JIS

A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research

Félix Moya-Anegón, Víctor Herrero-Solana and Evaristo Jiménez-Contreras

Department of Library and Information Science, University of Granada, Spain

Received 22 January 2004
Revised 13 July 2005

Abstract.

The visualization of scientific field structures is a classic of scientometric studies. This paper presents a domain analysis of the library and information science discipline based on author co-citation analysis (ACA) and journal co-citation analysis (JCA). The techniques used for map construction are the self-organizing map (SOM) neural algorithm, Ward's clustering method and multidimensional scaling (MDS). The results of this study are compared with similar research developed by Howard White and Katherine McCain [1]. The methodologies used allow us to confirm that the subject domains identified in this paper are, as well, present in our study for the corresponding period. The appearance of studies pertaining to library science reveals the relationship of this realm with information science. Especially significant is the presence of the management on the journal maps. From a methodological standpoint, meanwhile, we would agree with those authors who consider MDS, the SOM and clustering as complementary methods that provide representations of the same reality from different analytical points of view. Even so, the MDS representation is the one offering greater possibilities for the structural representation of the clusters in a set of variables.

Correspondence to: Víctor Herrero-Solana, University of Granada, E-18071 Granada, Spain. E-mail: victorhs@ugr.es

Keywords: domain analysis; author co-citation analysis (ACA); journal co-citation analysis (JCA); library and information science; multidimensional scaling (MDS); self-organizing map (SOM)

1. Introduction

The perception and representation of reality are complicated undertakings. Tufte [2] expressed it well with these words: '. . . the world is complex, dynamic, multidimensional; paper is static, flat. How are we to represent the rich visual world of experience and measurement on mere flatland?' The visualization of information is an activity that humans have developed over time in a two-dimensional form, conditioned by the restrictions of the traditional document forms [3]. The arrival of the digital computer meant a brave new tool for processing, visualizing and analysing information structures that could not be comprehended outside a computerized context. Williams [4] sums up the possibilities to date:

> visualizations have usually taken the form of sketches, drawings, graphics, and images that depict natural phenomena, theoretic structures of invisible objects, flows of different forms of matter and mixtures, systems of interactions, models of relationships in multivariate data sets, and volumetric information in medical images. More recently, visualizations representing business data, the contents of a library, and abstract concepts such as force fields have been produced.

White and McCain [5] introduce the problem of the visualization of literatures, stating, 'The trend now is to combine computerized graphics – visualizations – with computerized document retrieval, thereby making literatures seem more responsive.' This would represent the convergence of the two main lines of information science (IS) research into a single area, that of the visualization of the specialized literature of the discipline. The first of these lines of research comprises informetric-bibliometric studies, and is also known as the domain analysis [6]; the second would be that of information retrieval [7, 8].

Thus, we find ourselves with two clearly differentiated functions – analysis and interface – which we must attempt to integrate by introducing the methodology of visualization. There is, however, a gap between the two spheres. This gap has been described, though not resolved, by several authors, including Doszkocs et al. [9] and Polanco et al. [10]. While each sphere has its own objectives, they share the visualization methodologies, which we could group under two major headings:

(1) those of a statistic nature (based on multivariate analysis); and
(2) those of a connectionist nature (usually, but not exclusively, based on neural networks).

Within the techniques of multivariate statistical analysis, we find three basic methods:

(1) cluster analysis,
(2) principal component analysis (PCA), and
(3) multidimensional scaling (MDS) [11].

According to Kinnucan [12], 'These methods are referred to as dimensionality-reduction methods because this function is to simplify what might at first appear to be a complex pattern of associations among many entities.'

We described them in a previous study [13] as follows:

- **Cluster analysis.** This technique is used to create a two-dimensional display (e.g. dendrograms) of clusters of different objects whose relationships are represented by matrix values. This type of automatic classification, also known as numerical taxonomy, currently comprises more than 150 different techniques that are grouped in families according to shared procedures. In general, the IS discipline involves polythetic clustering hierarchies to create trees illustrating the hierarchy of relationships among elements on the basis of individual characteristics.
- **Principal component analysis (PCA).** The basic premise of PCA is that the linear relation between any two variables is best summarized by a regression line. In other words, the variable that represents the regression line as a point cloud contains essential information about both variables. The two variables are thus combined into a single factor. This mechanism can be used to reduce pairs of variables to single dimensions in order to simplify the graphic display of the elements included in the matrix.
- **Multidimensional scaling (MDS).** This multivariate analysis technique is used to identify the dimensions that best explain similarities and differences between variables. Because the purpose of MDS is to generate a map of objects, this approach can be considered an alternative to PCA.

Among the methodologies of a connectionist nature, the artificial neural networks (ANN) are now the most widely used, in particular the model called the self-organizing map (SOM, also known as the Kohonen model). It is based on the principle of the self-organization and grouping of *n*-dimensional vectors in a bidimensional space [14], and has been used to reduce dimensions in a wide variety of document spaces of diverse nature [15–23]. The working principle behind this model is simple: it consists of establishing a correspondence between the entered data and an output space or topological map. In this way, input data sharing common characteristics will activate adjacent areas on the map [24]. According to Kaski [25], the SOM presents four important properties for data exploration:

- **Ordered display.** This characteristic helps us to understand the underlying structures in the series of data, although the problem is that analysts must 'familiarize themselves' with the way the SOM presents the results.
- **Visualization of clusters.** We are able to perceive the *clustering density* of the different regions of the map. The disadvantage in this case is that the shape and borderlines of each cluster cannot be distinguished.
- **Missing data.** It is possible to work with missing vectors.

- **Outlier.** This enables us to detect unusual cases caused by input errors or similar anomalies.

There is very little bibliographic imformation comparing these techniques in the area of IS, with the exception of a recent study by White et al. [26] in which the MDS and SOM maps are descriptively compared in the following terms:

> . . . they are in effect the same map, differing mainly in matters of nuance. Both also seem reasonably accurate visualizations of a discipline, in the sense of *retrievable literatures represented by a display of interrelated authors*. 'Retrievable literatures' could mean works by the authors – their *oeuvres* – or works that cite them, or both. It could also mean works that mention them, such as one might find on the World Wide Web. Thus, one may see the SOM map, depicting Lin's Website, as part of the current transition from the static, print-based literature mapping of the past to the more dynamic bibliographic linkages possible through an interactive computer interface.

2. Objective

The objective of this study is to apply the connectionist approach provided by the SOM to the analysis of library and information science (LIS), to compare it with a multivariate approach. Our main point of reference was a similar study recently published by Howard White and Katherine McCain [1]. These authors examine the IS structure through author co-citation analysis (ACA), focusing on the production of 12 specialized journals selected by McCain from the *Social Science Citation Index* (SSCI). Our study involves a more comprehensive coverage, which includes, besides, the domain of library science (LS). This set of multivariate data is visualized using two statistical techniques, the PCA and the MDS. The study we present complements MDS and cluster analysis with the addition of the SOM. Moreover, we tried to widen the objective of White and McCain by introducing another vantage point: that of journal co-citation analysis (JCA), as a complement to ACA. We found several relevant studies by McCain [27–30], one of which contains the following observation:

> journal co-citation mapping is potentially of interest both to the researcher studying the structure of scholarly specialities through the published literature and to the collection manager concerned with developing core journals lists, selecting journals and evaluating collections that serve particular research-oriented constituencies. It is likely to complement journal network analyses, particularly those studies focusing on journal similarity based on patterns of citations received or sent. [27]

The analysis offered by White and McCain [1] with regard to IS is thorough and conclusive, for which reason we will limit the scope of our study to the methodology involved. Some comments on the structure and organization of the discipline are included, and in the conclusions we will refer with special emphasis to the work of authors comparing the two methods, always from the viewpoint of ACA and JCA as complementary elements in the analysis of any scientific area.

