Showing the Essential Science Structure of a Scientific Domain and its Evolution

Benjamin Vargas-Quesada, Félix de Moya-Anegón, Zaida Chinchilla-Rodríguez and Antonio González-Molina

Information Visualization 2010 9: 288
DOI: 10.1057/ivs.2009.33

Original Article

Showing the essential science structure of a scientific domain and its evolution

Benjamin Vargas-Quesada[a,b,*]
Félix de Moya-Anegón[c]
Zaida Chinchilla-Rodríguez[c]
and Antonio González-Molina[b]

[a]Library and Information Science, University of Granada Communication and Information Science Faculty, Colegio Maximo de Cartuja, Granada 18071, Spain.
E-mail: benjamin@ugr.es
[b]CSIC, Unidad Asociada Grupo SCImago, 18071 Granada, Spain.
[c]CSIC/CCHS/IPP, 28037 Madrid, Spain.

*Corresponding author.

Abstract Category cocitation and its representation through social networks is proving to be a very adequate technique for the visualization and analysis of great scientific domains. Its combination with pathfinder networks using pruning values $r=\infty$ and $q=n-1$ makes manifest the essence of research in the domain represented, or what we might call the `most salient structure'. The possible loss of structural information, caused by aggressive pruning in peripheral areas of the networks, is overcome by creating heliocentric maps for each category. The depictions obtained with this procedure become tools of great usefulness in view of their capacity to reveal the evolution of a given scientific domain over time, to show differences and similarities between different domains, and to suggest possible new lines for development. This article presents the *scientogram* of the United States for the year 2002, identifying its essential structure. We also show the scientograms of China for the years 1990 and 2002, in order to study its particular national evolution. Finally, we try to detect patterns and tendencies in the three scientograms that would allow one to predict or flag the evolution of a scientific domain.
Information Visualization (2010) **9**, 288–300. doi:10.1057/ivs.2009.33; published online 24 December 2009

Keywords: information visualization; ISI categories cocitation; scientific domains; network visualization; network evolution; pathfinder network

Introduction

In Moya Anegón *et al*,[1] we proposed the cocitation of ISI[1] Journal Citation Report categories as a new technique for building maps of vast scientific domains. Later, in Vargas-Quesada[2] we introduced Pathfinder Networks (PFNETs)[3] as the pruning algorithm in order to show the most 'salient' structure of a domain, as well as using non-normalized cocitation values to give rise to the grouping of ISI categories in major thematic areas, as clusters. More recently, our work has brought us to use this same methodology for the visualization and analysis of the scientific structure of the worldwide scientific domain,[4] and for the Spanish domain.[5] This allows us to clearly identify the macrostructure, the microstructure and the marrow of research. This methodology, as it evolves, may also be applied to the comparison of scientific domains, to show their evolution and even their developmental trends, as we shall see below. The objective pursued is to visualize the scientific structure of a developed country like the United States, and then compare it with that of a quickly developing nation such as China – which has undergone a strong increase in scientific production over the past decade – so as to detect trends and patterns of scientific connection. The choice of the year 2002 owes simply to the fact that this was the nearest complete year for

Received: 29 June 2009
Revised: 23 October 2009
Accepted: 9 November 2009

[1] Actually registered as Thomson Reuters.

which we could download documents at the time we began to experiment with this notion (in June of 2003); and 1990, because it struck us as a date remote enough to reveal changes, but still within the realm of our current understanding and interpretations.

Below we offer an overview of work related with this aim, and a succinct description of the methodology that we applied. Then, we show the results obtained in the form of scientograms, which is how we chose to denominate the graphic representations obtained from scientometric information, to finally draw a few brief conclusions.

Related Works

The generation of a *big picture* is something implicit in the process of visualizing scientific information. Ever since the earliest work involving bibliometrics and research, at the beginning of the 1960s, Garfield[6] and Garfield *et al*[7] began to construct historical maps based on citation. Shortly thereafter, Price[8] showed that the patterns of author citation of scientific articles served to define research fronts and that they could moreover be used to reflect the scientific structure of a domain. In the 1970s, Small and Griffith[9] and Griffith *et al*[10] represented the specialized fields of the natural sciences, demonstrating that science is a network of fields that are mutually interconnected. A year later, Aaronson[11] 'x-rayed' biomedicine over a 2-year period and showed its evolution. With the arrival in the 1980s of the generation of visualizations or maps of scientific domains, this sort or work became much more prolific: we see the maps of biochemistry and molecular biology,[12] biotechnology and molecular genetics,[13] biochemistry, immunology and animal and plant biology[14] and, finally, pharmacology.[15] In the meantime, Small, continued to fine-tune the techniques used for his first maps.[16–21,27,23]

In the 1900s, new information retrieval methods come onto the scene, as well as new techniques for the analysis, visualization and spatial positioning of information,[24] and there is a proliferation of work based on the visualization of the structure of small scientific domains oriented to the classification and/or retrieval of information. Thus for instance, the Centre for Science and Technology Studies[25] research involves the generation of maps of science with emphasis on their structural and dynamic aspects.[26–28,5] Lin *et al*[29] develop a self-organizing map (SOM) that can be used as a bibliographic interface for the retrieval of information online. White and McCain[30,31] propose visualization as a model for information analysis and retrieval. Garfield[32] advocates the use of new visualization techniques for the generation of global-sequential maps of science. White *et al*[33] compare visualizations obtained through Multidimensional Scaling to those created by the SOM, concluding that results are very similar but that the SOM are better at integrating bibliographic information plus retrieval. Chen[34] incorporates PFNETs into the field of Documentation in order to prune links when visualizing social networks.[35] For Ding *et al*,[36] visualization favors the simplification of an area of knowledge down to its main elements, as well as its use for a better comprehension of the domain on the part of the user.

The new millennium incorporates the challenge of dynamically generating maps that can be used as interfaces for the access and retrieval of information. Merton[37] argues that what was once conceived by Garfield to retrieve information is, in fact, a magnificent tool for the study of the sociology of science. White[38] proposes using networks centered on a subject (CAMEOs) as interfaces for the access and retrieval of bibliographic information for non-expert users, putting forth as well the possibility that these depictions may be generated in a dynamic manner. Noyons *et al*[39]; Buter and Noyons[40] and Noyon[41] analyze the use of maps as metaphors of a scientific discipline, their use as interfaces and their limitations, providing some solutions that allow for a better exploration of the domain in question. Chen and Paul[42] and Chen *et al*[43] manage to represent the structural patterns of scientific literature in 3D maps. Ding *et al*[44] show the intellectual structure of the field of information retrieval, indicating models, patterns and trends. Guerrero Bote *et al*[45] use a SOM to classify, browse and retrieve information. White *et al*[46] implement a dynamic system for visualization: *Authorlink*, based on author cocitation, which allows browsing and information retrieval, in real time.[47,48] Small[49] theorizes about the design of a web tool capable of detecting and monitoring changes in research fronts of an area in real time. Chen and Kuljis[50] study the appearance and evolution of research fronts in the field of Physics. Again Chen, but this time with Morris,[51] compares the visualization of citation networks derived using two reduction algorithms: minimum spanning trees (MSTs) and PFNETs. Morris *et al*[52] work on the visualization, detection and identification of temporal changes in the lines of research. Boyack and Börner[53] study the relationship between government financing and the number of citations received. The SCImago group[1,54] proposes ISI category cocitation as units of analysis and representation for the generation of maps of vast scientific domains, and compares three of these. In that same year, Boyack *et al*[55] present a map representing the structure of science on the whole, providing a bird's-eye view of today's scientific landscape. Later, Klavans and Boyack[56,57] propose a framework for a quantitative assessment of the performance of relatedness measures and visualization algorithms, a method for generating maps based on the relationship between hundreds of thousands of documents, and quantitative techniques for evaluating these vast maps. Samoylenko *et al*[58] propose an approach to visualizing the scientific world and its evolution by constructing MSTs and a 2D maps of scientific journals. Finally, Leydesdorff[59] combines the *Journal Citation Report* (*JCR*) of the *Science Citation Index* 2004 and the *Social Science Citation Index* 2004, in order to map journals and specialties.

