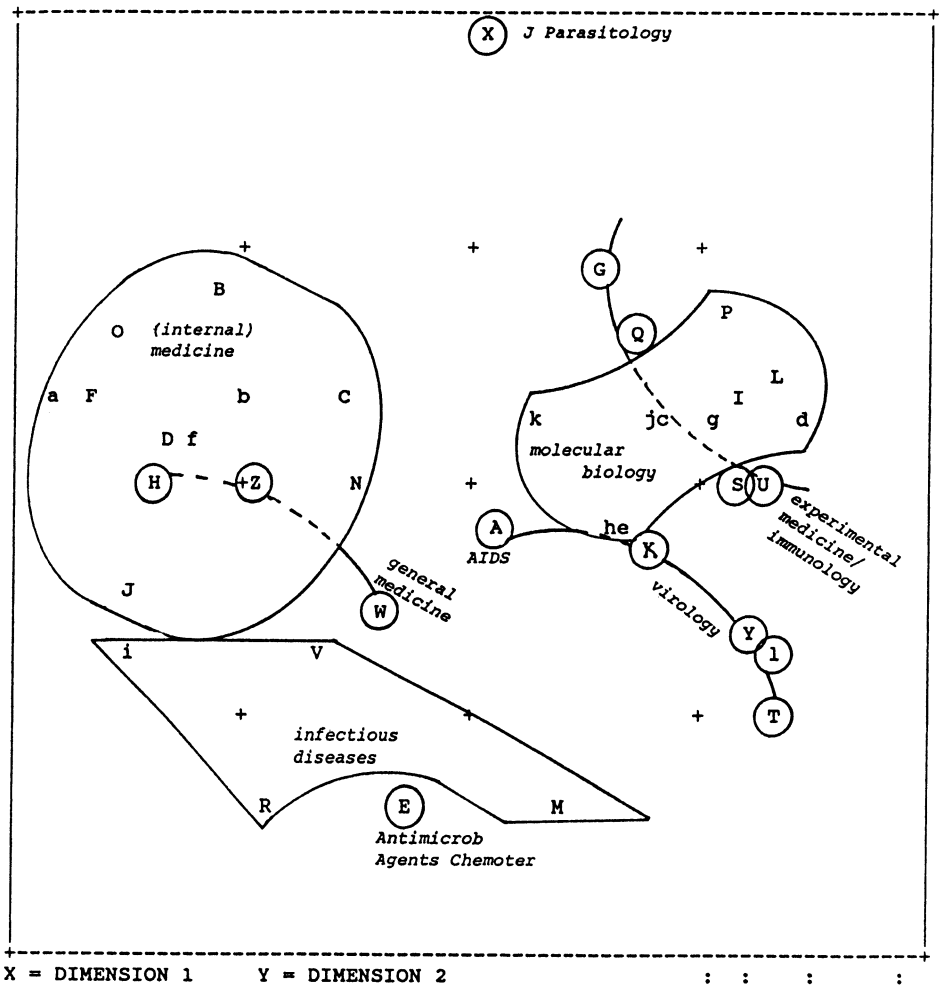

Fig. 7a. Multi-dimensional scaling for the citation environment of the journal *AIDS* 1993, citing patterns. Threshold 1.00%. Journal name abbreviations:

A. AIDS
B. AMER J GASTROENTEROL
C. AMER J MED
D. ANN INTERN MED
E. ANTIMICROB AGENTS CHEMOTHER
F. ARCH INTERN MED
G. BLOOD
H. BRIT MED J
I. CELL
J. CLIN INFECT DIS
K. CHEM IMMUNOL
L. EMBO J
M. INFEC IMMUNITY
N. J ACQ IMMUN DEFIC SYNDROME
O. JAMA-J AM MED ASSN
P. J BIOL CHEM
Q. J CLIN INVEST
R. J CLIN MICROBIOL
S. J EXP MED
T. J GEN VIROL
U. J IMMUNOL
V. J INFEC DIS
W. J MED VIROL
X. J PARASITOL
Y. J VIROL
Z. LANCET
a. MED J AUSTRALIA
b. N ENGL J MED
c. NATURE
d. NUCL ACID RES
e. MICROBIOL REV
f. MED CLIN
g. PROC NAT ACAD SCI USA
h. RES VIROLOGY
i. SCAND J INFEC DIS
j. SCIENCE
k. PHARMACOL THER
l. VIROLOGY

Groupings are based on the VARIMAX factor solutions.