3. Material

Our source of data was the list of specialized publications from *Journal Citation Reports*, 1996 edition. Out of the 24 journals listed as having the greatest impact, we chose 17 (see Appendix A). Seven journals were rejected (Appendix B), as they have editorial scopes related to the application of IS to a specific technique or area of knowledge (medicine, geography, telecommunications), with LIS as a secondary interest. Moreover, *Knowledge Acquisition* was also rejected because no issue appeared in the last year of *Journal Citation Reports*, and its impact factor was artificially high. Out of the excluded publications, only *Bull Med Lib Assoc* reappeared on our map as one of the most cited, something to be taken into account in future studies of this sort.

Because this study was conceived as a validation and extension of previous research aimed at describing the characteristics of LIS, we hoped to confirm or refute some methodological uncertainties that our reading of the aforementioned studies had raised. For example, to what degree might the a priori selection of journals bias the results? A set of journals chosen on the basis of an external criterion such as the *impact factor* would provide for a non-random selection of sources. At the same time, the criterion of relevance would be met, and we could conclude that the results are independent of source selection, and vice versa. The fact that we collected data from the period 1992–7 also meant that we could update the descriptions given by previous studies, and point out what aspects had evolved along with LIS dynamics.

The information was obtained from the SSCI, in ASCII files, in ISI full record format. These files were downloaded into a relational database using an ad hoc program. The structure of this database was designed for the purpose of generating co-citation matrixes both of authors and journals, by means of chained queries. Appendix C offers a detailed description of all the journals appearing in the representation. A manual authority control was performed to correct the multiple variant names of journals (both ISI and non-ISI) in the CW field. The abbreviated title is followed by the complete title, ISSN, country of origin, and, for the non-ISI journals, the corresponding subject category.

4. Methodology

We used co-cites not only to study authors, but also to study the journals themselves. This point must be underlined, because earlier research, for the most part, used direct counts of cites ('journal to journal citation'; Campanario [31]). We believed co-cites to be preferable. First, they maintain the methodological coherence of the author analysis. Secondly, the additional information obtained about the journals could be interpreted in the same methodological context, at least in so far as data collection was concerned. Both the authors and journals are the most highly cited of each set (77 in the first case and 73 in the second). The threshold was indicated by the total citation received over the time, and was 61 for the authors set and 91 for the journals set. In both sets, we do a full authority control to normalize the data.

The resulting list of co-cited journals makes it possible to create a matrix (the symmetrical co-citation matrix) integrated by the journals most frequently used as intellectual references for researchers. This matrix would represent the interrelationships that, according to researchers, exist among the body of literature. The resulting vectors, then, would characterize each journal on the basis of its associations with the rest. After the preceptive normalization (Pearson's *r*) of the vector components, they are subjected to the simulator of the SOM network in order to obtain a matrix of neural activation. This output matrix constitutes the input for a second process, through which the SOM of authors or journal titles is constructed from the activation rates of each vector in each neuron. Then, in order to compare the output from the clustering techniques and that from the SOM, we used Ward's method to project the resulting groupings on the maps over the neural topology of authors and journals. This agglomeration rule, based on variance, allows us to avoid the chaining of objects that other methods produce (e.g. single link). Some authors use complete link method; we think that the grouping results are very similar to Ward's results. We use Statistica software to obtain cluster analysis. The complete link method and the Ward method produce different dendogram layouts. However, the groups that they give are very similar. The choosing of one or the other method does not change the final result. For more details about the advantages of Ward's method, see Egghe and Rousseau [11]. We used the same co-citation matrix to generate the MDS maps, whose groupings matched those of the SOM when the clustering was based on Ward's method. In both statistical methods (clustering and MDS) we used the software Statistica 6.0.

The SOM maps obtained constitute a combined representation of what Campanario [31] calls 'relations maps and domain maps', in his case applied only to journal titles. In our case, the labels are placed over the winning neurons in a 15×15 grid. The zones of influence of each vector are represented by fine contours around the winning neuron. These areas of influence are configured using the lowest activation values found for each neuron. Therefore, the shape, size and location of these zones depends on the topology of the whole set. So to speak, it is as if each author or journal invited over all the neighbours he associates with. The resultant topology, which respects neighbourly contact while keeping outsiders from trespassing, offers a graphic expression of the intellectual structure of library and information science. Above this layer, a second level, represented through thick lines and patterns, illustrates the groupings obtained using Ward's method.

Similarly, the MDS maps reduce *n*-dimensional space to just two dimensions. The elements – authors or journals – are organized in this reduced space in such a way that the distance between them is an indicator of the degree of relationship. Unlike the SOM, MDS tends to form clouds of common elements that may be visually and intuitively grouped by the observer. However, we once again applied clustering based on Ward's method for visualization.

For MDS and clustering, the Statistica software package was used, and for the SOM, a specific module called Statistica Neural Networks was applied.

5. Results and discussion

Before analysing the maps obtained, we must take note that three separate levels of interpretation are possible: depending on whether we use the point of view of the clustering, that of the SOM and MDS, or that of the opinion of experts in the material that we attempt to conjugate in our analysis of the results.

In the first SOM map (Figure 1), we see a certain coincidence in the groupings of the authors with those described in the White and McCain study. The upper third section of the map constitutes the bibliometric domain, whereas the rest belongs to the information

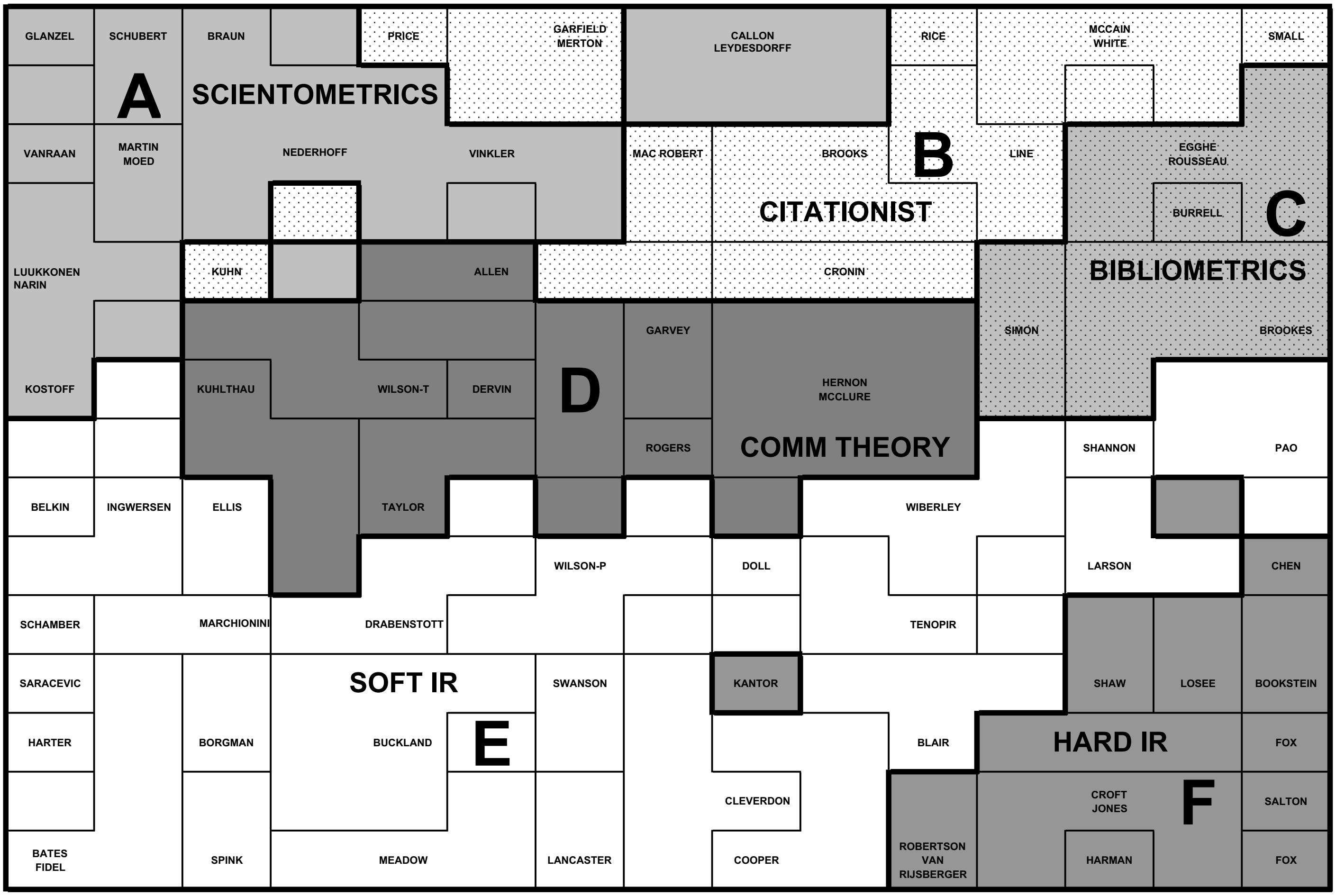

Fig. 1. SOM map of ACA.