In an attempt to sum up what has taken place to date, we can say that nowadays there are two proposals for tracking down the *big picture*. One adopts the traditional units of analysis (authors, documents and journals) and, through their grouping, scientific disciplines are identified following a bottom-up process. The alternative uses the ISI categories to the same end, and shows the scientific structure from them in a top-down manner. The first proposal presents all the pros of fine-grained character, but runs into difficulties in representing the totality of the panorama on a single plane, and in tagging the disciplines. The second option has its strong points where the former shows weaknesses, and vice versa. That is, it is relatively simple to represent on a single plane the scientific structure of a domain by means of a maximum of 218 categories and their interrelation, and problems with tagging are non-existent. Here, the difficulties revolve around how to descend to smaller units of analysis, departing from the categories, with no loss of information. We consider that both the former and the latter are valid alternatives for the achievement of the *big picture*, yet we decided to experiment in this case with the latter to delimit scientific disciplines. This means that the work presented here is based on the assumption that the ISI databases represent a certain scientific reality that can be used to recreate the structure of science. It is also acknowledged, however, that the database is fuzzy, that journal categories overlap and that they may be sometimes erroneous.

Material and Methods

For strictly investigative purposes, we downloaded from the ISI Web of Science® (Thomson Reuters)[60] and more precisely from the SCI, SSCI and A&HCI databases, all the records for the year 2002 that contained 'USA' in the address field. This gave us a total of 316 878 documents (comprising articles, biographical items, book reviews, corrections, editorial materials, letters, meeting abstracts, news items and reviews). Likewise, we downloaded the records of scientific publication involving China for the years 1990 and 2002, obtaining 9603 and 58 981 documents, respectively.

Units of analysis

As we indicated above, we resorted to the ISI categories as units of analysis and representation. Each category agglutinates the journals that were categorized under that name, and likewise, the documents that were published in those journals. We do not use the category *Multidisciplinary Sciences*, and so all the documents ascribed to it were recategorized. The problem of recategorization of documents published in multidisciplinary journals has been dealt with in depth by Glänzel *et al*.[61,62] The solution they propose is to recategorize each one of those documents in view of the most referenced category. We

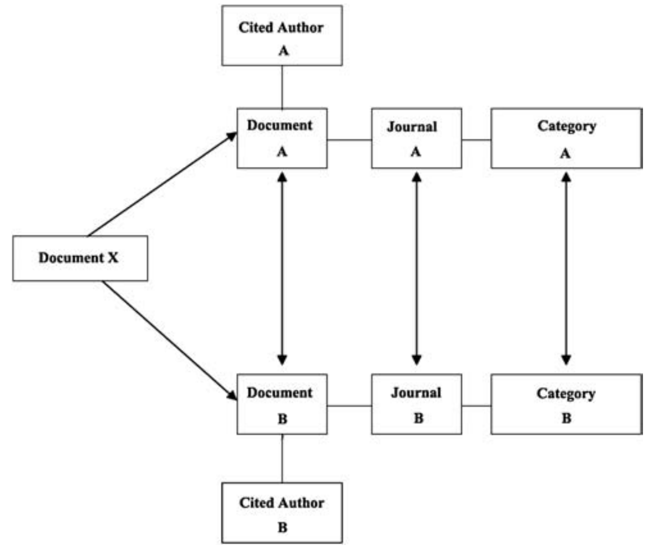

Figure 1: Cocitation scheme.

adopted this procedure, with very satisfactory results: only a few documents had to be recategorized manually because there was a lack of references. However, recategorization of multidisciplinary documents on the basis of the predominant category of the citing documents is an alternative with which we are now experimenting and we may possibly incorporate in the near future.

Unit of measure

As in previous works, we adopted cocitation as a unit of measure. Given this basis, we revisited the traditional scheme of cocitation and added a new element to it. Accordingly, if one document published in a journal with an ISI category in the *JCR* is cocited along with another document that has been published in a different journal having another ISI–*JCR* category, we can state that there is a relationship between these categories. The more cocitations there are, the stronger the relationship between the categories (see Figure 1).

Just like White,[63] we consider that PFNETs derived from raw cocitation counts appear to form more readily interpreted structures than do their normalized versions. And like Leydesforff and Vaugahn,[64] we have confirmed that the standardization of relational information causes distortions in the values of the units of analysis when cocitation is the unit of measure, making them less coherent as compared with raw data. Chen[51] does not agree with this idea, especially when the temporal factor is taken into account. Of similar opinion are Klavans and Boyack[56,57] who, after testing eight similarity measures, propose K50, a modified cosine index used as a unit of measure with quite good results. In our view, the standardization of relational information is necessary in units of analysis such as authors and documents. A limited capacity to concentrate citations in a short time period, for example one year, makes them particularly unstable units for temporal studies. Yet this is not the case of the ISI categories, which show great stability due to their greater capacity of agglutination.

Dimensionality reduction

The display of a scientific domain involving such a high number of units, yet easily identified by tags, and showing interactions by means of links, all in an intelligible and esthetically pleasing form for the human eye … is a formidable challenge. Bearing in mind the precautionary message of Hjørland and Albrechtsen:[65]

If users are provided with a system of too many possibilities, without giving priority to the essential connections, the user is overloaded, and the system is ineffective.

Then there is the advice of Small:[16]

Despite the loss of structural information … the gain in simplicity may for some purposes be worth the sacrifice.

And we fully agree again with White[63] in that:

Among techniques, two-dimensional PFNETs *made with raw cocitation counts, and visualized through spring embedders, appear to have considerable advantages.*

Although PFNETs has been used in the fields of bibliometrics, infometrics and scientometrics since 1990,[66] its application to citation was the work of Chen,[35,67] who introduced a new form of organizing, visualizing and accessing information. PFNETs is based on two elements – the Minkowski distance and on an extension of triangular inequality – and is defined by two parameters: r, associated with the Minkowski distance used; and q, related to length, understood as the number of links of the paths that are compared. Therefore, all the links that defy triangular inequality, having one associated distance that is lesser than another for the same points composed up to q links, and with the calculation of this global distance by means of the parametrical Minkowski equation with parameter r, will be eliminated.[4]

In our opinion, PFNETs set to the strongest pruning configuration $r = \infty$, and $q = n - 1$, is the prime option for preserving and highlighting the salient relationships between categories, and for capturing the essential underlying intellectual structure of a scientific domain in an economical way.

Scalar

There are many different methods for the automatic generation of graphs. The spring embedder type is most widely used in the area of Documentation. Spring embedders begin by assigning coordinates to the nodes in such a way that the final graph will be pleasing to the eye.[68] One of the most noteworthy of this type is the Kamada and Kawai algorithm.[69] Its foremost features are the capacity to minimize differences with respect to theoretical distances in the entire graph, good computation times and the fact that it subsumes multidimensional scaling when the technique of Kruskal and Wish[70] is applied. As Cohen[71] and Krempel[72] indicate, the Kamada and Kawai algorithm uses an energy similar to the *stress* of multidimensional scaling as the measure for adaptation to theoretical distances.

The combination of PFNETs, cocitation, and Kamada and Kawai makes the most interdisciplinary elements of a depiction tend to situate themselves toward the center, as a result of the greater number of links. This creates an informational and intuitive effect that enhances analysis and interpretation.

Scientogram validation

We finally resorted to a method based on a statistical process – Factor Analysis (FA) – for our validation of findings. Its main features are:

- FA is conducted on the basis of raw data cocitation,
- the number of factors identified is extracted,
- each factor is tagged, and
- The factors identified are transferred to the scientogram.

We stopped extracting factors upon arriving at an eigenvalue greater or equal to one,[2] which was done with the Scree test.[3] In order to capture the nature of each factor so as to tag it, the factors were first ordered according to their index of weight – factor loading – in a decreasing order, and a cutoff of 0.5 was established for membership; though for denomination, we took into account only those categories of each factor that had a value of 0.7 or more. In order that each one of the subject areas, along with the categories integrating it, can be easily identified, all the categories comprising a common factor were given the same color. Thus, for instance, the categories identified in Factor 1 (Biomedicine) appear in Light Purple, while those of Factor 2 (Psychology) are colored Emerald Green, etc. Those that belong to more than one subject area are red, the 'hot' points of interaction among the subject areas. Finally, dark grey shows the 'cold' ones that were not identified by FA and therefore belong to no subject area.

Results

Applying the methodology described here to the ISI data gives us a network of categories whose form is reminiscent of a human neuron, with a great axon or central neurite. In order to favor interpretation of the scientogram – which

[2] This simple criterion works quite well, giving results much in accord with the expectations of researchers (Ding *et al*).[44]

[3] The *scree* test consists of the examination of the line obtained in the graphic representation of the eigenvalues of the identified factors. The extraction of factors comes to a halt when the line of eigenvalues begins to level out, practically forming a line parallel to the axis, with hardly any slant.

