

A new approach to institutional domain analysis: Multilevel research fronts structure

SANDRA MIGUEL,^a FÉLIX MOYA-ANEGÓN,^b VÍCTOR HERRERO-SOLANA^b

^a *Library Science Department, National University of La Plata, La Plata (Argentina)*

^b *Library and Information Science Department, University of Granada, Granada (Spain)*

The intellectual structure and main research fronts of the Faculty of Natural Sciences and Museum of the National University of La Plata, Argentina is studied, based on the cocitation analysis of subject categories, journals and authors of their scientific publications collected in the Science Citation Index, CD-ROM version, for the period 1991–2000. The objective of this study is to test the utility of those techniques to explore and to visualize the intellectual structure and research fronts of multidisciplinary institutional domains. Special emphasis is laid on the identification of multilevel structures, by means of arrangements of subject categories cocitation analysis and journal cocitation analysis.

Introduction

It is generally assumed that cocitation analysis of documents, authors, journals, specialties or fields yield valid representations of intellectual structure of scientific domains. The fundamental premise of cocitation analysis is that among two or more documents, authors, journals or other entities that are cited together in a third posterior work there exists, at least from the perspective of the citing author, a subject similarity; and that the greater the frequency of cocitation the more closed is the relation among them. This technique, initially presented by Small and Marshakova for the cocitation

Received September 19, 2005

Address for correspondence:

SANDRA MIGUEL

Library Science Department, National University of La Plata

48 e/ 6 y 7 (1900) La Plata, Argentina

E-mail: sandra@fcnym.unlp.edu.ar

0138–9130/US \$ 20.00

Copyright © 2007 Akadémiai Kiadó, Budapest

All rights reserved

analysis of documents [1,2], and by White and Griffith for cocitation analysis of authors (ACA) [3] was used in numerous later works with the object of revealing different aspects of the intellectual structure and the composition of scientific fields, identifying scientific specialties, research fronts and changes in the intellectual structures over time.

Most noteworthy among the methods and statistical techniques used for this sort of study are: clustering, multidimensional scaling (MDS), factor analysis, or some combination thereof [4]. These techniques can be used to make groupings of the entities (whether they be documents, authors, journals) and to construct maps that portray the intellectual structure of the scientific domains in two- or three-dimensional space. In order to interpret these maps, one must bear in mind the localization of these entities as single points and groups of points on the map – their central or else peripheral position, and the proximity of points within the groups as well as their relationship with other groups. In general, the interpretation and validation of the maps may entail inspection, comparison with preexisting data, consultation with specialists, and statistical tests[5].

Among the many studies of thematic domains we might mention Small's work with Particle Physics [1]; those of White and others, Moya-Anegón and others, and Ding and others in the field of Library and Information Science [3,6,7]; Karki in Sociology of science and Information Science [8]; McCain in Macroeconomics, Drosophila Genetics and Biotechnology [9,10,11,12]; Paisley in the area of Communication [13]; Bayer and others in literature about marriage and family [14]; Culnan and others about Organizational Behaviour [15]; Eom and others regarding Decision Support Systems (DSS) [16,17]; Kreuzman in Philosophy [18], and Tsay and others on literature concerned with semiconductors [19,20]; Andrews in medical informatics[21].

There are also studies about the intellectual structure underlying scientific journals, such as that of Persson applied to the *Journal of the American Society for Information Science* [22], and that of Moya-Anegón and Herrero-Solana on the Mexican journal *Investigación Bibliotecológica*[23].

A quick glance at the analysis of institutional domains reveals its scarcity and moreso when Latin American countries are involved. One previous study offers an initial approach to the intellectual structure of the National University of Mar del Plata, Argentina [24].

The objective of this study is to test the utility of techniques cocitation analysis to explore and to visualize the intellectual structure and research fronts of multidisciplinary institutional domains. In the present case the Faculty of Natural Sciences and Museum (*Facultad de Ciencias Naturales y Museo, or FCNyM*) of the National University of La Plata (*Universidad Nacional de La Plata, or UNLP*), Argentina is studied, based on the cocitation analysis of subject categories, journals and authors of their international publications collected in the Science Citation Index CD-ROM version (SCI-CD), over the period 1991–2000.

The FCNyM is one of the academic units of the UNLP with the greatest scientific tradition. Its origins can be traced back to 1877, with the creation of the Museo de La Plata. In the 1890's it already possessed a structure for investigation in the field of the Natural Sciences; and in 1906, with the creation of the National University of La Plata, it also incorporated superior education functions, thus becoming the first Faculty of Natural Sciences in the country [25]. At present there are three undergraduate degree programs: Biology (with orientations in Zoology, Botany, Ecology and Paleontology); Geology and Geochemistry; and Anthropology (including Archaeology, Physical Anthropology and Social Anthropology); as well as the graduate program of Ph.D. in Natural Sciences. The number of undergraduate students is around 2000, with some 230 additional graduate students. Their scientific structure is conformed by Divisions, Institutes, Laboratories where the scientific activities are carried out by more 400 professors and reches in several disciplines and scientific specialties in the field of natural sciences.