retrieval domain. This division is clearly corroborated thanks to the groupings produced by clustering. The bibliometric domain would be represented by zones A, B and C, which would respectively correspond with what White and McCain call 'scientometrics,' 'citationist' (closely interrelated) and 'bibliometrics.' Meanwhile, the information retrieval (IR) domain embraces zones E and F, which respectively correspond to the soft or cognitive IR and the hard or algorithmic IR. Finally, we find a zone D, located in the centre of the map, which does not appear to belong to either of the two major domains described above: its authors proceed from different disciplines, although we might provisionally assign them the generic denomination 'communication theory'. In contrast with MDS representations, the central position does not mean that these authors play the role of 'obligatory' intellectual referents in their field. Rather, the explanation would be that these authors are related to an equal extent to both of the two major domains, and that the SOM uses them to clearly separate the authors of the upper and lower map sections. It is interesting to note that a considerable generational renewal has 'uprooted' LIS classics, such as Bradford, Lotka and Zipf in bibliometrics, as well as IR authors like Bush or Vickery. We believe that this is a direct result of the difference in the chronological periods included in the White and McCain study (1972–95) and in ours (1992–7).

In the first MDS map (Figure 2), we can clearly see the same groupings that appeared earlier. They fit the two-dimensional distribution quite well, with a stress value of 0.15. In MDS the centre, indicated with a circle, is important (whereas in the SOM it is not); and in this case, the empty centre indicates there are no intellectual referents common to the discipline as a whole. Authors Shannon and Pao occupy a point midway between the

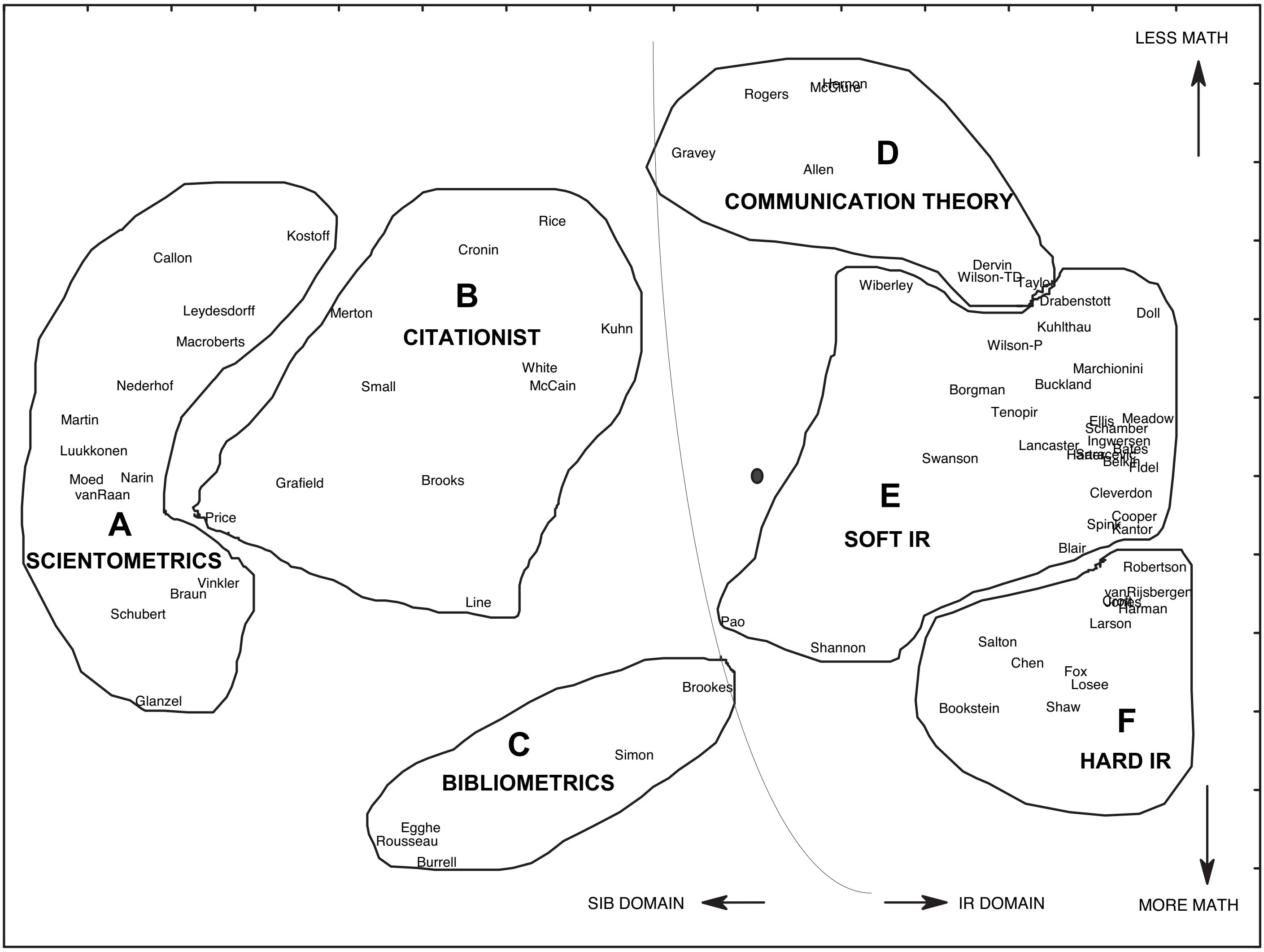

Fig. 2. MDS map of ACA.

SIB (scientometrics, informetrics and bibliometrics) domain and the IR (retrievalist) domain, perhaps as referents common to both. In this map, group D (communication theory) appears more closely related with the retrievalists, something which is not made clear on the SOM. Both maps illustrate the capacity of group E (soft IR) to relate with most of the other groups. Figure 3 completes author analysis with a dendogram of the ACA clustering based on Ward's method.

The second SOM map (Figure 4) illustrates the structure obtained from the journal co-citation matrix. Four distinct areas are clearly seen, which we will call A (information science, with a specific nucleus in information retrieval), B (library science, which includes library management and collection development), C (science studies) and D (management). There is also a series of journals with enough in common that we might logically group them together as 'technological' (*Database*, *Online*, *Cd-Rom Prof*, *Byte*, *Computer*, etc.). The SOM places these in an approximately central position. The clustering, however, was not able to identify these journals as a group, and desegregates them amid areas A and B. In contrast, a high degree of coincidence can be seen between the allotment of winning neurons and the clustering of the journals included in area D, which provides a compact, coherent map representation of this group. The *J Educ Libr Inform Sci* appeared in the library science cluster.