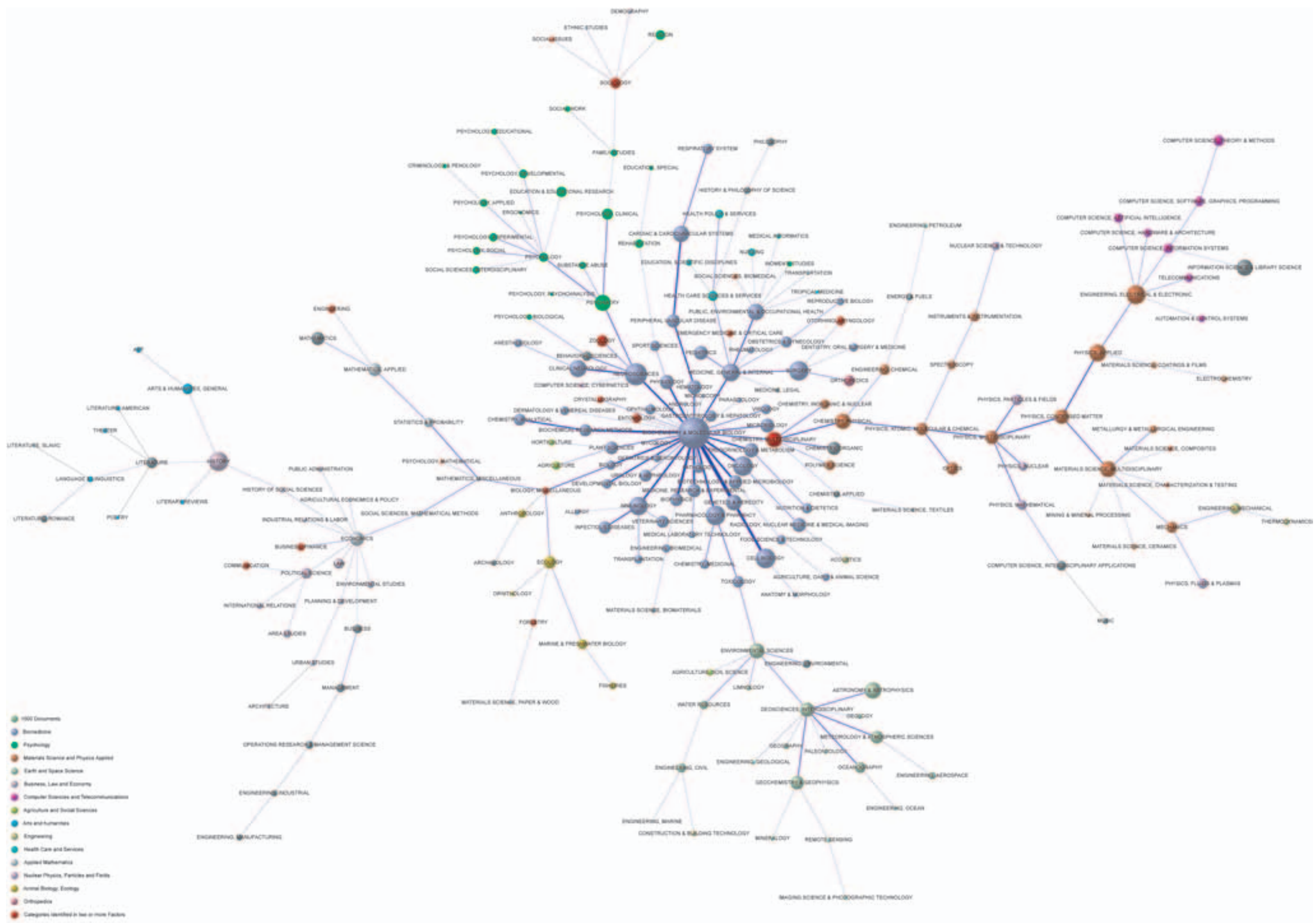

Figure 2: Scientogram of USA 2002.

is how we prefer to denote this type of representation built from scientometric information – each sphere was labeled with the name of the ISI–*JCR* category it represents, and was given a size directly proportional to the volume of documents it comprises. To help visually establish the relationship between the size of each category and its actual output, in the lower left corner of the scientogram there is a sphere of reference with a size reflecting the specified number of documents. The lines that connect the different spheres show the salient relations of cocitation among categories. These associations are thicker or thinner depending on the intensity of cocitation: the greater the intensity, the greater the thickness of the link. They represent, then, the salient consensual opinion of authors of documents as expressed by means of their use of citation.

Brief description of a domain

The scientogram in Figure 2 shows the synthesis of US Scientific output and its interrelations for the year 2002. Using FA, a total of 14 factors were identified. Each factor or thematic area was assigned a color and a number, listed in the lower left part of the scientogram.

The US scientogram stands as the model of science of a developed nation. The basic features, from a macrostructural viewpoint, can be summed up as: a central zone featuring what we might call Biomedical Sciences and Earth Sciences; toward the right are the Hard Sciences; and toward the left, the Soft Sciences are configured. This scheme of vertebration on the macrostructural level of science is a classic arrangement, which persists in the scientograms of developed countries, and shares little in common with the incipient backbones of other less-developed countries.[4]

The US scientogram shows a domain well advanced in research, with a structure or backbone typical of countries with a high level of socioeconomic well-being. Biomedicine occupies a central position, indicating the importance that its research and development hold for the community, sustained on its right side by research in science and technology, and by the social and human sciences to the left. This interdisciplinary position of biomedicine in the United States coincides with that detected in the Spanish domain of science in the year 2000,[1] where the Anep[73] classification was used to concentrate all the ISI categories into 24 areas of knowledge. The same position is revealed when Boyack

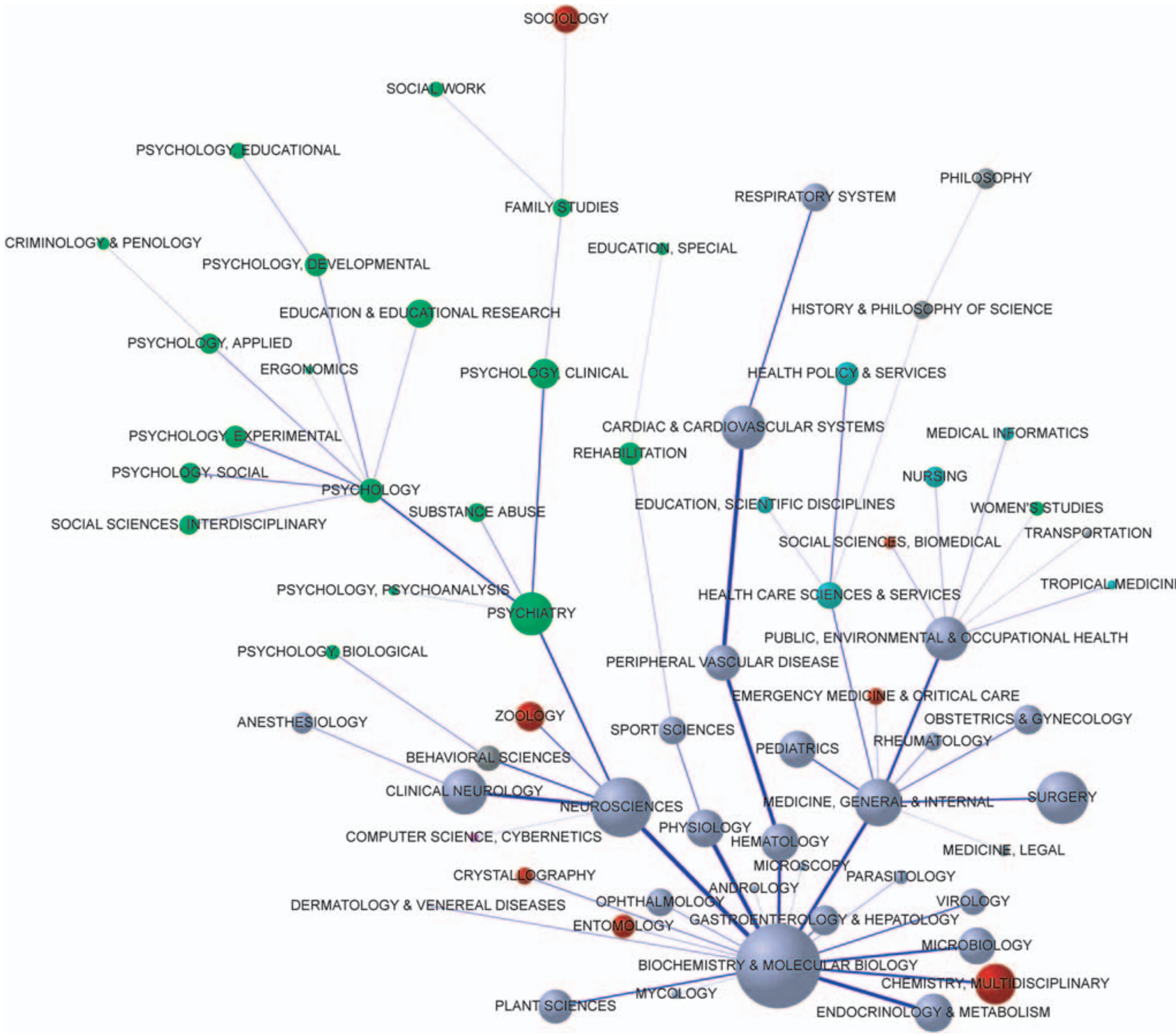

Figure 3: Enlargement of the center of the US Domain 2002.

et al[55] show the backbone of the World science for the year 2000.

