A new technique for building maps of large scientific domains based on the cocitation of classes and categories

FÉLIX MOYA-ANEGÓN, BENJAMÍN VARGAS-QUESADA, VICTOR HERRERO-SOLANA,
ZAIDA CHINCHILLA-RODRÍGUEZ, ELENA CORERA-ÁLVAREZ,
FRANCISCO J. MUNOZ-FERNÁNDEZ

University of Granada, Library and Information Science Faculty, Granada (Spain)

Our objective is the generation of schematic visualizations as interfaces for scientific domain analysis. We propose a new technique that uses thematic classification (classes and categories) as entities of cocitation and units of measure, and demonstrate the viability of this methodology through the representation and analysis of a domain of great dimensions. The main features of the maps obtained are discussed, and proposals are made for future improvements and applications.

Introduction

Scientific information is spread out over disciplines which, to the outside observer, may seem to have little in common. For this reason, when traditional methods are used to study a domain pertaining to one specific field of knowledge, one is sometimes left with a sensation of not grasping the domain as a whole. It is like trying to complete a puzzle and not knowing where to put the piece held in the hand, not seeing which puzzle pieces it fits in with.

The representation of scientific information in ways easier for the human mind to embrace is nothing new. To make visible to the mind that which is not visible to the eye, or to create a mental image of something that is not obvious (e.g. an abstraction)[1], are two definitions of the word "visualization" that point to the intrinsic need to represent information in a non-traditional manner. To paraphrase Costa[2], visualizing is neither the implicit result of the act of seeing, nor a spontaneous product of the individual receiving visible input. To visualize is a task of the communicative process, through which abstract data and complex phenomena of reality are transformed into visible messages. This enables individuals to apprehend with their own eyes certain data and phenomena that cannot be directly retrieved from a hidden body of knowledge.

The realm of graphic languages for visualizing these "invisible effects" constitutes a new science of visual communication – schematics – which Costa has defined as the "third language" after the image and the sign.

The present study proposes a new technique for schematic visualization applied to the analysis of large scientific domains. The scientific domain is understood in the terms put forth by Hjørland and Albrechten[3], as the reflection of interactions between authors, and their role in science, through citation. The methodology applied allows the analysis of data and can serve as an intermediate step in information retrieval.

We shall begin with an overview of different contributions to, or attempts at, the representation of large scientific domains, to see how they approached the visualization of small domains while available tools and techniques matured. The elements and units of measure that could be used for the graphic representation of vast domains are then discussed, as is the methodology developed for the generation of maps that make browsing easier. Finally, our results and conclusions are presented, and the particular features of several examples of maps are mentioned.

A brief bibliographic review of contributions to the representation of large scientific domains

The first author to articulate this need was Doyle, in 1961.[4] He underlined the importance of computers in producing maps similar to those that the brain generates, and showed how they can be projected in multidimensional spaces. He also offered his opinions as to how to construct such maps. To get "the big picture" of a large scientific domain has allured many researchers since. Garfield,[5] in an article published in American Documentation, showed vivid interest in the construction of historical maps by means of citation. This interest, taken up by Sher in 1964, led to the creation of a historical map showing the development of DNA from Mendel to Nirenberg. That same year Garfield, Sher and Torpie[6] generated manually distinct historic-topological science maps, on the basis of citation of DNA research, using bibliographic coupling as a variable. DeSolla Price[7] showed that the patterns of citations by the authors of scientific articles would define research fronts, and could also be used to sketch a topology reflecting the intellectual structure of a scientific domain. But the giant leap forward in the possibilities for building domain maps or graphs came, in our opinion, from Small[8] and Marshakova,[9] each of whom proposed the use of document cocitation as a variable of study in the analysis of citations of scientific production. Science maps showing all the special fields of the natural sciences, elaborated by Small and Griffith[10] or Griffith, Small, Stonehill and Dey,[11] using the *Science Citation Index* (SCI) as their source of information and cocitation as a variable of relation, stand as a landmark in the development of the representation of scientific domains.

The most important point about the methodology used by the above authors is that the groups of documents with common intellectual interests were identified. This was proof that science is a network of interconnected special fields that can be contemplated through the quantitative analysis of written production. In 1975, along these same lines, Aaronson[12] X-rayed the biomedical publications from the years 1972 and 1973 to observe their evolution over this time and highlight, on the 1973 map, what the author called a *supercluster*, showing the convergence of special fields. The maps by Aaronson are another turning point in the history of domain representation; not only do they show its evolution and how the different disciplines mutually interact, but they also supply information: the documents conform a map of *clusters*. Each cluster is characterized by a name, a precise number of documents comprising it, and their degree of interrelation. This is represented by connecting lines that indicate the value of cocitation. The most important clusters stand out very clearly, as they have a greater number of relationships with the rest.