Another noteworthy finding is the entry of a considerable number of journals that do not pertain to

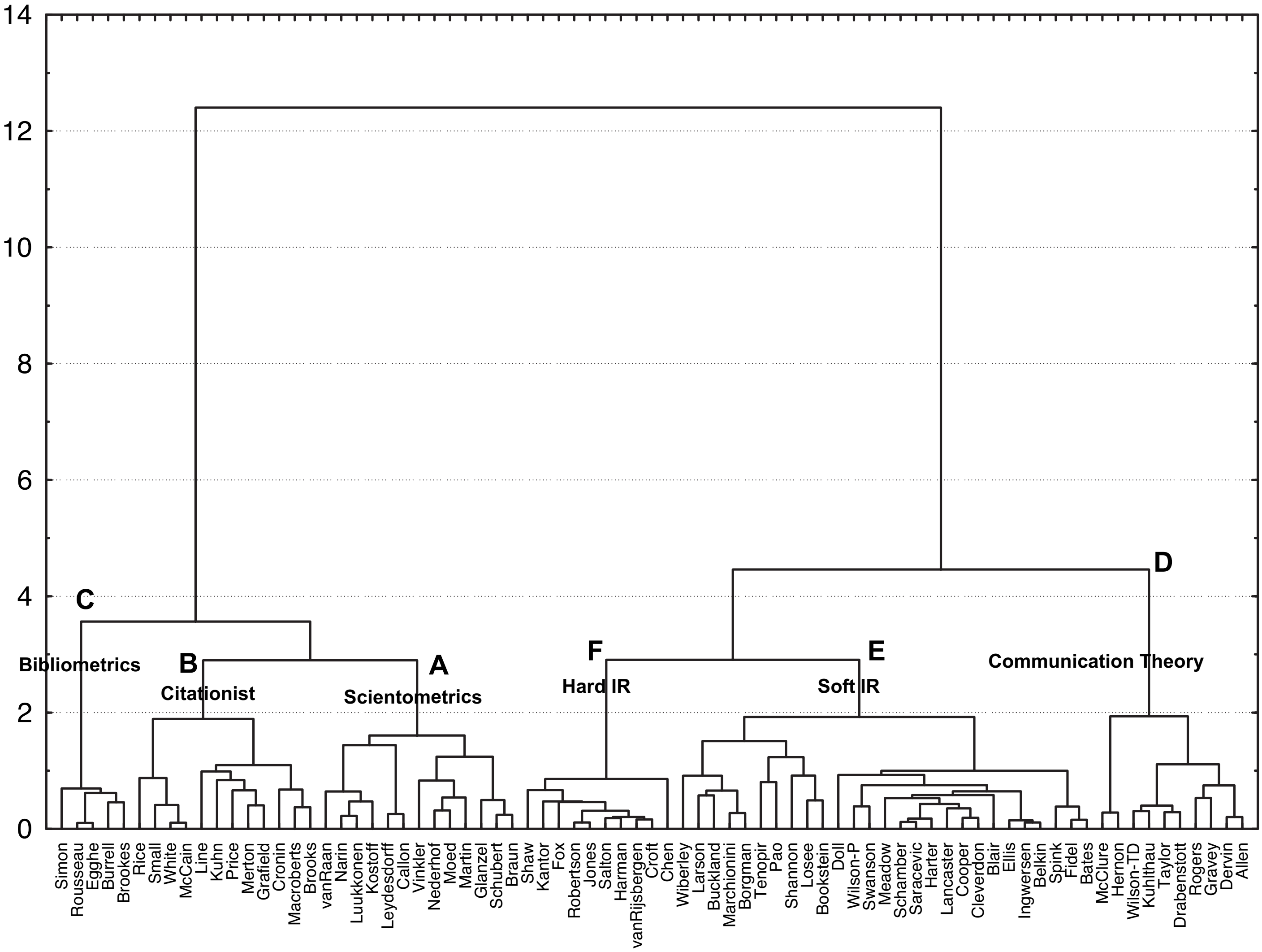

Fig. 3. Clustering of ACA.

the discipline (*New Engl J Med*, *Nature*, *Cell*, *Science*, *Jama*, *Am Psychol*, etc.). They are found concentrated in area C, around the journal *Scientometrics*, the only one that really belongs to our speciality, according to ISI categories. Because no similar phenomenon is seen in the rest of the groups, we can speculate that the area of knowledge dedicated to 'science studies' (bibliometric domain) is the most open and interdisciplinary of the entire field. Moreover, the origin of its components would show it to be 'the most international'.

Also, we would like to point out the disparity between the structures defined in the two maps. Nearly all the authors who appear on the first map have their works published by the journals located in groups A and C of the journal map. Meanwhile, the journals in groups B and D do not seem to have a nucleus of authors who act as intellectual referents in the sense observed for the other groups. This suggests that the methodology of author co-citation analysis, as applied in our study at least, may not necessarily detect all the research fronts in a discipline. For this reason it is advantageous to apply both approaches – ACA and JCA – to this kind of research.

In Figure 5 we can see the MDS of the journals which, like the authors, served to group the different variables. These groupings coincide with those of the clustering. In the bottom area of groups A and B we found the 'technological' journals (*Online*, *Database*, *Cd-Rom Prof*, *Computer* and *Byte*). The two groups that appear to be most closely related (just as in SOM) are

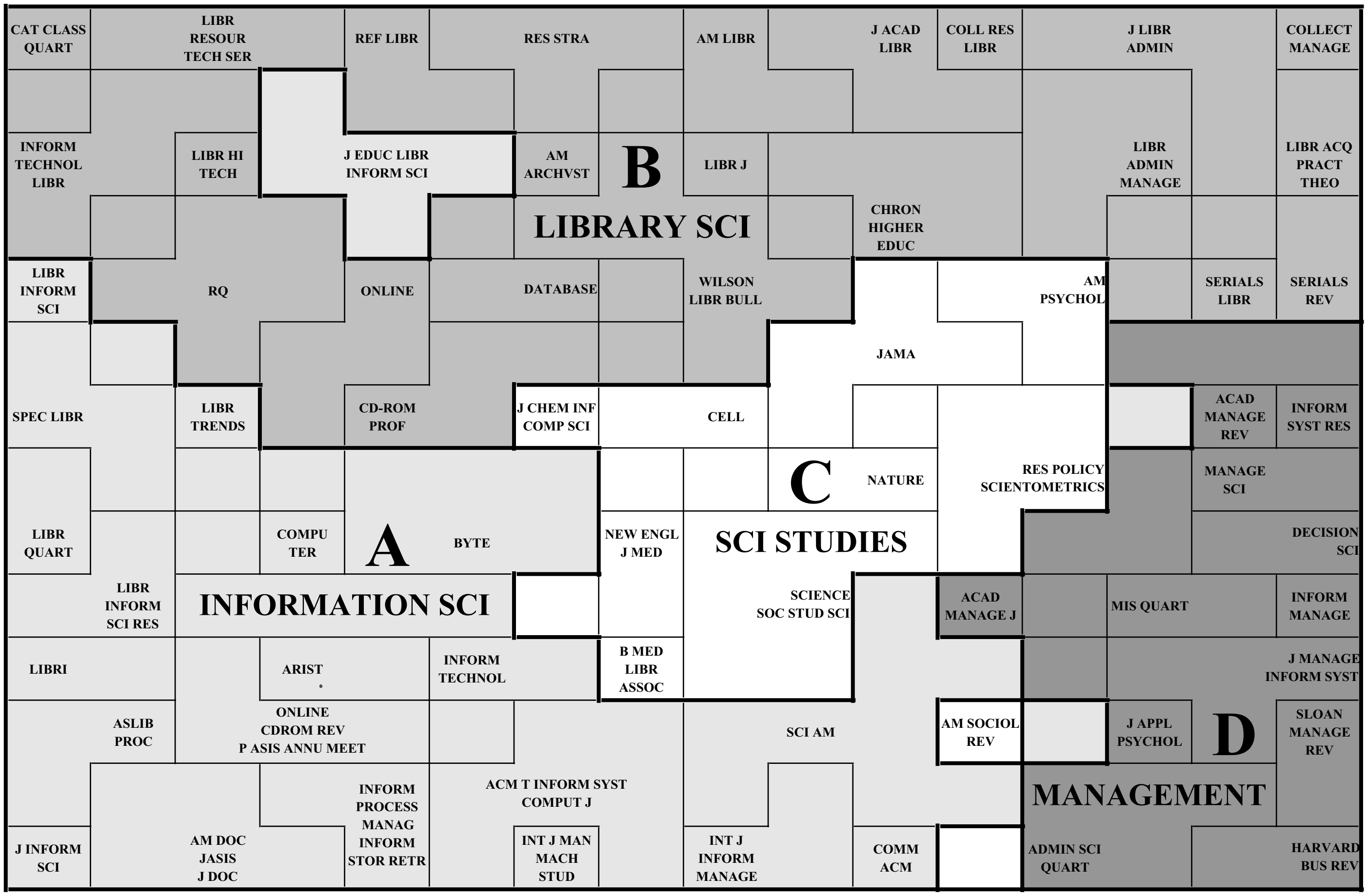

Fig. 4. SOM map of JCA.