The way that the different thematic areas are related over category paths is logical, though it also sheds some added light on the scientific structure of a domain. It could even be used to establish differences according to the domain involved. For instance, if we look closely at the way the areas of Biomedicine and Psychology are connected (zooming into Figure 2), we see that their path of connection goes from *Biochemistry & Molecular Biology* ⟷ *Neurosciences*⟷ *Psychiatry* ⟷ *Psychology* (Figure 3).

This is not the case in China, for the same time span (Figure 6 and its central zoomed area: Figure 4) where the path of connection runs: *Biochemistry & Molecular Biology* ⟷ *Neurosciences*⟷ *Clinical Neurology* ⟷ *Psychiatry* ⟷ *Psychology*, making us suppose that in China, research into Psychology is more focused on clinical and pathological studies. Hence, its intermediary link with Clinical Neurology; whereas in the United States, the branch is more theoretical, dedicated to the study of the individual psyche. The categories seen in red reveal the points of confluence among different thematic areas as a result of their multiple assignment.

Brief description of the evolution of a domain

Figures 5 and 6 show the scientograms of the structure of the Chinese scientific domain for the years 1990 and 2002, respectively. At a glance, there is little similarity with the US model. Yet a certain structural resemblance seems to grow over time. Let us look more closely at each example.

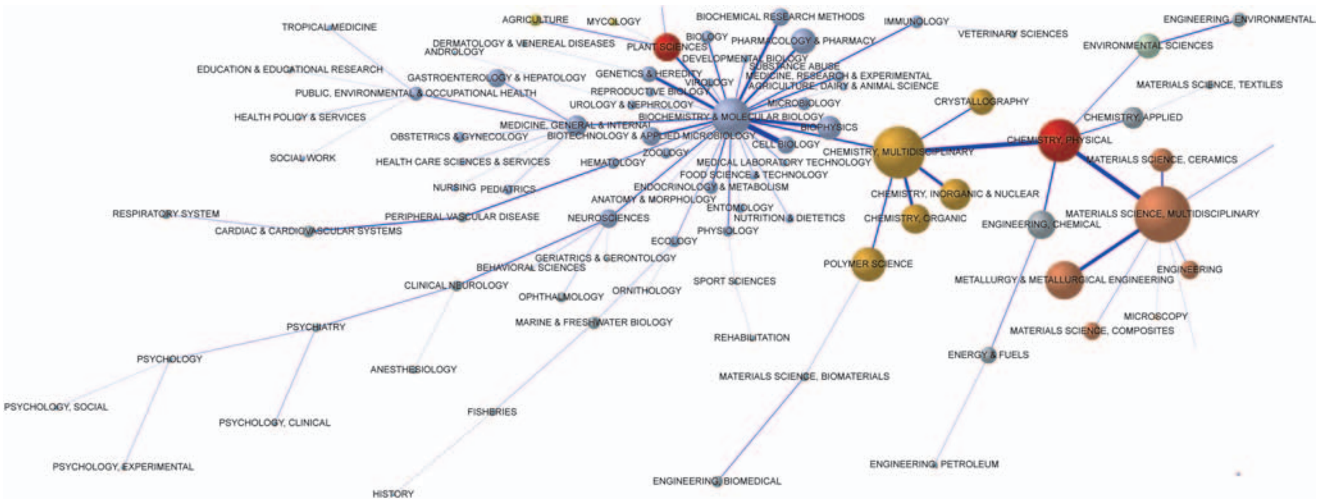

Figure 4: Enlargement of the center of the Chinese Domain 2002.

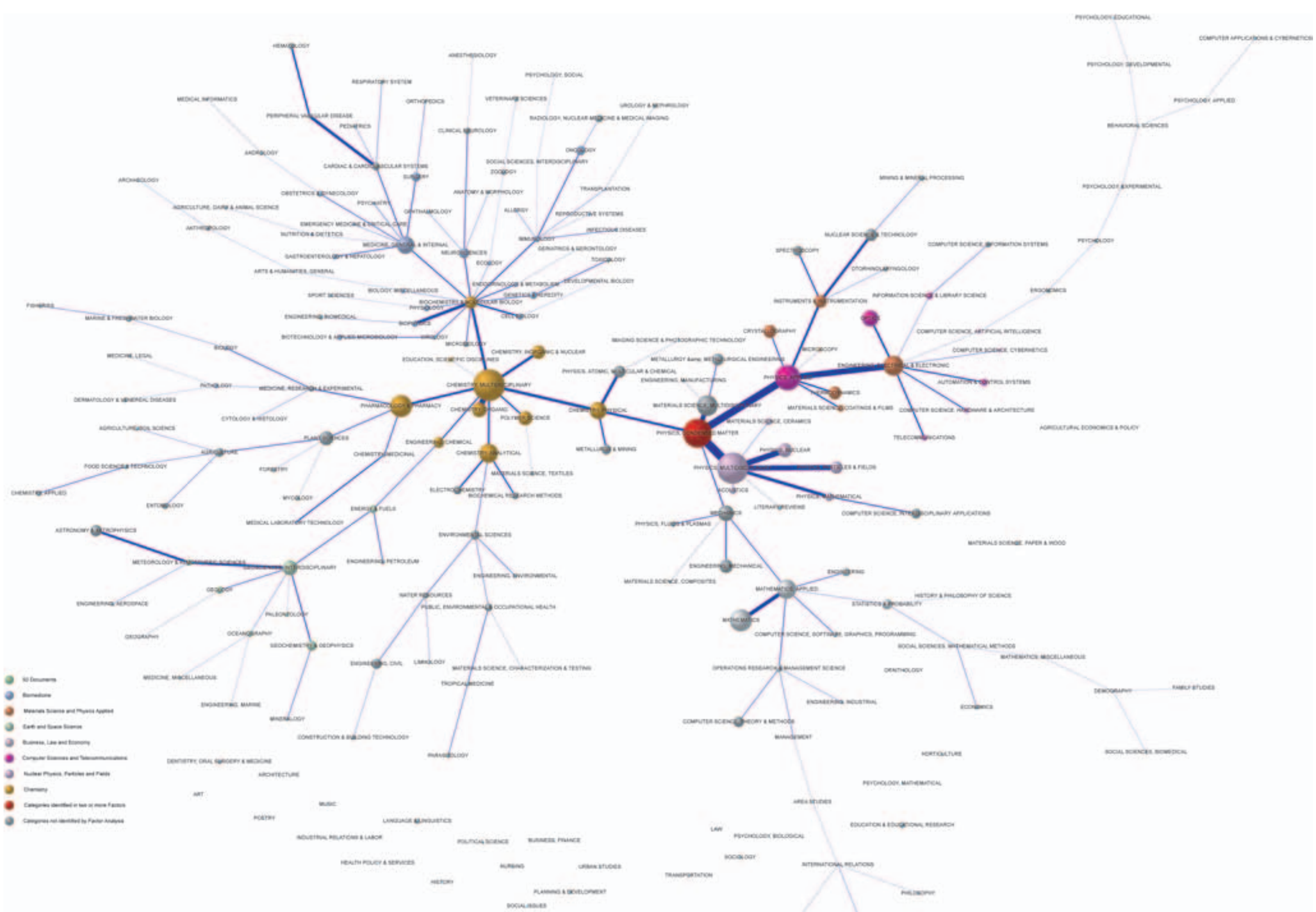

Figure 5: Scientogram of China 1990.