Shortly after the appearance of the Aaronson maps, Garfield[13] reported that the Institute for Science Information (ISI) was working on the elaboration of an *Atlas of Science*. This project took six years to materialize. It was finally in 1981[14] when the first two volumes of the Atlas appeared, on biochemistry and molecular biology. The techniques used for the generation of the Atlas maps are quite similar to those used by Small and Griffith in 1974 (that is, based on the cocitation of documents of a specific discipline), yet a new spatial positioning of the clusters is achieved by applying Multidimensional Scaling (MDS). Over time, new volumes of the *Atlas of Science* have presented biotechnology and molecular genetics,[15] and biochemistry, immunology, and plant and animal biology,[16] then drew to a close, to date at least, after the volume on Pharmacology.[17] Meanwhile, Small has continued to work on the design of maps of scientific domains, refining the techniques used in his early maps. He can be considered the ISI´s top specialist in the research and development of science maps.[18-25]

After the 90's, new methods of information retrieval and new techniques for the analysis, visualization and spatial positioning of information (well reviewed by Börner, Chen and Boyack[26]), studies based on techniques for visualizing the structure of small scientific domains begin to proliferate. So, for instance, Braam, Moed and van Raan[27,28] propose the combined use of cocitation with co-word analysis for the generation of science maps, emphasizing their structure and dynamic aspects. Lin, Soergel and Marchionini[29] develop a Self-Organizing Map (SOM) that represents the semantic relationships among documents and can be used as a bibliographic interface for the retrieval of online information. Hjørland and Albrechtsen[3] put forth a new model based on the way of thinking or speaking of the society, in turn determined by the social, economic, or work setting, and which they call domain analysis. White and McCain,[30] on the basis of this domain analysis, propose graphic representation or visualization as a model for information retrieval and analysis. To corroborate their theory, in 1988 they

analyzed information science[31] in terms of the authors of the 12 journals that, according to their criteria, were most important between 1972 and 1995. Garfield[32] states that the new techniques of visualization make possible the generation of global science maps which, by zooming over or representing different time periods, allow us to identify the emerging research fronts, revealing the interests of researchers now at work and allowing us to associate author names to each front.

White, Lin and McCain[33] compare the relatively traditional mode of visualizing scientific domains using MDS with the Self-organizing Maps (SOM), to conclude that the latter make it slightly easier to integrate and retrieve bibliographic information. Ding, Chowdhury, Foo and Qiang[34] use bibliometric techniques to break down an area of knowledge into its main elements, and represent the areas and subareas graphically. White[35] presents networks revolving around on a subject, and proposes that maps be made with an author name supplied by a non-expert user. This involves a lesser cognitive load (for the user) the development of interfaces for inexpert users facilitate the retrieval of information from the bibliographic information, and the possibility that these interfaces can be generated in a dynamic way. Noyons, Moed and Luwel,[36] Buter and Noyons,[37] and Noyons[38] point to the great potential of science maps as an interface for information retrieval, and suggest some ways to aid the user in exploring and comprehending what he sees. Chen and Paul[39] describe a method that broadens and transforms traditional coauthor analysis into structural patterns of scientific literature that can be represented in 3D maps. For Chen, Paul and O'Keefe,[40] the essence of knowledge, whether in a geographical, thematic, or intellectual context, is the key to visualization. The proliferation of techniques for the visualization of information allows the core of knowledge to be represented. This is a key role in the process of modelization and representation of the structure or intellectual map of a given domain, be it geographical, thematic or intellectual.

Again, Ding, Chowdhury and Foo[41] make a map of the intellectual structure of the field of information retrieval, this time over a ten-year period (1987–1997), showing models, patterns and trends of the field as well as different measurements of the degrees of association among the most relevant terms of the document produced under the heading "information retrieval." Ingwersen and Larsen[42] use MDS to create a map of the production of 17 European countries in nine areas of the social sciences. Guerrero Bote, Moya Anegón and Herrero Solana[43] classify database documents automatically using a SOM, and describe how it can be used for browsing and retrieving information from the database. White, Buzydlowski and Lin,[44-46] on the basis of the Pathfinder networks (PFNET) used by Chen,[47] and White's CAMEOs,[48] devise a dynamic system of visualization called *Authorlink* (using author cocitation) that allows browsing and information retrieval in real time from a database with records of the *Arts and Humanities Citation Index* (A&HCI). Small[49] theorizes about the design of a web tool able to detect and monitor, in real time, the changes on research fronts resulting from its

interactions. Chen and Kuljis,[50] using citation and cocitation, study the appearance and evolution of new research fronts in the field of physics. Morris, Yen, Wu and Asnake[51] work on visualization, detection and identification of changes on research fronts over time. Boyack and Börner,[52] in an evaluative study, generate maps of scientific publications with government grants to reveal connections between funding and the number of citations received.

What we can clearly derive from this general overview is that domain maps or visualizations are primarily used, thus far, to reveal relationships among documents, to detect the most important authors within a given discipline, or to analyze the structure of an area of knowledge and its evolution. The methodology may involve clustering, MDS, factor analysis, or social networks based on models of graphs, or some combination thereof.

At present, a great deal of research shares the common objective of producing an initial map of a domain (a blueprint of sorts) that is general and informative enough to orient the non-expert user, while also capable of panning over or zooming in on levels of a discipline using multivariate techniques or network analysis.

Our proposal

In our opinion, cocitation is to date the best tool for obtaining relational information on documents with which to schematically represent the image of a domain. Thus, depending on the type of variable analyzed (words, documents, authors, journals), it is possible to offer different snapshots of a domain to reflect the relationships existing among the component elements. The use of one variable or another in cocitation will depend largely on the size of the domain to be represented, as well as on the possibility or interest in obtaining dynamic visualizations – that is, either online or offline maps.[30] But there may also be persistent physical limitations, such as the representation of information in a low-resolution reduced and static space, which has drawn the attention of Tufte[53,54] for over a decade.