A (information science) and B (library science). This effect can be seen more clearly in MDS, owing to their 'intermediate' position in the discipline (*J Educ Libr Inform Sci*, *Libr Trends*, *Libr Inform Sci* and *RQ*). Nonetheless, MDS is better able to reflect the proximity of group A (information science) with D (management). Clustering (Figure 6) shows group D to be far apart from the rest, much more so than MDS would suggest. This journal map has a slightly higher level of stress than that of the authors (0.18), the main difference residing in that the centre of representation is occupied by journals that tend to be intellectual references for the discipline, and belong to the most numerous group, i.e. group A (information science).

Finally, we believe that it is important to point out that the groups from author and journal clustering (Figures 3 and 6) do not coincide for MDS and the SOM, despite being generated from the same database. The journal groupings, which are more generic by nature, appear to be more appropriate than the author groupings in representing the structure of the discipline overall: group A (information science) appears in a central position, quite closely related to B (library science). Groups C (science studies) and D (management) are noteworthy in that most of the publications are not included in the subject category 'library and information science'.

If we compare these with the ACA groupings, we correlate groups E (soft IR) and F (hard IR) with group A (information science) of JCA. On the other hand, groups A (scientometrics), B (citationist), and C (bibliometrics) of ACA are roughly parallel to group C (science studies) of the JCA. Group D (management) of the JCA bears little resemblance to the ACA clustering, which is not surprising given the difference of thematic areas involved. What is remarkable is the absence in the ACA of the groups corresponding to group B (library science) of the JCA. This is a very

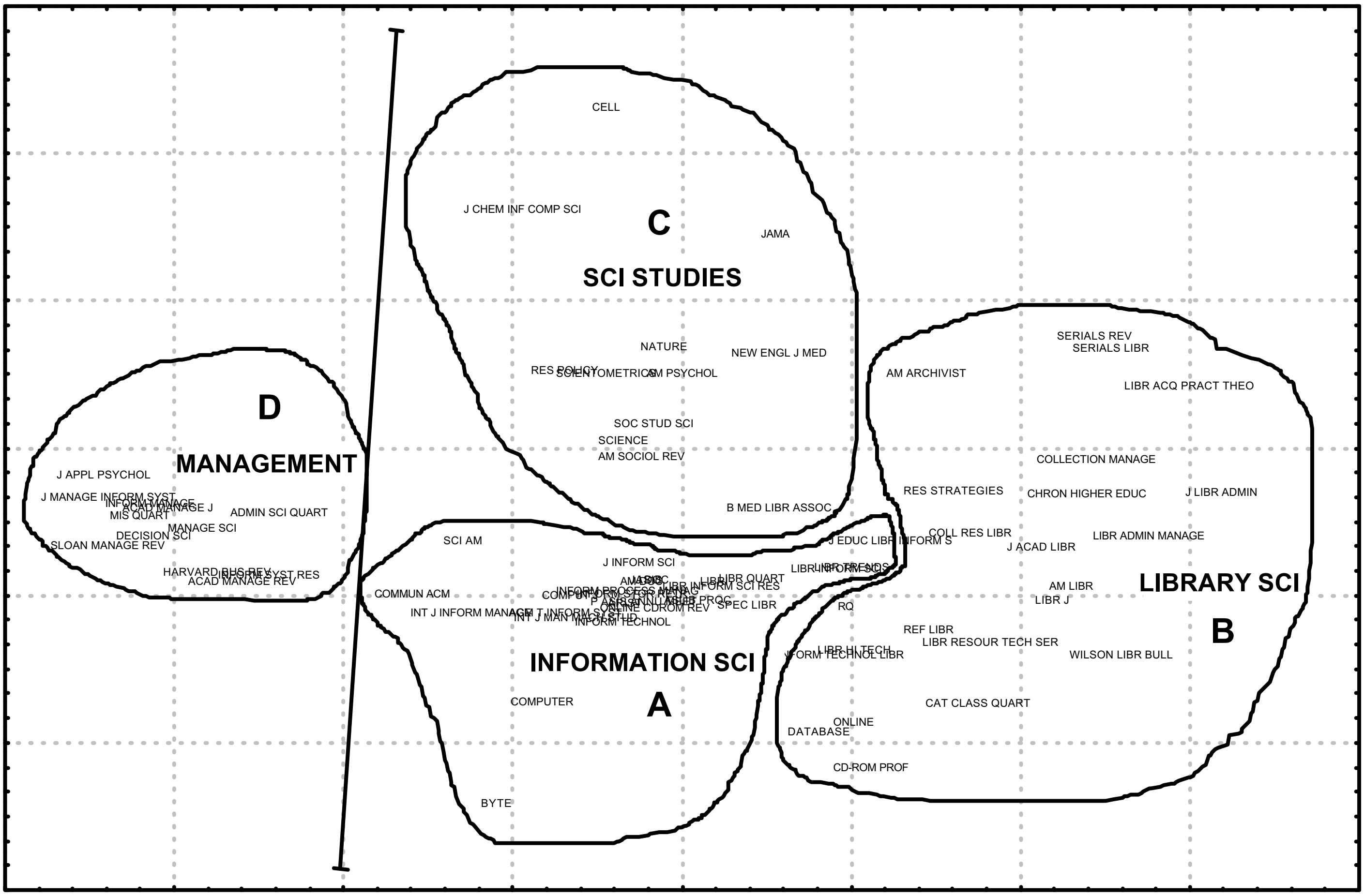

Fig. 5. MDS map of JCA.

important constituent of the LIS category, and the publishing policy of this group is directed precisely at library-related problems. The reason for this absence is that some authors did not reach the threshold number of citations that would make them 'eligible' for ACA representation. Authors who are working on one very specific research area in library science, for example, may cite different bibliographic references; yet this is not the case with bibliometrics (A, B and C) or retrievalists (E and F). Perhaps the most important reference in the library realm is Wilfred Lancaster, who is also a reference for retrievalists, where he finally appears as belonging to research front E (soft IR).

6. Conclusions

The connectionist and multivariate analysis methods (SOM and MDS), as analysed here, allow a reduction of dimension along with clustering. The main difference is that the SOM tries to present a locally corrected projection, whereas MDS attempts to preserve all the distances between the points. That is, MDS is *distance-preserving*, while the SOM is *topology-preserving*.

Generally speaking it is not possible to assign variables to the Cartesian axes in a SOM [25]. It is not the case in MDS, but with ACA it is possible to interpret the vertical axis of MDS (Figure 5). We may assert that the authors found in the lower part of the map are those who most actively use mathematical methods in their studies (group C, bibliometrics; and F, hard IR), while conversely, those in the upper part apply them to a lesser extent (group D, communication theory). In this way, as the map indicates, the vertical axis would show the relationship of the authors with the mathematical methodology (more mathematics and less mathematics).