The scientogram of 1990 identifies seven thematic areas by means of FA. Its scheme of vertebration appears quite distinct from the American model, which we called exemplary of developed scientific domains. The center of research is clearly conformed by the area of Chemistry, underlining its importance for this time and place. The research into Biomedicine is just emerging. It would seem to be an offshoot from Chemistry rather than an

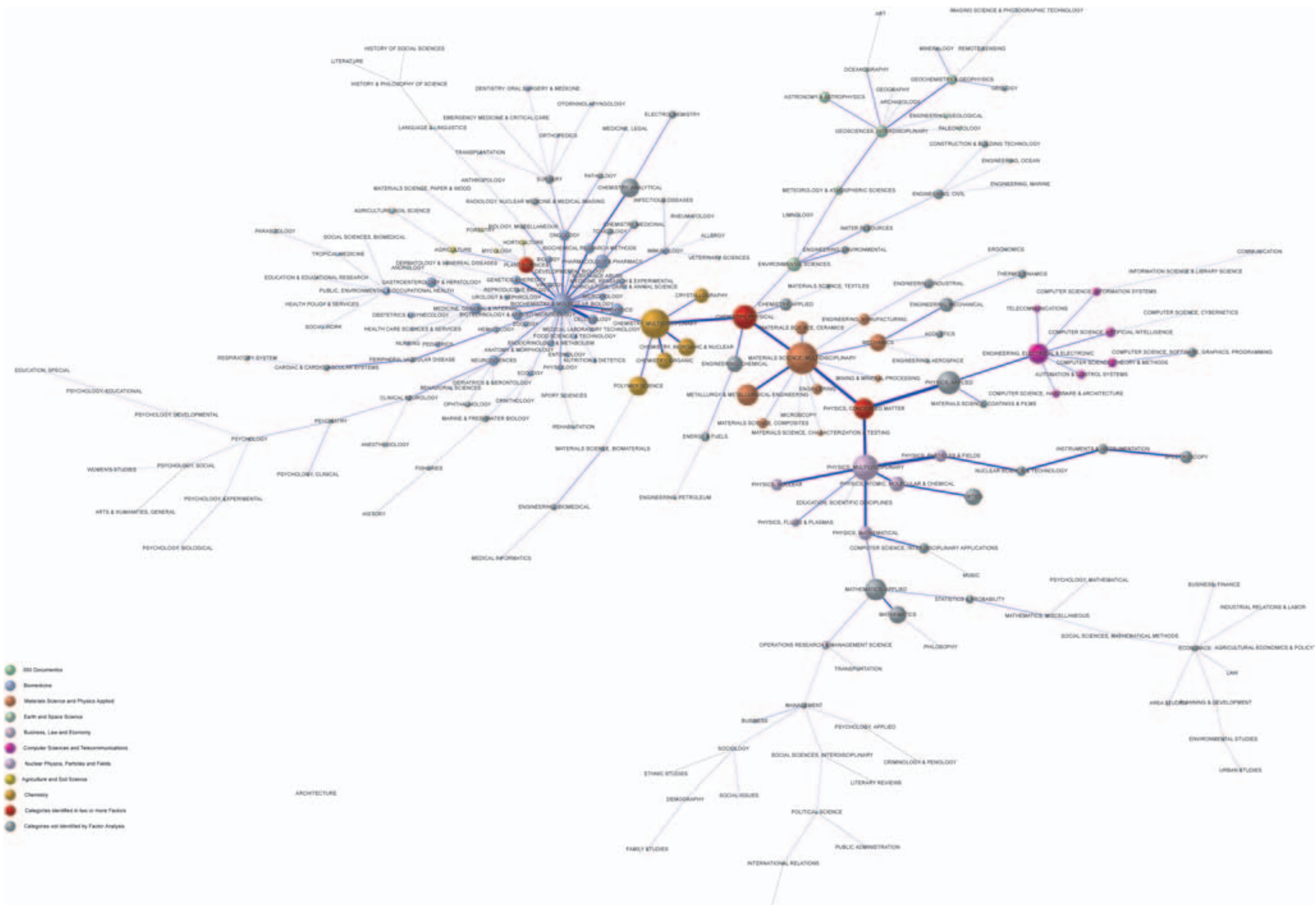

Figure 6: Scientogram of China 2002.

autonomous area of knowledge. This is evidenced by the category *Biochemistry and Molecular Biology*, the nucleus of Biomedical research in developed domains, which here belongs to the area of Chemistry (not Biomedicine). The research output in Science and Technology is beginning to stand out. There are few categories involved, and its structure is weak and somewhat disconnected. The Materials Sciences appear more closely linked to the area of Computer Science and Telecommunications than to Chemistry or Physics. We find only one multidisciplinary category – *Physics Condensed Matters* – which is responsible for interaction among the areas of Materials Sciences, Computer Science and Telecommunications, and Nuclear Physics and Particle Physics.

In the scientogram of 2002 (Figure 6) a total of eight thematic areas can be identified, one more than in the year 1990 – Soil and Earth Sciences. Its backbone is also quite different from the US model. However, the area Biomedicine has shifted toward a more central position, and there is a strong increase in the number of categories that integrate it. Again unlike 1990, the nucleus of research has ceased to be exclusively in the realm of Chemistry and is now shared by this category and the Materials Sciences, making the category denoted *Chemistry Physical* the point of interconnection of the two. There are two multidisciplinary categories in this particular year: *Physics Condensed Matter*, which continues to be the link between the Materials Sciences and Nuclear and Particle Physics; and *Plant Sciences*, which is here the point of confluence of Biomedicine with Agriculture and Soil and Earth Sciences.

Generally speaking, we can say that the scientific domain of China is evolving toward a model that is structurally similar to that of the United States. That is, it has joined the ranks of the developed countries. Its emphasis on research in the areas of Science and Technology as the nucleus of the scientogram of 2002, and the greater presence of the categories of the biomedical area, point to this evolution. The key difference with respect to the US image is the lack of any structural grouping in the area of Social Sciences and Arts and Humanities in Chinese research.

Discussion

Scientograms constitute the intellectual and consensual opinion of hundreds of thousands of researchers

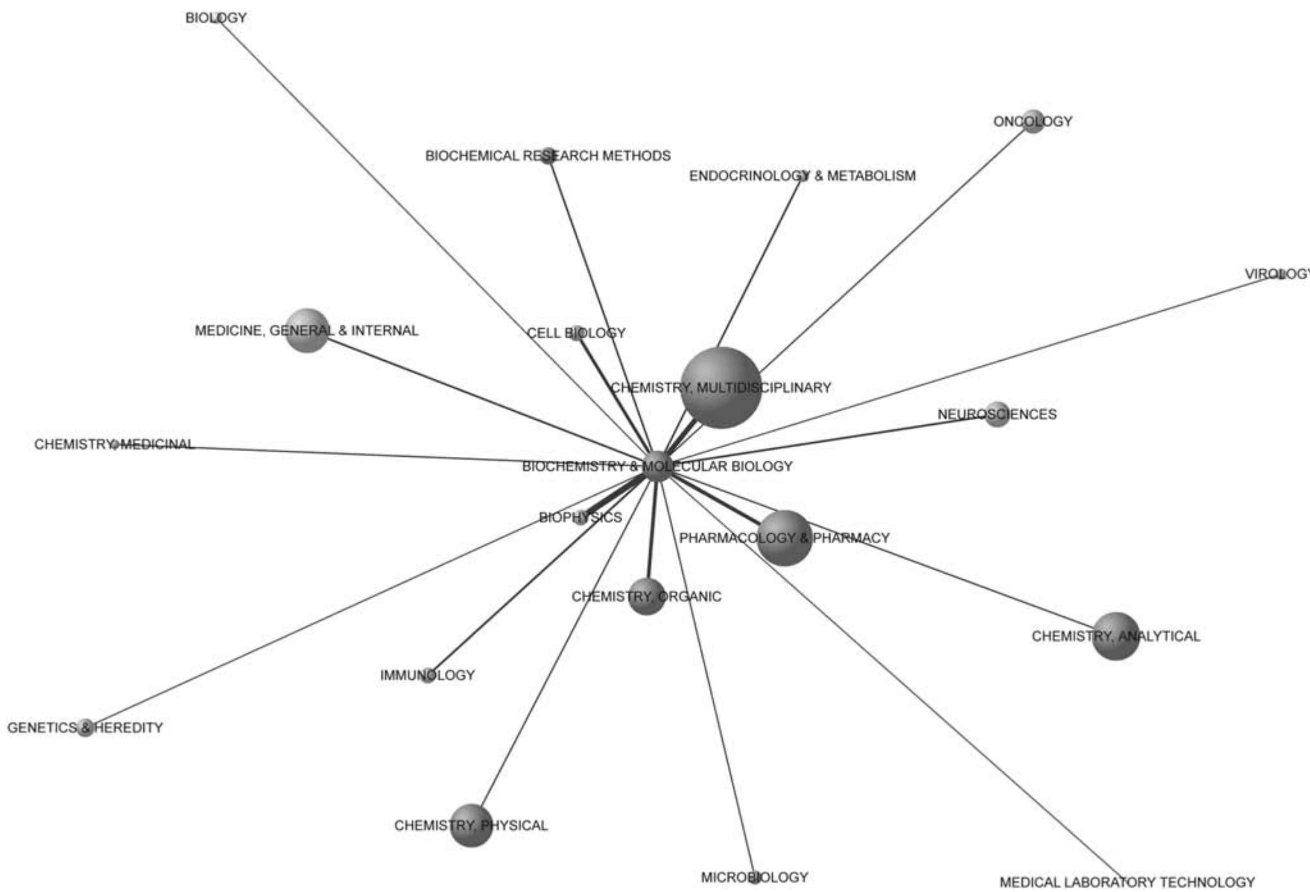

Figure 7: Heliocentric map of Biochemistry and Molecular Biology for China 1990.

worldwide, yet simplified to an extreme degree. Heliocentric maps[54] are a sort of magnifying glass, at a lower level, to show the relationships of one category with the rest and where some relations have been eliminated by PFNET.