Our proposal entails the graphic and schematic representation of vast thematic, geographical or institutional domains based on cocitation. It is obviously necessary to perform some type of clustering, if we hope to corral the intellectual structure of a large domain on a computer screen, making it quickly intelligible for the human eye and mind. We propose the cocitation of classes and categories based on the following scheme:

It is generally accepted by the research community that the frequency with which any two documents are cocited reflects their degree of affinity in the opinion of the citing authors.

In Figure 1, documents A and B are cocited in document X. The intensity with which A and B are mutually related depends on the number of times that they are cocited.
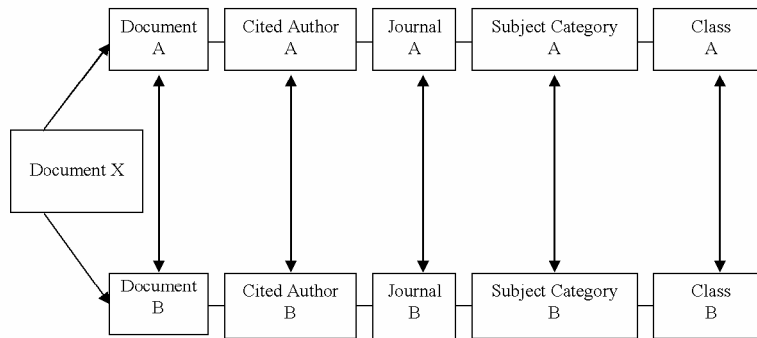

Figure 1. Cocitation scheme

Just as the reseach community accepts the fact that documents A and B are an entity of cocitation valid for the representation of the structure of a domain, making evident their semantic and intellectual relationships, it also accepts the use of authors as the entity of cocitation and as the unit of measure for representing this domain. The same happens in the case of journals.

The ISI, in the *Journal Citation Report* (JCR),[55] places every journal in at least one subject category. For example, the journal *Scientometrics* is situated in the category Computer Science, Interdisciplinary Applications. By analogy, the cocitation of categories assigned by the ISI-JCR could be used as the entity of measure following the order established in Figure 1, that is: documents, authors, journals, and ISI-JCR subjects. To go back to our example, if document A was published in *Scientometrics* and therefore has been assigned to the above category, while document B appears in *Pattern Recognition Letters*, assigned to the ISI-JCR categories of Computer Science, Artificial Intelligence, the interrelationship of all these categories becomes evident. The intensity of the connections will depend on the number of times that documents published in journals of these subject categories are cocited.

There are different classifications that group ISI-JCR subject categories in superior conglomerates – heretofore referred to as classes. Because we attempt to represent the structure of Spain's scientific production, we have adopted the classification of the Agencia Nacional de Evaluación y Prospectiva (ANEP),[56] which is a taxonomy carried out by experts in the evaluation of science for the technical and scientific assessment of

the action plans of the National Plan of Spanish Scientific Research, adapting the ISI-JCR categories as a distinctive classifying element of scientific output in Spain, to twenty-five classes. As is the case in the ISI-JCR, one same category can belong to different subject areas. Accordingly, the classes may perfectly well be used as the entity of cocitation and as a unit of greater measure, valid for representing the intellectual structure of a domain, as we will show later on.

In short, we propose the use of the cocitation of classes and categories as entities of cocitation and units of measurement for the generation of great schematic visualizations that act as graphic interfaces for domain analysis.

Methodology

For strictly academic purposes we downloaded from the *Web of Science*[57] – specifically from: the *Science Citation Index-Expanded* (SCI-E), the *Social Science Citation Index* (SSCI) and the *Arts and Humanities Citation Index* (A&HCI) – the records with at least one Spanish address in the field "address" from the year 2000, and put them into an ad hoc database for consultation. The database held 172,562 author names, who published a total of 26,062 documents (articles, biographical items, book reviews, corrections, editorial materials, letters, meeting abstracts, news items and reviews) in 3,838 different ISI journals. When these were broken down into the 243 categories established by the ISI-JCR for the year 2000, 222 categories were covered. (Spain did not produce any scientific communications pertaining to the remaining 21 categories within the year 2000.) These 222 categories were grouped, based on the ANEP classification, in 25 classes, again taking into account that one single category may belong to different classes.

Because we try to show the relationships existing among diverse disciplines in the natural sciences, social sciences, arts and humanities, we must first solve the problem of uneven level of citation, as suggested by Small and Garfield.[19] For this reason, when carrying out the cocitation queries corresponding to the classes or categories, we normalize this measurement of association by dividing the cocitation by the square root of the product of the frequency of the cites of the cocited documents:

$$Cc = \frac{Ccij}{\sqrt{ci \cdot cj}}, \text{ (Measurement of normalized cocitation}^{58}\text{)}$$

where:
- Cc is cocitation
- c is citation.

Kamada & Kawai's algorithm[59] was used to automatically produce representations on a plane, starting from a circular position of the nodes. It generates social networks with aesthetic criteria such as the maximum use of available space, the minimum number of crossed links, the forced separation of nodes, building balanced maps, etc.