The quality of a SOM map or an MDS map should be evaluated by experts in the area studied, as no

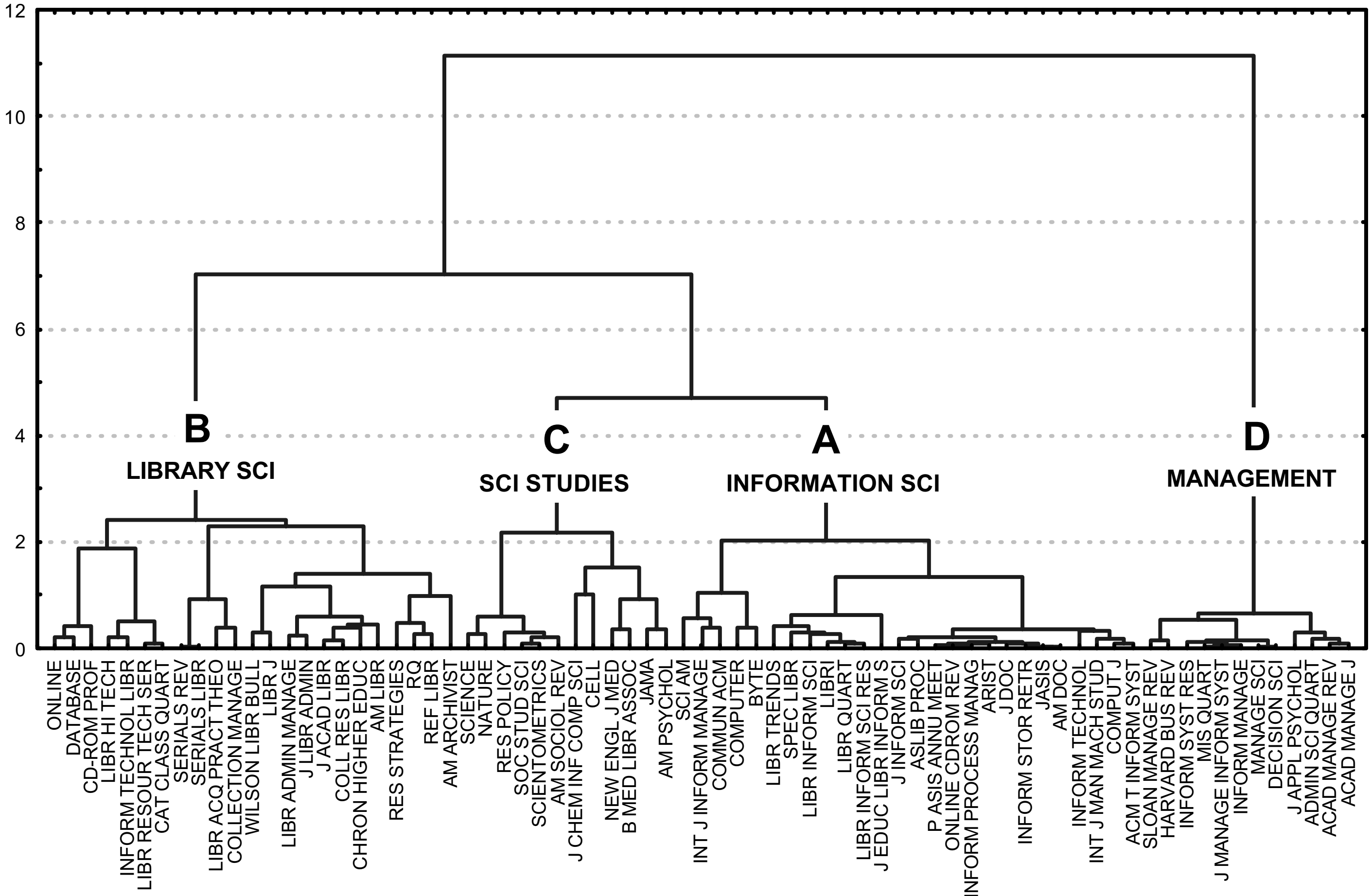

Fig. 6. Clustering of JCA.

objective means exist for assessing unknown domains. This opinion is shared by Tijssen [32], who adds that quality is the most determinant element for credibility. Nonetheless, he offers empirical data to show that the cognitive perception of a group of experts in one subject area with respect to the same map can be very diverse. A quote from Garfield [33] is used to exemplify this problem: 'The ontological status of maps of science or other cognitive maps will perhaps remain speculative until more has been learned about the structure of the brain itself.' Tijssen goes one step further to say that the mental representations of the experts, on an individual micro-level, and the bibliometric maps, on a macro-level, are inherently different. Tijssen concludes that MDS maps may reflect the opinions of experts, yet that additional complementary approaches such as the SOM should be used.

Though we may affirm that the SOM is comparable with MDS, some authors such as Flexer [34, 35] analyse, point by point, the advantages of the SOM for clustering and reducing dimensions. Goodhill [36] emphasizes the importance of the reduction methods that preserve neighbourhood, though only in certain circumstances. Flexer criticizes the SOM as a vague concept for data visualization, while concluding that it cannot replace clustering techniques and MDS, merely complement them. We feel, however, that the potential of the SOM resides in its usefulness not just as a visualization tool, but as a graphic, dynamic interface for a database. This potential is still under study.

In this area, one of the advantages that the SOM would have with respect to MDS is that of perception. The basic postulates of Gestalt theory establish that rectangular forms are more 'natural' for human vision. The schematic SOM representation would therefore be more 'comfortable' to observe [37], and is related with what Kaski calls 'ordered display'. This feature may not be relevant when we are dealing with representations to be analysed by experts, but may be conclusive if we wish to use the map as an interface between a non-expert user and a collection of data.

Although the SOM is capable of clustering, it does not establish the area of influence of each neuron or of the cluster as a whole. Merkl and Rauber [38, 39] developed a technique to help solve this problem: cluster connections (CC). It uses the distance between the vectors of neighbouring neurons to determine whether these neurons belong to the same cluster and should therefore be connected visually, or not. Our methodological approach solved this problem by associating the empty neurons with those authors who had a higher value for activation (then associating a specific cluster). In some cases, very few, this caused some neurons to appear apart from their group; nevertheless, no noise was introduced in the map.

This apparent confrontation between the SOM and MDS as means of dimension reduction seems to be rooted in a more general dichotomy: that of statistical methods as opposed to methods based on connectionist techniques. In reality, this antinomy seems to affect persons more than the methodologies they use. Sarle [40] sums up this conflict in no uncertain terms: 'The marketing hype claims that neural networks can be used with no experience and automatically learn whatever is required; this, of course, is nonsense.'

In summary, the methodologies used allow us to confirm that the subject domains identified in the White and McCain studies are, as well, present in our study for the corresponding period. The appearance of studies pertaining to LS reveals the relationship of this realm with IS. Especially significant is the presence of a management category on our journal maps. From a methodological standpoint, meanwhile, we would agree with those authors who consider MDS, SOM and clustering as complementary methods that provide representations of the same reality from different analytical points of view. Even so, MDS representation is the one offering greater possibilities for the structural representation of the clusters in a set of variables. This approach may be complemented with other kinds of representation based on network analysis [41–43].

References

[1] H. White and K. McCain, Visualizing a discipline: an author co-citation analysis of information science, 1972–1995, *Journal of the American Society for Information Science* 49(4) (1998) 327–55.

[2] E. Tufte, *Envisioning Information* (Graphics Press, Chesire, CT, 1990).

[3] T. Cawkell, Progress in visualisation, *Journal of Information Science* 27(6) (2001) 427–38.

[4] J. Williams, K. Sochats and E. Morse, Visualization, *Annual Review of Information Science and Technology* 30 (1995) 161–207.

[5] H. White and K. McCain, Visualization of literatures, *Annual Review of Information Science and Technology* 32 (1997) 99–169.

[6] B. Hjørland and H. Albrechtsen, Toward a new horizon in information science: domain-analysis, *Journal of the American Society for Information Science* 46(6) (1995) 400–425.

[7] K. Jarvelin and P. Vakkari, The evolution of library and information science 1965–1985: a content analysis of journal articles, *Information Processing and Management* 29(1) (1993) 129–44.

[8] O. Persson, The intellectual base and research fronts of JASIS 1986–1990, *Journal of the American Society for Information Science* 45(1) (1994) 31–8.

[9] T.E. Doszkocs, J. Reggia and X. Lin, Connectionist models and information retrieval, *Annual Review of Information Science and Technology* 25 (1990) 209–60.