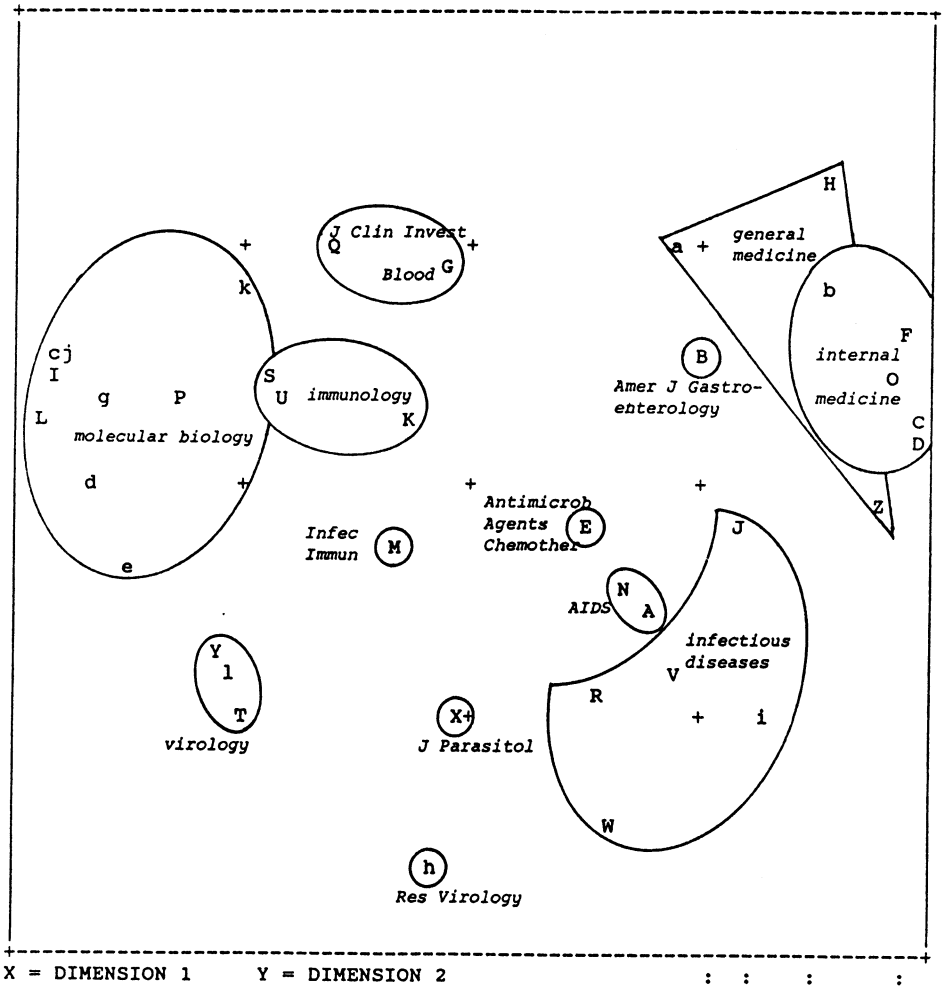

Fig. 7b. Multi-dimensional scaling for the citation environment of the journal *AIDS* 1993, cited patterns. Threshold 1.00%. Journal name abbreviations as in Fig. 7a

Case IV: *Artificial Intelligence*

In the same context of discussing indicators of emergence, we have also analyzed the position of *Artificial Intelligence* as a leading journal in a rapidly developing research specialty [cf. *Van den Besselaar* and *Leydesdorff* (1993)[12]; *Leydesdorff* et al.,

(1994)[8]. The conclusion on the basis of data for 1984, 1986, and 1988 was that this specialty exhibited a lack of stability in terms of its constitutive journals and relevant citation environment during this period. Therefore, it seems interesting to raise the question of how this journal developed further in relation to its citation environment between 1988 and 1993.

In comparison to 1988 (see Fig. 8), the journal cluster of artificial intelligence in 1993 (Figure 9) has retained a stable core which consists exclusively of the journals *Artificial Intelligence* and *AI Magazine*.* As in the period between 1984 and 1988, the environment for this journal group changed drastically between 1988 and 1993, with the sole exception of a cluster of "pattern recognition" journals, which has remained the major cluster in this citation environment. However, journals focussing on software engineering, robotics, and programming problems tend to be replaced by journals from computational chemistry, corrosion science, remote sensing, and other fields of application. These results suggest that application has become a central concern in the field of artificial intelligence.