The result is a tree structure with the following characteristics: a map representing the Spanish scientific structure as a whole dividing it in 25 big classes. 25 maps, one for every ANEP class, each of which containing the ISI-JCR categories that ANEP evaluation experts have considered appropriate. And finally, 222 maps of ISI-JCR categories, one for every ISI-JCR category with its nearest neighbours.

Class cocitation map (first level)

To obtain the basic data needed for a graphic representation of the whole domain of Spain, we carry out a cocitation query of classes under the ANEP classification, as described above.

The result is a symmetrical class cocitation matrix of 25×25. Of course, the degree of intellectual connections shown by the cocitation matrix among certain classes is very high, making it difficult in some zones to clearly visualize the structure of the domain. In following the advice of Small,[25] therefore, we believe it better to eliminate some connections, as "the loss of information of the structure implies a gain in simplicity, justifying the sacrifice in some cases." We therefore prune the relations between classes using the Minimum Spanning Tree (MST): when relationships among classes are under a threshold value, they are successively deleted until only one is left, totally disconnected from the rest. Then the threshold value is re-established, leaving no class disconnected. Other available pruning tools include PFNET; but as White[60] points out, this algorithm prunes all the paths except those having the highest degree of meaningful cocitation, leaving a very reduced number of coincidences in the matrix, where $q = n - 1$ and $r = \infty$. Since our aim is to show the highest number of possible links to represent the structure of a domain as well as its semantic and intellectual relationships, we opted to use MST, because it allows us to reduce the minimum number of links in order to facilitate the domain visualization and analysis without loose information. Similar results might be obtained using PFNET with a lower q, such as $q = 2$ for example, but this will be addressed in the future.

We assign each class a different color to distinguish it from the rest, and a size proportional to the totality of Spanish scientific output in the year 2000, as well as an ANEP label. This information, together with the cocitation matrix, is processed by the Kamada & Kawai algorithm to produce a social network where each class is represented by a node connected with other nodes by undirected links. The relationships among them and their intensity is seen in the thickness of the links. In just a few cases in which the tag of a node is partly superimposed on another, we manually modified the

location assigned by the algorithm a bit. The definitive network is exported to an Scalable Vector Graphic (SVG) format,[61,62] which allows us to zoom in on or shift the graphics in any direction on the screen.

Category cocitation map (secondary level)

The process involved is very similar to that used for classes, the only difference residing in details for improving the visualization of the intellectual relationships. As we commented earlier, each one of the classes is made up of a specific number of ISI categories, assigned according to the criteria of the ANEP. For each ANEP class, we consult the cocitation of ISI categories, normalized as explained earlier, to obtain a symmetrical cocitation matrix of $n \times n$ categories, based on the number of each in each class. After pruning by MST we assign a color to each category, which is the color of the class it belongs to. We adjust slightly in each level: the category with the greatest scientific output is the biggest one, and the rest are made proportional to this, reflecting their relative magnitude in the context of total publication. We observe that there is little difference among the categories of a single class, and those with a small yield, such as four works, are perfectly visible.

The Kamada & Kawai algorithm is supplied with the name of the categories that make up each class, their size, color, and the corresponding cocitation matrix, which is what establishes the relationships among categories. In this case, we indicate to the algorithm that these relationships are to be directed. Though it is a symmetric matrix (and therefore when relationships exist among categories, the categories will likewise be symmetrical), taking this small license helps us achieve a very clear and well structured visualization of the domain. The thickness of the links again indicates the intensity of their interrelationships.

The social network that we obtain, modified only in a few cases to avoid the overlap of tags on nodes or categories, is exported to an SVG format.

Map of neighbors (third level)

Here we do not take into account the ANEP classification, but use the ISI categories alone. Because we want to discover the documents hidden behind each category and each link, we create an egocentered star-shaped network of all the categories with scientific output from Spain in 2000.

We depart from a cocitation matrix of 222×222 categories. From there, we build a list of neighbors based on the specific subject area under consideration. We use the Kamada & Kawai algorithm to process the list of neighbors, the name of each category, color and size. As it would be impossible to clearly show the 222 categories, we prune, but this time not with MST. Rather, we take as the threshold the figure obtained from

calculating the average plus the standard deviation of those categories which, within the complete cocitation matrix, belong to the vector of the category that we are going to represent, and which have a value greater than zero. In other words, we use a pruning mechanism with a variable threshold, established by the idiosyncrasy of the links of each central category itself, with respect to the rest. We eliminate the links and vertices that are connected to the central node or category, with a threshold value under the average plus standard deviation. To represent this as a network we again use the Kamada & Kawai algorithm, but now specifying that the value of the distance between vertices is a similarity function of cocitation. Hence, the thickness of the lines is always the same, while their length varies. It is thus obvious which categories are closest to the central one, and share a greater topic affinity. In some instances, when links or node tags overlap, they are manually shifted a bit. The map obtained is exported to the SVG format.