The representation of the salient structure of a scientific domain involves a process of clearing around structural elements. If we cannot see the forest for the trees, the trees will have to be pruned. The pruning needed to make manifest the most salient structure of a scientific domain ($r = \infty$ and $q = n - 1$) is appropriate for the categories situated in the central area of the depiction, where greater cocitation between categories is found. This gives rise to bunches or clusters made up of links and categories which, in themselves, supply sufficient information for the analysis and evaluation of a thematic area or even of a given category. However, as we move away from the center of the depiction toward the periphery, where categories cocitation is more limited, the pruning tends to be very aggressive, leaving paths of connection between categories that hardly have bunches. This makes the analysis of the domain less informative and more difficult to interpret. To avoid such a drawback, and in order to delve deeper into the structure of the domain, we resort to heliocentric mapping that offers a second level of representation that is more detailed, and where the selected category appears in the center, surrounded by its neighbors or 'satellites'. For the sake of clarity in the visualization, up to a maximum of 20 nodes are shown. The closer they lie to the central category, or the thicker their links, the greater the relationship of informational interchange between or among categories. To better reflect the thematic areas at a glance, each category is shown with the same color (or shade thereof) representing the factor or thematic area to which it pertains.

Just as we saw with the scientograms of vast scientific domains, these heliocentric maps can be used for a richer and more detailed analysis and comparison of scientific domains, or even to detect patterns of behavior and trends in interchange from category to category over time.

Figures 7–9 show three heliocentric maps for the category *Biochemistry & Molecular Biology*. The first two maps pertain to the Chinese domain, and the third shows the United States.

The heliocentric map of 1990 shows a much more detailed form in the interchange of information between *Biochemistry & Molecular Biology* and its most akin categories. We easily detect a nucleus made up of this category

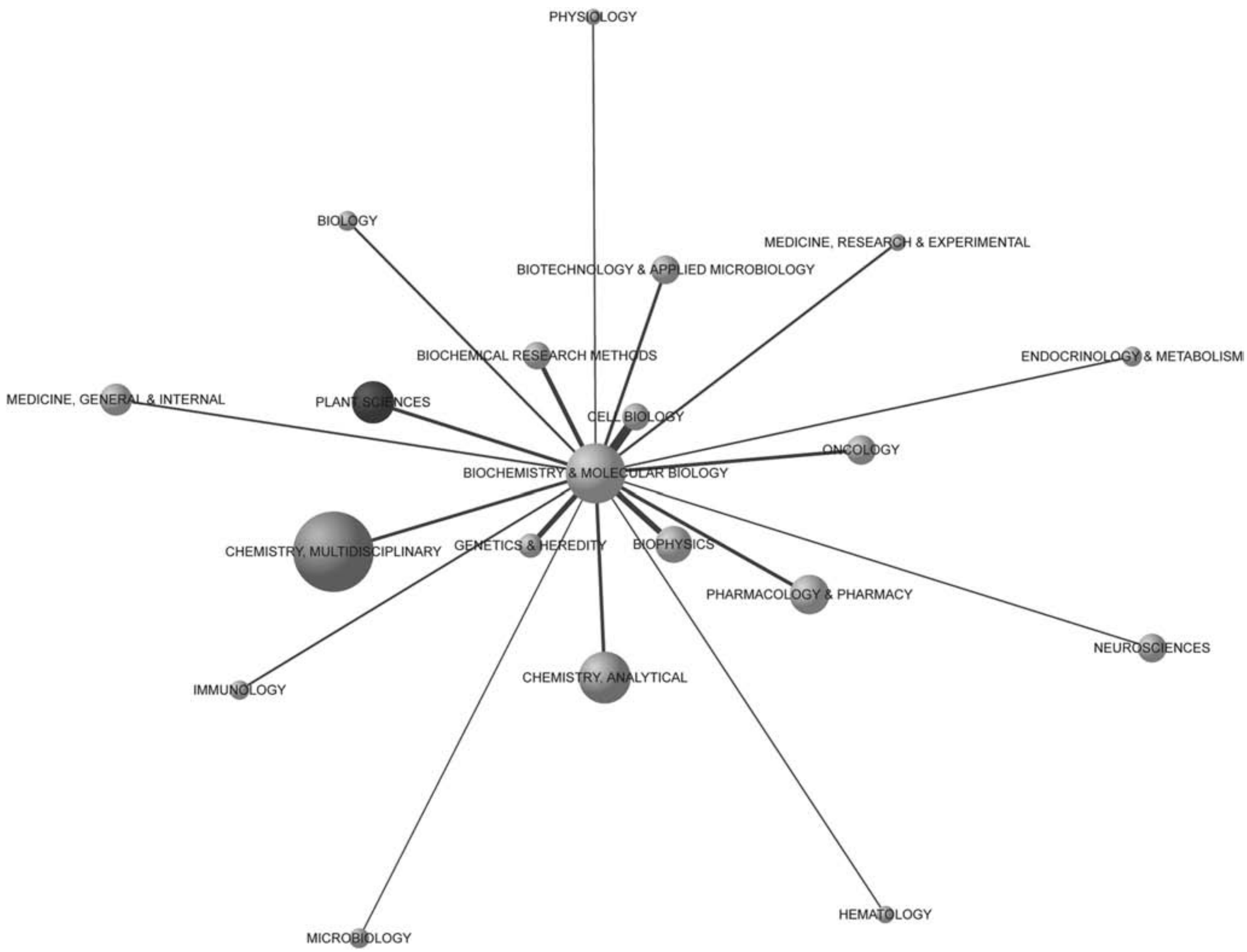

Figure 8: Heliocentric map of Biochemistry and Molecular Biology in China 2002.

and others from the areas Chemistry and Biomedicine. The gray tone code tells us that the central category and the ones with which it shares the most information all belong to the area of Chemistry. In turn, this indicates that here *Biochemistry & Molecular Biology* in particular, and the area of Biomedicine in general, are more focused on chemical and pharmacological research than on strictly medical or clinical studies. Yet the later scientogram tells a different story.

In the heliocentric map of 2002, we observe that the nucleus is relocated amid a conglomerate of categories of the area of Biomedicine. Only two are factorized in other areas: *Plant Sciences*, belonging to Biomedicine and Agriculture and Soil Science (the dark gray indicates the convergence of two areas), revealing the strong and lasting influence of traditional medical practices on modern medicine in China; and *Chemistry Multidisciplinary*, which serves as a bridge between Biomedicine and Materials Science and Physics Applied.

If we now look at the heliocentric map of the US for 2002, we see that practically all the categories, whether in the nucleus or orbiting outside it, belong to the area of Biomedicine.