Note that the study for 1988[7] covered a larger data set including also journals listed in the *Social Science Citation Index*, the *Arts and Humanities Citation Index*, and the *Compumath Citation Index*. Particularly, the 1988 data indicated an interesting relation of *Artificial Intelligence* with psychological journals and *The International Journal of Man-Machine Studies*. These relations cannot be further analyzed using a reconstruction of the *JCR* for 1993, which is based on only the *Science Citation Index*.

* *IEEE Expert* was also part of this core group of AI journals in 1988 and 1990 (*Van den Besselaar* and *Leydesdorff* 1993). However, beginning 1993 this journal was no longer listed in the CD-ROM version of the *Science Citation Index*. It is now part of the *Social Science Citation Index*.

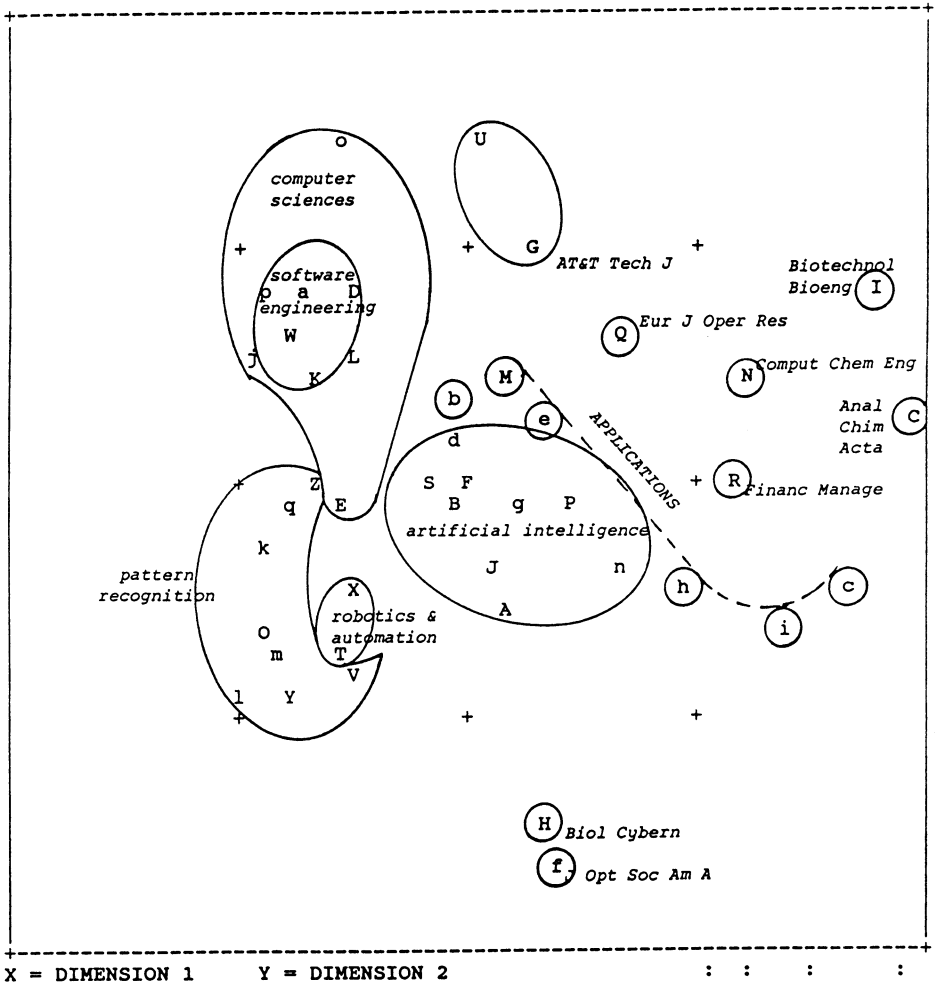

Fig. 8. Multi-dimensional scaling for the citation environment of the journal *Artificial Intelligence*, 1988, citing patterns. Journal name abbreviations:

A. ACTA PSYCHOL
B. AI MAG
C. ANAL CHIM ACTA
D. ANGEW INFORM
E. ANNU REV COMPUT SCI
F. ARTIF INTELL
G. AT&T TECH J
H. BIOL CYBERN
I. BIOTECHNOL BIOENG
J. COGNITIVE SCI
K. COMMUN ACM
L. COMPUT ARTIF INTELL
M. COMPUT BIOMED RES
N. COMPUT CHEM ENG
O. COMPUT VISION GRAPH
P. DAEDALUS
Q. EUR J OPER RES
R. FINANC MANAGE
S. IEEE EXPERT
T. IEEE J ROBOTIC AUTOM
U. IEEE J SEL AREA COMM
V. IEEE T PATTERN ANAL
W. IEEE T SOFTWARE ENG
X. IEEE T SYST MAN CYB
Y. IMAGE VISION COMPUT
Z. INFORM SCIENCES
a. INFORM SOFTWARE TECH
b. INT J MAN MACH STUD
c. J ECONOMETRICS
d. J LOGIC PROGRAM
e. J MOL GRAPHICS
f. J OPT SOC AM A
g. J PHILOS LOGIC
h. J ROY STAT SOC B MET
i. J STAT PLAN INFER
j. LECT NOTES COMPUT SC
k. P IEEE
l. PATTERN RECOGN
m. PATTERN RECOGN LETT
n. PROG PHYS GEOG
o. SIAM J COMPUT
p. SIGPLAN NOTICES
q. TSI-TECH SCI INF

Groupings are based on the VARIMAX factor solutions

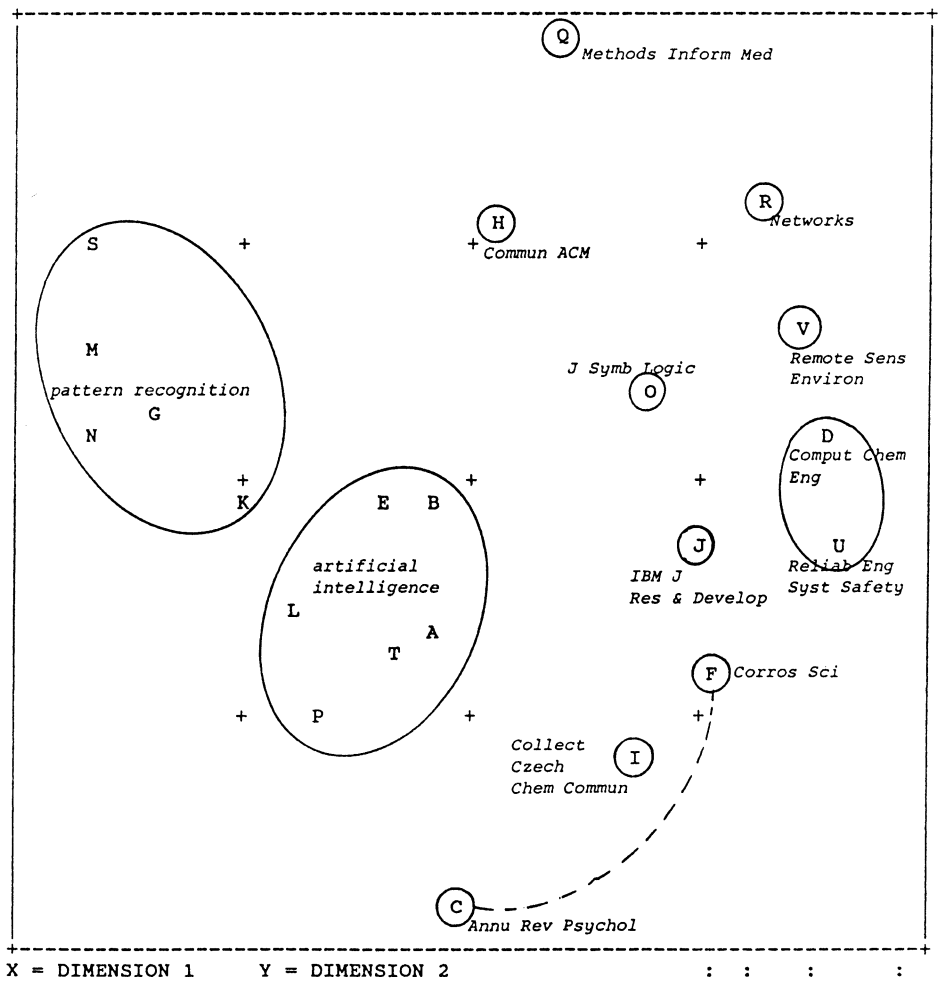

Fig. 9. Multi-dimensional scaling for the citation environment of the journal *Artificial Intelligence*, 1993, citing patterns. Journal name abbreviations:

A. AI MAG	I. COLLECT CZECH CHEM COMMUN	O. J SYMB LOGIC
B. ARTIF INTELL	J. IBM J RES DEVELOP	P. MACH LEARN
C. ANNU REV PSYCHOL	K. IEEE TRANS PATT ANAL MACH INT	Q. METHODS INFORM MED
D. COMPUT CHEM ENG	L. IEEE TRANS SYST MAN CYBERN	R. NETWORKS
E. COMPUTER	M. IMAGE VISION COMPUT	S. PATT RECOG
F. CORROS SCI	N. INFORM SCIENCES	T. PHIL TRANS ROY SOC LONDON B
G. CVGIP-IMAGE UNDERSTAND		U. RELIAB ENG SYST SAFETY
H. COMMUN ACM		V. REMOTE SENS ENVIRON

Groupings are based on the VARIMAX factor solutions

Conclusions and discussion

The challenge in this project has been primarily a technical one: is it feasable to construct a database equivalent to the *Journal Citation Reports* of the *Science Citation Index* using the CD-ROM version of this database? The answer is: "yes, but there remain important sources of error". The various sources of potential error have been discussed, and strategies to counteract them have been suggested. In particular, we explained why it is not necessary from the perspective of scientometric mapping to correct all misspellings, etc., at the level of the whole database, since such problems can be handled locally. Our results suggest that error-correction is necessary, but that the impacts of misspellings, etc., will be limited if one is performing statistical analysis of the data in terms of (eigen-)structures.

We have shown that the resulting figures can be informative about new developments. The suggested methods, however, can be extended to other files like the *Social Science Citation Index* and the *Arts & Humanities Citation Index*. Thus, more comprehensive journal mappings can be created, on a quarterly (or yearly) base, and without much delay. Furthermore, using the essentially raw data stored on the CD-ROMs by ISI allows one to develop one's own standards for data processing and error correction, independently from the database producer.

Among other things, this technique may prove to be important for the purpose of creating longer time-series of complex network data. The journal-journal network data provide us with a rich domain, including the emergence and death of individual units and clusters of journals at various levels of aggregation [*Leydesdorff* (1991)[5]]. This data exhibits both stability, propensities to cluster, and various forms of change, and therefore it may offer opportunities for empirical research on the non-linear dynamics of scientific communication networks [cf. *Leydesdorff* (1992)[6]]. As I understand the licence agreement, one is allowed to use the data for these academic purposes on campus: but I suppose that ISI would expect to receive payment of royalties whenever the data is used for the creation of scientometric mappings on a commercial basis.

*

I would like to acknowledge the help of Suzanne *Bakker* in organizing the data.

References

1. M. P. CARPENTER, F. NARIN, Clustering of scientific journals, *Journal of the American Society of Information Science*, 24 (1973) 425–35.
2. S. E. COZZENS, L. LEYDESDORFF, Journal systems as macro-indicators of structural change in the sciences, In: A. F. J. VAN RAAN et al. (Eds), *Science and Technology in A Policy Context*, DSWO, Leiden, pp. 219–233.
3. P. DOREIAN, T. J. FARARO, Structural equivalence in a journal network, *Journal of the American Society of Information Science*, 35 (1985) 28–37.
4. L. LEYDESDORFF, The Development of frames of references, *Scientometrics*, 9 (1986) 103–25.
5. L. LEYDESDORFF, The static and dynamic analysis of network data using information theory, *Social Networks*, 13 (1991) 303–345.
6. L. LEYDESDORFF, Irreversibilities in science and technology networks: An empirical and analytical approach, *Scientometrics*, 24 (1992) 321–357.
7. L. LEYDESDORFF, S. E. COZZENS, The delineation of specialties in terms of journals using the dynamic journal set of the SCI, *Scientometrics*, 26 (1993) 135–156.
8. L. LEYDESDORFF, S. E. COZZENS, P. VAN DEN BESSELAAR, Tracking areas of strategic importance using scientometric journal mappings, *Research Policy*, 23 (1994) 217–219.
9. H. F. MOED, The use of on-line databases for bibliometric analysis. In: L. EGGHE, R. ROUSSEAU, *Informetrics 87/88*, Elsevier, Amsterdam etc., 1988, pp. 133–46.
10. F. NARIN, *Evaluative Bibliometrics*, CHI, Cherry Hill, NJ, 1976.
11. R. TIJSSEN, J. DE LEEUW, A. F. J. VAN RAAN, Quasi-correspondence analysis on square scientometric transaction matrices, *Scientometrics*, 11 (1987) 347–61.
12. P. VAN DEN BESSELAAR, L. LEYDESDORFF, Artificial intelligence or expert systems: The emergence of AI as a research tradition, In: H. AKKERMANS, J. BREUKER (Eds), *Proceedings Nederlandse AI Conferentie 1993*, Petten, ECN/NVKI, 1993, pp. 317–328.