Results

We believe that this type of representation allows any user to grasp, clearly and quickly, the structure of a domain by observing its nodes and links, and their thickness or distance.
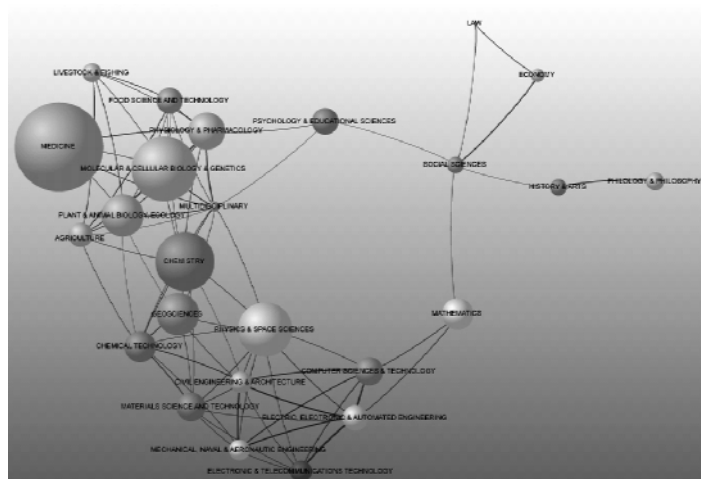
Map of the first level or class cocitation

Figure 2. Class cocitation map

Each class is represented by a node in the network with its name. At first glance, Spain's scientific output for the year 2000 is widely distributed, over the length and width of the graphic area. We note that there is a large "gap" in the center, indicating there is no subject area associated directly with all the others. This is also what happens with MDS representations. Even Multidisciplinary, the class with the greatest number of relationships, is found at a roughly left-center position, closer to the area of Sciences.

From a general standpoint, we can clearly distinguish the three ISI databases. In the leftward area SCI-E is vaguely contoured, wheras in the upper right the SSCI database is reflected, and pendent from it, the contents of what would be the A&HCI. Within the SCI-E zone, there are three prominent blocks: one we could call Life Sciences, including Livestock and Fishing, Food Sciences and Technology, Medicine, Physiology and Pharmacology, Psychology and Educational Sciences, Molecular & Cellular Biology & Genetics, Plant & Animal Biology & Ecology, and Agriculture. Another block would be that of Physics, Chemistry and Earth and Space Sciences, containing Chemistry, Geosciences, Chemical Technology, Physics and Space Sciences and Materials Science & Technology. Finally, the group of Engineering and Computer Sciences would contain Civil Engineering & Architecture, Computer Sciences & Technology, Electric, Electronic & Automated Engineering, Mechanical, Naval and Aeronautical Engineering, Mathematics and Electronic & Telecommunications Technology. We have deliberately left the Multidisciplinary Sciences out of this classification because its contents could belong to all three blocks. In the upper right zone of Social Sciences and Art and the Humanities we quickly infer the SSCI database represented by Social Sciences, Economy and Law, as well as the A&HCI database, with History & Arts and Philology & Philosophy. It is noteworthy that Psychology and Educational Sciences, Mathematics and Social Sciences act as a bridge for the network as a whole, connecting the three major component groups, which coincide with the three ISI databases.

The nodes with a greater number of links occupy more or less central positions easily related with the rest (for example Multidisciplinary Sciences, Physics & Space Sciences or Chemistry), whereas those with fewer links are situated in the periphery, (among others, Philology & Philosophy, Economy, Law and even Medicine). At a glance we note that the link between Social Sciences and Economy is thicker than Social Sciences' links with History & Arts, Law, Psychology & Educational Sciences or Mathematics. Similarly, in the lower part of the Graph, this happens with Electronic & Telecommunications Technology, Electric, Electronic & Automated Engineering, Computer Sciences & Technology, Civil Engineering & Architecture and Mechanical, Naval and Aeronautic Engineering. The thicker links serve as evidence that the use of common sources is greater.

Maps of the second level or cocitation of categories

Each one of the nodes in the class cocitation maps gives rise to the generation of a new map of ISI category cocitation, which gives us a total of 25 maps, as explained in Methodology. For reasons of research affinity, we shall use as an example the map of Social Sciences, which includes the category Library and Information Sciences.
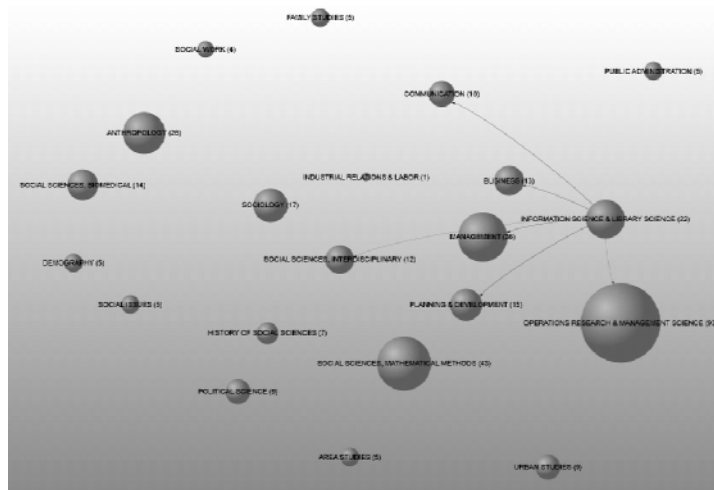