Bearing in mind the two scientograms of China, we witness the shift of Biomedicine toward the central zone of the scientogram for 2002, a situation very similar to that of the scientogram for US 2002. We must remember that scientograms based on category cocitation and PFNET pruning tend to place the most interdisciplinary categories and thematic areas in the center of the depiction; and we should take note of the evolution of the nuclear category of the area of Biomedicine (*Biochemistry & Molecular Biology*) in the two respective heliocentric maps: there is a considerable increase in output in 2002 (a noticeable difference in size), a change of thematic assignment (the departure of Chemistry and adoption of Biomedicine), and the most adjacent categories no longer belong to Chemistry but rather to Biomedicine. These developments suggest that China is roughly taking after the US domain. Indeed, the heliocentric map of China 2002 is essentially the same as that of US 2002, except for the distance of its categories with respect to the central

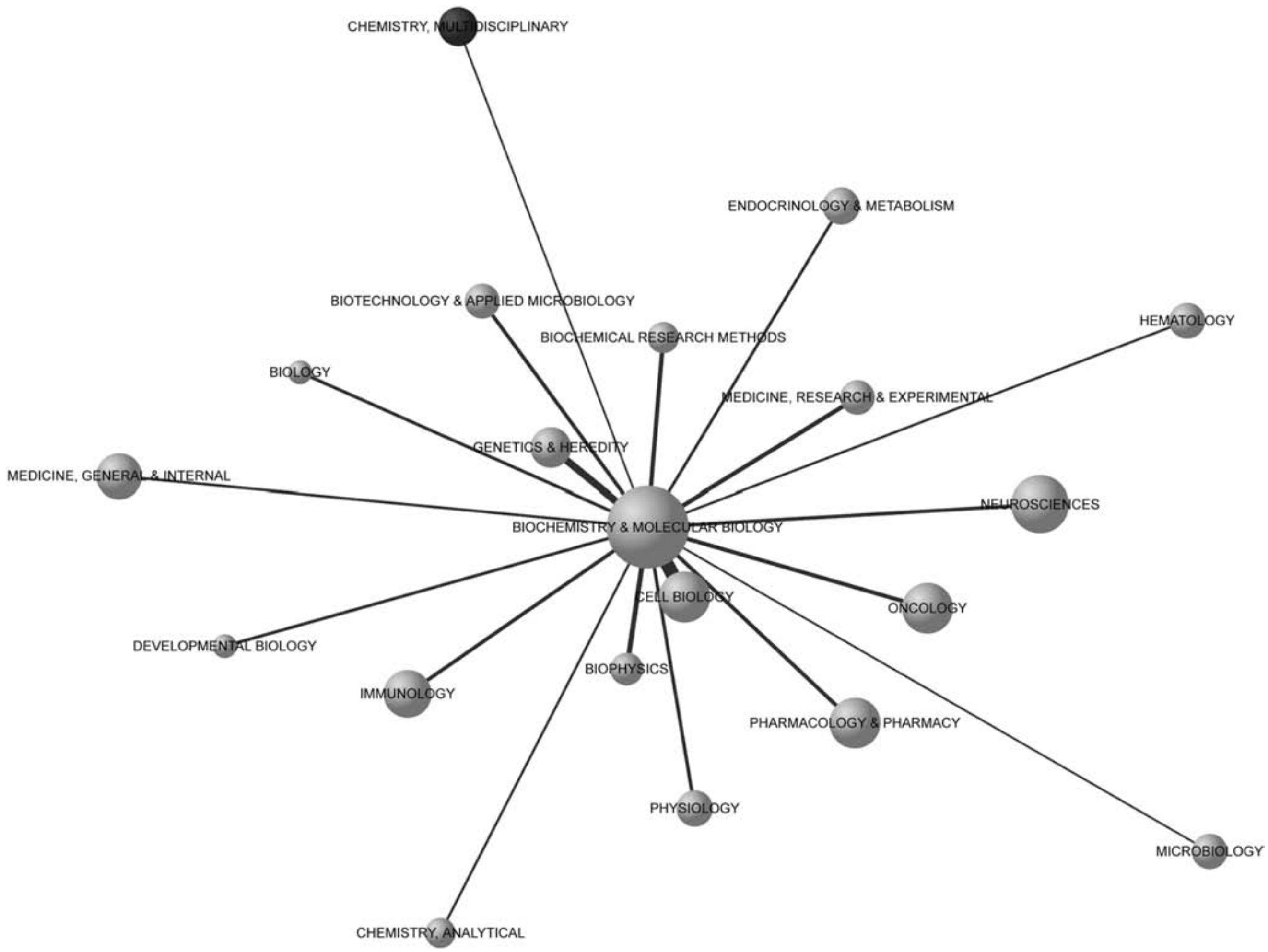

Figure 9: Heliocentric map of Biochemistry and Molecular Biology in US 2002.

one, and the divergence of a single category in each map: *Plant Sciences* in China and *Developmental Biology* in the United States.

Conclusions

With the foundation of a previously developed methodology, we have shown how the scientograms of major scientific domains are very useful tools for the representation, analysis, comparison and evolutionary study of these domains. They can even be used as models to predict the behavior of other domains. The limitations that they entail, as a consequence of pruning the weakest relationships, are compensated by the heliocentric maps, which serve to enhance domain analysis.

The remaining weak points of scientograms and heliocentric maps are those inherent to domain analysis. That is, a certain awareness of the philosophy of science is a pre-requisite, as is some familiarity with the economic, cultural and socio-political aspects of a domain in order to make a proper analysis and adequate interpretation of the underlying scientific structure. Yet a lack of such knowledge can be overcome to a great extent by the representations themselves. As is the case with scientograms, heliocentric maps (such as *Authorlink*[48] and *Citespace*[74] can be used as interfaces to access documents, which are ultimately the elements responsible for establishing the relationships of cocitation among categories, and therefore, the scientific structure of the domain under study. Furthermore, as can be seen in the Atlas of Science (http://www.atlasofscience.net), they are also the point of access for the generation of new maps of journals or authors, which allow the identification of research fronts within each category.

References

1 Moya-Anegón, F.D., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Herrero-Solana, V., Corera-Álvarez, E. and Munoz-Fernandez, F.J. (2004) A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics* 61(1): 129–145.