[10] X. Polanco, C. François and J.P. Keim, Artificial neural network technology for the classification and cartography of scientific and technical information, *Scientometrics* 41(1–2) (1998) 69–82.

[11] L. Egghe and R. Rousseau, *Introduction to informetrics* (Elsevier, Amsterdam, 1990).

[12] M. Kinnucan, M. Nelson and B. Allen, Statistical methods in information science research, *Annual Review of Information Science and Technology* 22 (1987) 147–78.

[13] F. Moya-Anegón and E. Jiménez-Contreras, Research fronts in library and information science in Spain (1985–1994), *Scientometrics* 42(2) (1998) 229–46.

[14] T. Kohonen, *Self-organizing Maps* (Springer, Berlin, 1997).

[15] T. Honkela, V. Pulkki and T. Kohonen, Contextual relations of words in Grimm tales, analysed by self-organizing map, *Proceedings of the International Conference on Artificial Neural Networks (ICANN-95)* (EC2, Paris, 1995) 3–7.

[16] T. Honkela, S. Kaski, K. Lagus and T. Kohonen, *Newsgroup exploration with WEBSOM method and browsing interface* (Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, 1996) [Technical Report, A32].

[17] T. Honkela, S. Kaski, K. Lagus and T. Kohonen, Self-organizing maps of document collections, *Alma* 1(2) (1996).

[18] S. Kaski and T. Kohonen, Exploratory data analysis by the self-organizing map: structures of welfare and

poverty in the world. In: A. Refenes et.al. (eds), *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets (London, 11–13 October 1995* (World Scientific Publishing, London, 1996) 498–507.

[19] S. Kaski, T. Honkela, K. Lagus and T. Kohonen, Creating an order in digital libraries with self-organizing maps. In: *Proceedings of the World Congress on Neural Networks, WCNN'96* (INNS Press, Mahwah, NJ, 1996) 814–7.

[20] K. Lagus, T. Honkela, S. Kaski and T. Kohonen, Self-organizing maps of document collections: a new approach to interactive exploration. In: E. Simoundis, J. Han and U. Fayyad (eds), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, 1996) 238–43.

[21] X. Lin, Searching and browsing on map displays, *Proceedings of ASIS'1995, 9–12 October, Chicago*, 13–18.

[22] X. Lin, Graphical table of contents, *Digital Library* (March 23–27) 45–53.

[23] X. Lin, Map displays for information retrieval, *Journal of the American Society for Information Science* 48(1) (1997) 40–54.

[24] F. Moya-Anegón, V. Herrero-Solana and V. Guerrero-Bote, La aplicación de redes neuronales artificiales (RNA) a la recuperación de la información, *Anuario SOCADI de Documentación e Información* 2 (1998) 147–64.

[25] S. Kaski, *Data exploration using self-organizing maps* (Thesis for the degree of Doctor of Technology, Helsinki University of Technology, 1997).

[26] H. White, X. Lin and K. McCain, Two modes of automated domain analysis: Multidimensional Scaling vs. Kohonen Feature Mapping of information science authors. In: M. Widad, J. Maniez and S. Pollitt (editors), *Structures and Relations in Knowledge Organization* (Ergon Verlag, Wurzburg, 1998) 57–63.

[27] K. McCain, Mapping economics through the journal literature: an experiment in journal citation analysis, *Journal of the American Society for Information Science* 42(4) (1991) 290–96.

[28] K. McCain, Core journal networks and cocitation maps: new bibliometric tools for serials research and management, *Library Quarterly* 61(3) (1991) 311–36.

[29] K. McCain, Biotechnology in context: a database-filtering approach to indentifying core and productive non-core journals supporting multidisciplinary R&D, *Journal of the American Society for Information Science* 46(4) (1995) 306–17.

[30] K. McCain, Neural networks in context: a longitudinal journal cocitation analysis of an emerging interdisciplinary field, *Scientometrics* 41(3) (1998) 389–410.

[31] J.M. Campanario, Using neural networks to study networks of scientific journals, *Scientometrics* 33(1) (1995) 23–40.

[32] R.J.W. Tijssen, A scientometric cognitive study of neural network research: expert mental maps versus bibliometric maps, *Scientometrics* 28(1) (1993) 111–36.

[33] E. Garfield, M.V. Malin and H. Small, Citation data as science indicators. In: Y.J. Elkana et al. (eds), *Toward a Metric of Science* (Wiley, Chichester, 1978).

[34] A. Flexer, *Limitations of Self-organizing Maps for Vector Quantization and Multidimensional Scaling* (The Austrian Research Institute, Vienna, 1996) [Technical Report: oefai-tr-96-23].

[35] A. Flexer, Connectionists and statisticians, friends or foes? In: J. Mira and F. Sandoval (eds), *Proceedings of the International Workshop on Artificial Neural Networks (IWANN'95), Spain, 1995* (Springer, 1995) 454–61.

[36] G. Goodhill and T. Sejnowski, Quantifying neighbourhood preservation in topographic mappings. In: *Proceedings of the 3rd Joint Symposium on Neural Computation, Pasadena, 1996* (California Institute of Technology, Pasadena, 1996) 61–82.

[37] J. Costa, *La esquemática: visualizar la información* (Paidós, Barcelona, 1998).

[38] D. Merkl and A. Rauber, Alternative ways for cluster visualization in self-organizing maps. In: T. Kohonen (ed.), *Proceedings of the Workshop on Self-Organizing Maps (WSOM'97), Helsinki, June 4–6, 1997* (Helsinki University of Technology, Espoo, 1997) 106–11.

[39] D. Merkl and A. Rauber, Cluster Connections: a visualization technique to reveal boundaries in self-organizing maps. In: M. Marinaro and R. Tagliaferri (eds), *Proceedings of the 9th Italian Workshop on Neural Nets (WIRN'97), Vietri sul Mare, Italy, 1997* (Springer, London, 1998).

[40] W.S. Sarle, Neural networks and statistical models. In: *Proceedings of the Nineteenth Annual SAS Users Group International Conference* (SAS Institute, Cary, 1994) 1538–50.

[41] H. White, Author-centered bibliometrics through CAMEOs: characterizations automatically made and edited online, *Scientometrics* 51(3) (1995) 607–37.

[42] C. Chen, Bridging the gap: the use of pathfinder networks in visual navigation, *Journal of Visual Languages and Computing* 9 (1998) 267–86.

[43] C. Chen, Visualising semantic spaces and author co-citation networks in digital libraries, *Information Processing and Management* 35 (1999) 401–20.

Appendix A: Selected journals

Abbreviated name	Full name	ISSN
ARIST	*Annual Review of Information Science and Technology*	0066-4200
COLL RES LIBR	*College & Research Libraries*	0010-0870
INFORM MANAG	*Information Management*	0019-9966
INFORM PROC MANAG	*Information Processing & Management*	0306-4573
INT J INF MANAG	*International Journal of Information Management*	0268-4012
J DOC	*Journal of Documentation*	0022-0418
J INF SCI	*Journal of Information Science*	1352-7460
JASIS	*Journal of the American Society for Information Science*	0002-8231
LIBR HI TECH	*Library Hi Tech Journal*	0737-8831
LIBR INFORM SCI RES	*Library & Information Science Research*	0740-8188
LIBR QUART	*Library Quaterly*	0024-2519
LIBR RES TECH SERV	*Library Resources & Technical Services*	0024-2527
LIBR TRENDS	*Library Trends*	0024-2594
ONLINE	*Online*	0146-5422
RQ	*Reference Quarterly*	0033-7072
SCIENTIST	*Scientist*	0890-3670
SCIENTOMETRICS	*Scientometrics*	0138-9130

Appendix B: Rejected journals

Abbreviated name	Full name	ISSN
BULL MED LIB ASSOC	*Bulletin of the Medical Library Association*	0025-7338
INFORM SYST RES	*Information Systems Research*	1047-7047
INT J GEOGR INF SYST	*International Journal of Geographical Information Systems*	0269-3798
INTERLENDING DOC SUPPLY	*Interlending and Document Supply*	0264-1615
J AM MED INFORM ASSN	*Journal of the American Medical Informatics Association*	1067-5027
TELECOMMUN POLICY	*Telecommunications Policy*	0308-5961
KNOWL ACQUIS	*Knowledge Acquisition*	1042-8143