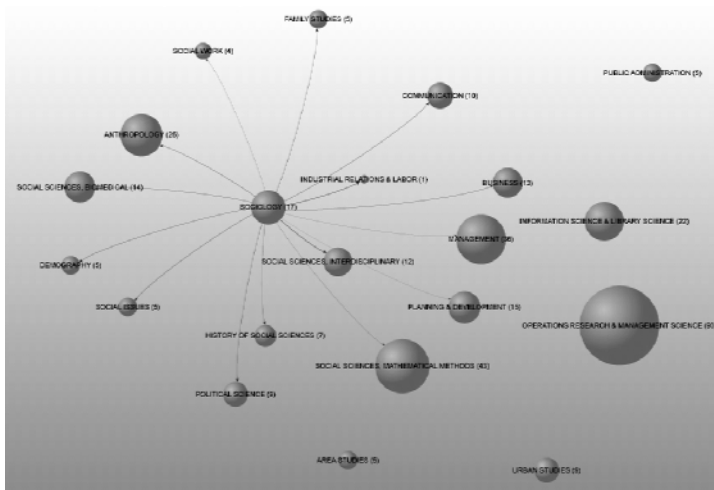Figure 3a. Category cocitation map: Social Sciences

Figure 3b. Category cocitation map: Social Sciences

Every node in the network represents an ISI-JCR category. The numeral alongside is the number of published works in that field produced by Spanish research institutions within the year 2000. In the case of the Social Sciences, the most central nodes are, in this order: Sociology, Planning & Development, Social Sciences-Mathematical Methods, Social Sciences-Interdisciplinary and Management. The rest, including Library & Information Sciences, are in more or less peripheral positions.

As in class cocitation maps, the intensity of the links is shown by their thickness. In the case of the maps of category cocitation alone, however, we have included a small application that allows us to hide the links in order to enhance visibility. The links become visible when the cursor is positioned over one of the nodes or tags, as seen in Figures 3a and 3b.

Of the six links that Library & Information Sciences shares with other nodes, two are very weak – those with Operation Research & Management Science and Social Sciences-Interdisciplinary – while the other four are somewhat stronger. Sociology has fifteen different relationships, with two particularly strong ones: Social Sciences-Interdisciplinary and Industrial Relations & Labor.

Neighbor maps

Each node from the category cocitation maps becomes the starting point of a new map, giving rise to a total of 222 graphs we shall call Neighbor Maps. Tracking down Library & Information Sciences once again, we show its neighbors in Figure 4.
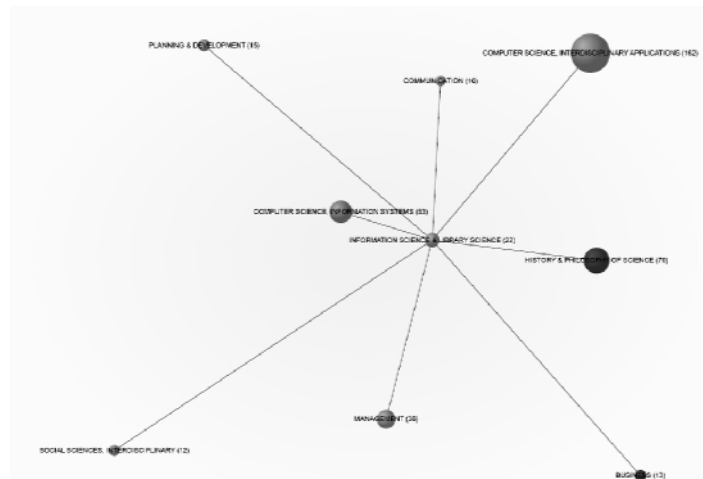

Figure 4. Neighbor map. Social Sciences

The characteristic feature of this map type is that it depicts an egocentered network, where the node studied is always situated in the center, and the rest orbit around it. Although the representation is balanced and tends to occupy all the available space, the intensity of the relationships is reflected here by the distance between nodes. Thus, the most closely related categories are, respectively, Computer Sciences & Information Systems, Communication, History & Philosophy of Science, Management, Computer Sciences & Interdisciplinary Applications, Planning & Development, Business and Social Sciences-Interdisciplinary. The central node or category "attracts" those nodes with which it maintains a closer relationship (in terms of common sources) regardless of the ANEP class it belongs to. This makes it very easy to perceive category groupings and their intensity of correlation without referring to any external classification system.

Conclusions

We have shown that the cocitation of classes or categories can be used to build maps of large geographic domains. This methodology could be applied to comparative studies of similar domains with a geographic reference, or even with a thematic or institutional point of reference. Moreover, the evolution of research fronts within a single domain can be inferred through sequences of representations over time.

In our view, domain analysis is best organized as a hierarchy that would begin with maps of classes, followed by category maps, and then by neighbor maps. The logical continuation of this proposal would be for the latter to be followed by maps of journal cocitation and author cocitation, in that order. Our current work takes us in this direction.

While it is true that we present off-line maps here, we are also testing a set of tools that would allow the automatic generation of schematic representations for a domain the size of Spain. However, we must remember that the re-generation of these maps is required as new information is added to the database (to date, annually). At some future date we may even have the means for dynamic updatings of such maps. For the time being, however, a "demo" or thematic prototype of Spain is available online, as is as a prototype of the institutions of Andalucía, for any interested browsers at www.atlasofscience.net.