2 Vargas-Quesada, B. (2005) Visualización y análisis de grandes dominios científicos mediante redes pathfinder. PhD Thesis, Universidad de Granada.
3 Schvaneveldt, R.W. (1990) *Pathfinder Associative Networks*. Norwood, NJ: Ablex.
4 Vargas-Quesada, B. and Moya-Anegón, F.D. (2007) *Visualizing the Structure of Science*. Berlin: Springer.
5 van Raan, A.F.J. and Tijssen, R.J.W. (1993) The Neuronal net of neuronal network research. An exercise in bibliometric mapping. *Scientometrics* 26: 169–192.
6 Garfield, E. (1963) Citation indexes in sociological and historical research. *American Documentation* 14(4): 289–291.
7 Garfield, E., Sher, I.H. and Torpie, R.J. (1964) *The Use of Citation Data in Writing the History of Science*. Philadelphia: Institute for Scientific Information.
8 Price, D.D.S. (1965) Networks of scientific papers. *Science* 149: 510–515.
9 Small, H. and Griffith, B.C. (1974) The structure of scientific literature, I: Identifying and graphing specialities. *Science Studies* 4: 17–40.
10 Griffith, B.C., Small, H., Stonehill, J.A. and Dey, S. (1974) The structure of scientific literature, II: Toward a macro and microstructure for science. *Science Studies* 4: 339–365.
11 Aaronson, S. ((March–April) 1975) The footness of science. *Mosaic* 6: 22–27.
12 Garfield, E. (1981) Introducing the ISI Atlas of Science: Biochemistry and molecular biology 1978–80. *Current Contents* (42): 5–13.
13 Garfield, E. (1984) Introducing the ISI Atlas of Science: Biotechnology and molecular genetics, 1981/82 and bibliographic update for 1983/84. *Current Comments* 7(41): 3–15.
14 Garfield, E. (1988) The encyclopedic ISI-Atlas of Science launches 3 new sections: Biochemistry, immunology, and animal and plant sciences. *Current Comments* 11(7): 3–8.
15 Seiden, L.S. and Swanson, D.R. (1989) ISI Atlas of Science: Pharmacology 1987, Vol 1 - Inst-SCI- Informat. *Library Quarterly* 59(1): 72–73.
16 Small, H. (2000) Charting pathways through science: Exploring Garfield's vision of a unified index to science. In: B. Cronin and H.B. Atkins (eds.), *The Web of Knowledge: A Festschrift in Honour of Eugene Garfield*. Medford, NJ: *Information Today*.
17 Small, H. (1993) Macrolevel changes in the structure of cocitation clusters: 1983–1989. *Scientometrics* 26(1): 5–20.
18 Small, H. (1981) The relationship of information science to the social sciences: A co-citation analysis. *Information Processing & Management* 17(1): 39–50.
19 Small, H. (1994) A SCI-MAP case-study: Building a map of aids research. *Scientometrics* 30(1): 229–241.
20 Small, H. (1999) Visualizing science by citation mapping. *Journal of the American Society for Information Science (JASIS)* 50(9): 799–813.
21 Small, H. and Garfield, E. (1985) The geography of science: Disciplinary and national mappings. *Journal of Information Science* 11: 147–159.
22 Small, H. and Sweeney, E. (1985) Clustering the science citation index using co-citations. 1. A comparison of methods. *Scientometrics* 7(3–6): 391–409.
23 Small, H., Sweeney, E. and Greenlee, E. (1985) Clustering the science citation index using co.citations. 2. Mapping science. *Scientometrics* 8: 321–340.
24 Börner, K., Chen, C. and Boyack, K.W. (2003) Visualizing knowledge domains. *Annual Review of Information Science & Technology* 37: 179–255.
25 *Centre for Science and Technology Studies (CWTS)*, http://www.cwts.nl, accessed 2 February 2009.
26 Braam, R.R., Moed, H.F. and van Raan, A.F.J. (1991) Mapping of Science by combined co-citation and word analysis. I: Structural aspects. *Journal of the American Society for Information Science (JASIS)* 42(4): 233–251.
27 Braam, R.R., Moed, H.F. and van Raan, A.F.J. (1991) Mapping of Science by combined co-citation and word analysis. II: Dynamic aspects. *Journal of the American Society for Information Science (JASIS)* 42(4): 252–266.
28 Brandenburg, F.J., Himsolt, M. and Rohrer, C. (1995) An experimental comparison of force-directed and randomized graph drawing algorithms. *Lecture Notes in Computer Science* 1027: 76–87.
29 Lin, X., Soergel, D. and Marchionini, G. (1991) A self-organizing semantic map for information retrieval. In: Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval; Chicago, Illinois: SIGIR, pp. 262–269.
30 White, H.D. and McCain, K.W. (1997) Visualization of literatures. *Annual Review of Information Systems and Technology (ARIST)* 32: 99–168.
31 White, H.D. and McCain, K.W. (1998) Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science (JASIS)* 49: 327–355.
32 Garfield, E. (1998) Mapping the world of science (at the 150 Anniversary Meeting of the AAAS, Philadelphia, PA) [On-line], http://www.garfield.library.upenn.edu/papers/mapsciworld.html, accessed 7 May 2003.
33 White H.D., Lin X., and McCain K.W. (1998). Two modes of automated domain analysis: Multidimensional scaling v.s Kohonen feature mapping of information science authors. In: Mustafa el Haidi, J. Maniez and S.A. Polit (eds.), Proceedings of the Fifth International ISKO Conference; 57–61, Würzberg: Ergon Verlag.
34 Chen, C. (1998) Generalised similarity analysis and pathfinder network scaling. *Interacting with Computers* 10: 107–128.
35 Chen, C. (1998) Bridging the gap: The use of pathfinder networks in visual navigation. *Journal of Visual Languages and Computing* 9: 267–286.
36 Ding, Y., Chowdhury, G.G. and Foo, S. (1999) Mapping the intellectual structure of information retrieval studies: An author co-citation analysis, 1987–1997. *Journal of Information Science* 25: 67–78.
37 Merton, R.K. (2000) On the Garfield input to the sociology of science: A retrospective collage. In: B. Cronin and H.B. Atkins (eds.), *The Web of Knowledge: A Festschrift in Honour of Eugene Garfield*. Medford, NJ: *Information Today*.
38 White, H.D. (2000) Toward ego-centered citation analysis. In: B. Cronin and H.B. Atkins (eds.), *The Web of Knowledge a Festschrift in Honour of Eugene Garfield*. Medford, NJ: *Information Today*, pp. 475–498.
39 Noyons, E.C.M., Moed, H.F. and Luwel, M. (1999) Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science (JASIS)* 50(2): 115–131.
40 Buter, R.K. and Noyons, E.C.M. (2001) Improving the functionality of interactive bibliometric science maps. *Scientometrics* 51(1): 55–67.
41 Noyons, E.C.M. (2001) Bibliometric mapping of science in a science policy context. *Scientometrics* 50(1): 83–98.
42 Chen, C. and Paul, R.J. (2001) Visualizing a knowledge domain's intellectual structure. *Computer* 34: 65–71.
43 Chen, C., Paul, R.J. and O'Keefe, B. (2001) Fitting the jigsaw of citation: Information visualization in domain analysis. *Journal of the American Society for Information Science & Technology (JASIST)* 52(4): 315–330.
44 Ding, Y., Chowdhury, G.G. and Foo, S. (2001) Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management* 37: 801–817.
45 Guerrero Bote, V.P., Moya Anegón, F.D. and Herrero Solana, V. (2002) Document organization using Kohonen's algorithm. *Information Processing & Management* 38: 79–89.
46 White, H.D, Buzydlowski, J. and Lin, X. (2000) Co-cited author maps as interfaces to digital libraries: Designing pathfinder networks in the humanities. In: Proceedings of the International Conference on Information Visualization (IV2000): 25–30, London: Computer Society Press.
47 Buzydlowski, J., White, H.D. and Lin, X. (2002) Term co-occurrence analysis as an interface for digital libraries. *Lecture Notes in Computer Science Series* 2539: 133–144.

48 Lin, X., White, H.D. and Buzydlowski, J. (2003) Real-time author co-citation mapping for online searching. *Information Processing & Management* 39: 689–706.
49 Small, H. (2003) Paradigms citations and maps of science: A personal history. *Journal of the American Society for Information Science & Technology (JASIST)* 54: 394–399.
50 Chen, C. and Kuljis, J. (2003) The rising landscape: A visual exploration of superstring revolutions in physics. *Journal of the American Society for Information Science & Technology (JASIST)* 54: 435–446.
51 Chen, C. and Morris, S. (2003) Visualizing evolving networks: Minimum Spanning trees versus pathfinder networks. In: IEEE Symposium on Information Visualization. Seattle, Washington: IEEE Computer Society Press, pp. 67–74.
52 Morris, S.A., Yen, G., Wu, Z. and Asnake, B. (2003) Time line visualization of research fronts. *Journal of the American Society for Information Science & Technology (JASIST)* 54: 413–422.
53 Boyack, K.W. and Börner, K. (2003) Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science & Technology (JASIST)* 54: 447–461.
54 Moya-Anegón, F.D., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Herrero-Solana, V., Corera-Álvarez, E. and Munoz-Fernandez, F.J. (2005) Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Management* 41: 1520–1533.
55 Boyack, K.W., Klavans, R. and Börner, K. (2005) Mapping the backbone of science. *Scientometrics* 64: 351–374.
56 Klavans, R. and Boyack, K.W. (2006) Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science & Technology (JASIST)* 57: 251–263.
57 Klavans, R. and Boyack, K.W. (2006) Quantitative evaluation of large maps of science. *Scientometrics* 68: 475–499.
58 Samoylenko, I., Chao, T.-C., Liu, W.-C. and Chen, C.-M. (2006) Visualizing the scientific world and its evolution. *Journal of the American Society for Information Science & Technology (JASIST)* 57: 1461–1469.
59 Leydesdorff, L. (2007) Mapping interdisciplinarity at the interfaces between the science citation index and the social science citation index. *Scientometrics* 71: 391–405.
60 Thomson Reuters. ISI Web of Knowledge, http://www.isiwebofknowledge.com, accessed 9 September 2008.
61 Glänzel, W., Schubert, A. and Czerwon, H.-J. (1999) An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics* 44: 427–439.
62 Glänzel, W., Schubert, A., Czerwon, H.-J. and Shoepflin, U. (1999) An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis. *Scientometrics* 46: 431–441.
63 White, H.D. (2003) Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientist. *Journal of the American Society for Information Science & Technology (JASIST)* 54(5): 423–434.
64 Leydesdorff, L. and Vaughan, L. (2006) Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science & Technology (JASIST)* 57: 1616–1628.
65 HjØrland, B. and Albrechtsen, H. (1995) Toward a new horizon in information science: Domain analysis. *Journal of the American Society for Information Science (JASIS)* 46: 400–425.
66 Fowler, R.H. and Dearhold, D.W. (1990) Information retrieval using pathfinder networks. In: R.W. Schvaneveldt (ed.), *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, MA: Ablex NJ, pp. 165–178.
67 Chen, C. (1999) Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management* 35: 401–420.
68 Eades, P. (1984) A heuristic for graph drawing. *Congressus Numerantium* 42: 149–160.
69 Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. *Information Processing Letters* 31: 7–15.
70 Kruskal, J.B. and Wish, M. (1978) *Multidimensional Scaling*. London: Sage.
71 Cohen, J. (1997) Drawing graphs to convey proximity: An incremental arrangement method. *ACM Transactions on Computer-Human Interaction* 4: 197–229.
72 Krempel, L. (1999) Visualizing Networks with Spring Embedders: Two-mode and Valued Graphs. American Statistical Association, *Proceedings of the Section of Statistical Graphics*, Alexandria, VA: pp. 36–45.
73 ANEP (2004) *Agencia Nacional de Evaluación y Prospectiva*, http://www.mcyt.es/sepct/ANEP/anep.htm, accessed 2 February 2009.
74 Chen, C. (2004) *CiteSpace: Visualizing Patterns and Trends in Scientific Literature*, http://cluster.cis.drexel.edu/~cchen/citespace/, accessed 2 February 2009.