Appendix C: Full list of mapped journals

No.	Abbreviated title	Title	ISSN	Country	ISI subject category
1	ACAD MANAGE J	*Academy of Management Journal*	0001-4273	USA	Management; Business
2	ACAD MANAGE REV	*Academy of Management Review*	0363-7425	USA	Management; Business
3	ACM T INFORM SYST	*ACM Transactions on Information Systems*	1046-8188	USA	Computer Science
4	ADMIN SCI QUART	*Administrative Science Quarterly*	0001-8392	USA	Management; Business
5	AM ARCHIVIST	*American Archivist*	0360-9081	USA	Information Science & Library Science; History
6	AM DOC	*American Documentation*	Continued by: *Journal of the American Society for Information Science*		
7	AM LIBR	*American Libraries*	0002-9769	USA	Non ISI journal
8	AM PSYCHOL	*American Psychologist*	0003-066X	USA	Psychology
9	AM SOCIOL REV	*American Sociological Review*	0003-1224	USA	Sociology
10	ARIST	*Annual Review of Information Science and Technology*	0066-4200	USA	Information Science & Library Science
11	ASLIB PROC	*Aslib Proceedings*	0001-253X	England	Information Science & Library Science

Continued

Appendix C: *Continued*

No.	Abbreviated title	Title	ISSN	Country	ISI subject category
12	B MED LIBR ASSOC	*Bulletin of the Medical Library Association*	0025-7338	USA	Information Science & Library Science
13	BYTE	*Byte*	0360-5280	USA	Computer Science
14	CAT CLASS QUART	*Cataloging & Classification Quarterly*	0163-9374	USA	Non ISI journal
15	CD-ROM PROF	*CD-Rom Professional*	1049-0833	USA	Information Science & Library Science; Computer Science
16	CELL	*Cell*	0092-8674	USA	Biochemistry & Molecular Biology; Cell Biology
17	CHRON HIGHER EDUC	*The Chronicle of Higher Education*		USA	Non ISI journal
18	COLL RES LIBR	*College & Research Libraries*	0010-0870	USA	Information Science & Library Science
19	COLLECTION MANAGE	*Collection Management*	0146-2679	USA	Non ISI journal
20	COMMUN ACM	*Communications of the ACM*	0001-0782	USA	Computer Science
21	COMPUT J	*Computer Journal*	0010-4620	England	Computer Science
22	COMPUTER	*Computer*	0018-9162	USA	Computer Science
23	DATABASE	*Database*	0162-4105	USA	Information Science & Library Science
24	DECISION SCI	*Decision Sciences*	0011-7315	USA	Management
25	HARVARD BUS REV	*Harvard Business Review*	0017-8012	USA	Management; Business
26	INFORM MANAGE	*Information Management*	0019-9966	Holland	Computer Science
27	INFORM PROCESS MANAG	*Information Processing & Management*	0306-4573	USA	Information Science & Library Science
28	INFORM STOR RETR	*Information Storage and Retrieval*	Continued by: Information Processing & Management		
29	INFORM SYST RES	*Information Systems Research*	1047-7047	USA	Information Science & Library Science
30	INFORM TECHNOL	*Information Technology*	0971-233X	India	Non ISI journal
31	INFORM TECHNOL LIBR	*Information Technology and Libraries*	0730-9295	USA	Information Science & Library Science
32	INT J INFORM MANAGE	*International Journal of Information Management*	0268-4012	England	Information Science & Library Science
33	INT J MAN MACH STUD	*International Journal of Man Machine Studies*	Continued by: International Journal Of Human-Computer Studies		
34	J ACAD LIBR	*The Journal of Academic Librarianship*	0099-1333	USA	Information Science & Library Science
35	J APPL PSYCHOL	*Journal of Applied Psychology*	0021-9010	USA	Psychology, Applied
36	J CHEM INF COMP SCI	*Journal of Chemical Information and Computer Sciences*	0095-2338	USA	Computer Science
37	J DOC	*Journal of Documentation*	0022-0418	England	Information Science & Library Science
38	J EDUC LIBR INFORM SCI	*Journal of Education For Library and Information Science*	0748-5786	USA	Information Science & Library Science; Education & Educ. Research
39	J INFORM SCI	*Journal of Information Science*	0165-5515	England	Information Science & Library Science
40	J LIBR ADMIN	*Journal of Library Administration*	0193-0826	USA	Non ISI journal
41	J MANAGE INFORM SYST	*Journal of Management Information Systems*	0742-1222	USA	Non ISI journal
42	JAMA	*Journal of the American Medical Association*	0098-7484	USA	Medicine, General & Internal
43	JASIS	*Journal of the American Society for Information Science*	0002-8231	USA	Information Science & Library Science
44	LIBR ACQ PRACT THEO	*Library Acquisitions: Practice and Theory*	Continued by: Library Collections, Acquisitions, and Tech. Services		
45	LIBR ADMIN MANAGE	*Library Administration and Management*	0888-4463	USA	Non ISI journal
46	LIBR HI TECH	*Library Hi Tech*	0737-8831	USA	Information Science & Library Science
47	LIBR INFORM SCI	*Library and Information Science*	0373-4447	Japan	Information Science & Library Science
48	LIBR INFORM SCI RES	*Library & Information Science Research*	0740-8188	USA	Information Science & Library Science
49	LIBR J	*Library Journal*	0363-0277	USA	Information Science & Library Science
50	LIBR QUART	*Library Quarterly*	0024-2519	USA	Information Science & Library Science

Appendix C: *Continued*

No.	Abbreviated title	Title	ISSN	Country	ISI subject category
51	LIBR RESOUR TECH SER	*Library Resources & Technical Services*	0024-2527	USA	Information Science & Library Science
52	LIBR TRENDS	*Library Trends*	0024-2594	USA	Information Science & Library Science
53	LIBRI	*Libri*	0024-2667	Denmark	Information Science & Library Science
54	MANAGE SCI	*Management Science Series A-Theory*	****-****	USA	Management
55	MIS QUART	*MIS Quarterly*	0276-7783	USA	Management
56	NATURE	*Nature*	0028-0836	England	Multidisciplinary Sciences
57	NEW ENGL J MED	*New England Journal of Medicine*	0028-4793	USA	Medicine, General & Internal
58	ONLINE	*Online*	0146-5422	USA	Information Science & Library Science
59	ONLINE CDROM REV	*Online & CD-ROM Review*	1353-2642	USA	Information Science & Library Science
60	P ASIS ANNU MEET	*Proceedings of the ASIS Annual Meeting*	0044-7870	USA	Information Science & Library Science
61	REF LIBR	*Reference Librarian*	0276-3877	USA	Non ISI journal
62	RES POLICY	*Research Policy*	0048-7333	Holland	Planning & Development
63	RES STRATEGIES	*Research Strategies*	0734-3310	USA	Non ISI journal
64	RQ	*RQ*	0033-7072	USA	Information Science & Library Science
65	SCI AM	*Scientific American*	0036-8733	USA	Multidisciplinary Sciences
66	SCIENCE	*Science*	0036-8075	USA	Multidisciplinary Sciences
67	SCIENTOMETRICS	*Scientometrics*	0138-9130	Holland	Information Science & Library Science
68	SERIALS LIBR	*Serials Librarian*	164497XXX	USA	Non ISI journal
69	SERIALS REV	*Serials Review*	1645432XX	England	Non ISI journal
70	SLOAN MANAGE REV	*Sloan Management Review*	0019-848X	USA	Management; Business
71	SOC STUD SCI	*Social Studies of Science*	0306-3127	USA	History & Philosophy of Science
72	SPEC LIBR	*Special Libraries*	0038-6723	USA	Information Science & Library Science
73	WILSON LIBR BULL	*Wilson Library Bulletin*	0043-5651	USA	Information Science & Library Science