References

1. *Framework of a Visualization System* [Web page], Available at:
   http://www.siggraph.org/education/materials/HyperVis/abs_con1/main.htm [Review at: 08/08/2003], (1999).
2. COSTA, J., *La esquemática: visualizar la información, Barcelona*, Paidós, 1998.
3. HJØRLAND, B., ALBRECHTSEN, H., Toward a new horizon in information science: domain analysis, *Journal of the American Society for Information Science (JASIS)*, 46 (1995) 400–425.
4. DOYLE, L. B., Semantic roadmaps for literature searchers, *Journal of the Association for Computing Machinery*, 8 (1961) 553–578.
5. GARFIELD, E., Citation indexes in sociological and historical research, *American Documentation*, 14 (1963) 289–291.
6. GARFIELD, E., SHER, I. H., TORPIE, R. J., *The Use of Citation Data in Writing the History of Science*, Philadelphia, Institute for Scientific Information, (1964).
7. PRICE D., J. D., Networks of scientific papers, *Science*, 149 (1965) 510–515.
8. SMALL, H., Co-citation in the scientific literature: a new measure of the relationship between two documents, *Journal of the American Society for Information Science (JASIS)*, 24 (1973) 265–269.
9. MARSHAKOVA, I. V., System of document connection based on references, *Nauchno-Teknichescaya Informatisya*, Series II (6) (1973) 3–8.
10. SMALL, H., GRIFFITH, B. C., The structure of scientific literature, I: Identifying and graphing specialtyes, *Science Studies*, 4 (1974) 17–40.
11. GRIFFITH, B. C., SMALL, H., STONEHILL, J. A., DEY, S., The structure of scientific literature, II: Toward a macro and microstructure for science, *Science Studies*, 4 (1974) 339–365.
12. AARONSON, S., The footnotes of science, *Mosaic*, 6 (March-April) (1975) 22–27.
13. GARFIELD, E., ISI's Atlas of Science may help students in choice of career in science, *Current Contents*, 29 (July 21) (1975) 5–8.
14. GARFIELD, E., Introducing the ISI Atlas of Science: Biochemistry and molecular biology, 1978-80, *Current Contents*, (42) (1981) 5–13.
15. GARFIELD, E., Introducing the ISI Atlas of Science: Biotechnology and molecular genetics, 1981/82 and bibliographic update for 1983/84, *Current Contents*, (41) (1984) 3–15.
16. GARFIELD, E., The encyclopedic ISI-Atlas of Science launches 3 new sections: bochemistry, inmunology, and animal and plant sicences, *Current Contents*, (7) (1988) 3–8.
17. SEIDEN, L. S., SWANSON, D. R., ISI Atlas of Science: Pharmacology 1987, Vol 1, *Library Quarterly*, 59 (1989) 72–73.
18. SMALL, H., The relationship of information science to the social sciences: a co-citation analysis, *Information Processing & Management*, 17 (1981) 39–50.
19. SMALL, H., GARFIELD, E., The geography of science: disciplinary and national mappings, *Journal of Information Science*, 11 (4) (1985) 147–159.
20. SMALL, H., SWEENEY, E., Clustering the science citation index using co-citations. 1. A comparison of methods, *Scientometrics*, 7 (1985) 391–409.
21. SMALL, H., SWEENEY, E., GREENLEE, E., Clustering the science citation index using co.citations. 2. Mapping science, *Scientometrics*, 8 (1985) 321–340.
22. SMALL, H., Macrolevel changes in the structure of cocitation clusters: 1983-1989, *Scientometrics*, 26 (1993) 5–20.
23. SMALL, H., A SCI-MAP case-study: building a map of aids research, *Scientometrics*, 30 (1994) 229–241.
24. SMALL, H., Visualizing science by citation mapping, *Journal of the American Society for Information Science (JASIS)*, 50 (1999) 799–813.
25. SMALL, H., Charting pathways through science: exploring Garfield's vision of a unified index to science, In: B. CRONIN, H. B. ATKINS (Eds), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Medford, N. J. Information Today, 2000, pp. 449–473.
26. BÖRNER, K., CHEN, C., BOYACK, K. W., Visualizing knowledge domains, *Annual Review of Information Science & Technology*, 37 (2003) 179–255.

27. BRAAM, R. R., MOED, H. F., VAN RAAN, A. F. J., Mapping of Science by combined co-citation and word analysis. I: Structural aspects, *Journal of the American Society for Information Science (JASIS)*, 42 (1991) 233–251.

28. BRAAM, R. R., MOED, H. F., VAN RAAN, A. F. J., Mapping of Science by combined co-citation and word analysis. II: Dynamic aspects, *Journal of the American Society for Information Science (JASIS)*, 42 (1991) 252–266.

29. LIN, X., SOERGEL, D., MARCHIONINI, G., 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, 1991, pp. 262–269.

30. WHITE, H. D., MCCAIN, K. W., Visualization of literatures, *Annual Review of Information Systems and Technology (ARIST)*, 32 (1997) 99–168.

31. WHITE, H. D., MCCAIN, K. W., Visualizing a discipline: an author co-citation analysis of information science, 1972-1995, *Journal of the American Society for Information Science (JASIS)*, 49 (1998) 327–355.

32. GARFIELD, E., *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 [Consulted: 07/05/2003], 1998.

33. WHITE, H. D., LIN, X., MCCAIN, K. W., Two modes of automated domain analysis: multidimensional scaling vs. Kohonen feature mapping of information science authors, In: *Proceedings of the Fifth International ISKO Conference*, Würzberg, Ergon Verlag, 1998, pp. 57–61.

34. DING, Y., CHOWDHURY, G. G., FOO, S., QIAN, W., Bibliometric information retrieval systems (BIRS): a web search interface utilizing bibliometric research results, *Journal of the American Society for Information Science (JASIS)*, 51 (2000) 1190–1204.

35. WHITE, H. D., Toward ego-centered citation analysis, In: B. CRONIN, H. B. ATKINS (Eds), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield,* Medford, N. J. Information Today, 2000, pp. 475–498.

36. NOYONS, E. C. M., MOED, H. F., LUWEL, M., Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study, *Journal of the American Society for Information Science (JASIS)*, 50 (1999) 115–131.

37. BUTER, R. K., NOYONS, E. C. M., Improving the functionality of interactive bibliometric science maps, *Scientometrics*, 51 (2001) 55–67.

38. NOYONS, E. C. M., Bibliometric mapping of science in a science policy context, *Scientometrics*, 50 (2001) 83–98.

39. CHEN, C., PAUL, R. J., Visualizing a knowledge domain's intellectual structure, *Computer*, 34 (3) (2001) 65–71.

40. CHEN, C., PAUL, R. J., O'Keefe, B., Fitting the jigsaw of citation: information visualization in domain analysis, *Journal of the American Society for Information Science and Technology (JASIST)*, 52 (2001) 315–330.

41. DING, Y., CHOWDHURY, G. G., FOO, S., Bibliometric cartography of information retrieval research by using co-word analysis, *Information Processing & Management*, 37 (2001) 801–817.

42. INGWERSEN, P., LARSEN, B., Mapping national research profiles in social science disciplines, *Journal of Documentation*, 57 (2001) 715–740.

43. GUERRERO BOTE, V. P., MOYA ANEGÓN, F. d., HERRERO SOLANA, V., Document organization using Kohonen's algorithm, *Information Processing & Management*, 38 (2002) 79–89.

44. WHITE, H. D., BUZYDLOWSKI, J., LIN, X., Co-cited author maps as interfaces to digital libraries: designing Pathfinder Networks in the humanities, In: *IEEE International Conference on Information Visualization*, London, 2000, pp. 25–30.

45. BUZYDLOWSKI, J., WHITE, H. D., LIN, X., Term co-occurrence analysis as an interface for digital libraries, *Lecture Notes in Computer Science Series*, 2539 (2002) 133–144.

46. LIN, X., WHITE, H. D., BUZYDLOWSKI, J., Real-time author co-citation mapping for online searching, *Information Processing & Management*, (2003) 689–706.

47. CHEN, C., *Information Visualization and Virtual Environments*, Berlin, Springer, (1999).

48. WHITE, H. D., Author-centered bibliometrics through CAMEOs: characterizations automatically made and edited online, *Scientometrics*, 51 (2001) 607–637.

49. SMALL, H., Paradigms, citations and maps of science: a personal history, *Journal of the American Society for Information Science and Technology (JASIST)*, 54 (2003) 394–399.

50. CHEN, C., KULJIS, J., The rising landscape: a visual exploration of superstring revolutions in physics, *Journal of the American Society for Information Science and Technology (JASIST)*, 54 (2003) 435–446.

51. MORRIS, S. A., YEN, G., WU, Z., ASNAKE, B., Time line visualization of research fronts, *Journal of the American Society for Information Science and Technology (JASIST)*, 54 (2003) 413–422.

52. BOYACK, K. W., BÖRNER, K., 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 and Technology (JASIST)*, 54 (2003) 447–461.

53. TUFTE, E. R., *Envisioning Information*, Cheshire, Graphics Press, 1994.

54. TUFTE, E. R., *The Visual Display of Quantitative Information*, Cheshire, Graphics Press, 2001.

55. Thomson-ISI, *ISI Journal Citation Reports* [Web page], Available at: http://isi6.isiknowledge.com [Review at: 09/07/2003], 2003.

56. ANEP, *Agencia Nacional de Evaluación y Prospectiva* [Web page], Available at: http://www.mcyt.es/sepct/ANEP/anep.htm [Review at: 09/07/2003], (2003).

57. Thomson-ISI, *ISI Web of Science* [Web page], Available at: http://wos.cica.es/

58. SALTON, G., BERGMARK, D., A citation study of computer science literature, *Professional Communication, IEEE Transaction*, PC-22 (1979) 146–158.

59. KAMADA, T., KAWAI, S., An algorithm for drawing general undirected graphs, *Information Processing Letters*, 31 (1989) 7–15.

60. WHITE, H. D., Pathfinder networks and author cocitation analysis: a remapping of paradigmatic information scientist, *Journal of the American Society for Information Science and Technology (JASIST)*, 54 (2003) 423–434.

61. W3C, *Scalable Vector Graphics (SVG)* [Web page], Available at: http://www.w3.org/Graphics/SVG/ [Review at: 10/09/2003], 2003.

62. W3C, *Scalable Vector Graphics (SVG) 1.1 Especification* [Web page], Available at: http://www.w3.org/TR/SVG11/ [Review at: 12/09/2003], 2003.