Materials and methods

As sources for the collection of data we used: the Science Citation Index database, CD-ROM version (SCI-CD), from the Institute for Scientific Information (ISI) of Philadelphia, covering the period 1991–2000 (although data is not complete for the year 2000); and the Journal Citation Report (ISI-JCR).

We downloaded from SCI-CD the records with at least one Argentinian address in the field "address", and put them into an *ad hoc* database for consultation. Within this subset of records we identified the works where at least one author indicated La-Plata (the city considered the headquarters of the FCNyM) as part of the data of institutional affiliation. The records corresponding to authors from the FCNyM were identified and grouped, adding others from this institution whose location was not in La Plata.* The names of the authors and of the journals and other sources cited in the bibliography of the contributions were normalized. The bibliographic references were grouped in the subject categories assigned to the ISI-JCR.**

With the subject categories most frequently cited (minimum of 15 citations) we built a square and symmetric matrix containing the cocitation frequencies that each one had with others. The values of the diagonal of this matrix were calculated by multiplying

* To ensure greater reliability in the selection of records, the author names from the selected records were compared with a list of professors and researchers of the institution, provided by the Secretary of Research and Transference.

** The description of the scope of each subject category assigned by the ISI to the journals indexed can be consulted at <http://www.isinet.com>

the highest value of cocitation frequency of each one of these with the rest by 1.5.* Two techniques of multivariate analysis of data were then applied: MDS and cluster analysis (hierarchical clustering and the Ward's method), after conversion of the matrix of cocitation frequency into matrices of similarities and distances using Pearson's r and $1 - \text{Pearson's } r$ correlation coefficients** for the respective analyses. With the results obtained for a solution in two dimensions (2D), maps were constructed to give a rough representation of the composition and disciplinary structure of the Faculty. Having arrived at the thematic groupings represented in this structure, we applied a similar procedure for the journals and for the other sources cited frequently in each one of these groups (minimum among 9 and 15 citations), and the main research fronts were identified. Only in one of these subject groups was it possible to carry out the cocitation analysis of authors gathering a considerable number of authors frequently cited and co-cited.

Likewise, and as an indicator of the goodness-of-fit of the representations resulting from MDS, the statistical stress was measured, setting 0.2 as the maximum acceptable value.***

Results

The FCNyM has a total of 481 scientific publications collected in the SCI-CD for the period 1991–2000, which include a total of 15,933 bibliographic references. Some 11,797 references correspond to different documents, of which 88% were cited only one time. Moreover, a noteworthy dispersion of the journals or other types of cited sources in the bibliography is observed. The 15,933 references correspond to a total of 5,769 different titles of journals or other sources, 68% of which were cited just once (Table 1). In apparent contrast with this dispersion, 42% of all cites (6,748) are concentrated in 265 titles that receive between 10 and 254 mentions. Altogether, 67% (179) of these cited sources are journals indexed by the ISI, and they accumulate 83% of the citations of this group (Figure 1).

* This measurement was used to guarantee that the values of the diagonal would be consistently higher than any of the frequencies of cocitation of a discipline, journal or author, an objective that could not be achieved using other solutions proposed in the literature [3].

** The standardization of the matrices of cocitation frequencies using measures of similarity and distance is a widely used strategy for reducing the scale effects that result from the different cocitation frequencies among units to be studied.

*** The values of stress reported for an MDS solution are usually stress I or Stress II of Kruskal 1977. The best solution is the one with the minimal stress value. For a solution in 2 or 3 dimensions, a stress value of 0.2 is acceptable[5].

Table 1. Distribution of cites for different journals or other sources included in the reference list

Cites (#)	Journals (#)	%
1	3938	68.3
≥2 <10	1566	27.1
≥10 <20	157	2.7
≥20 <30	55	1.0
≥30 <50	28	0.5
≥50 <100	19	0.3
≥100	6	0.1

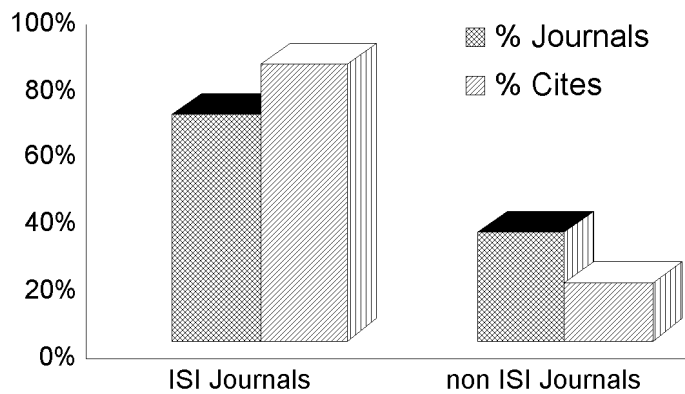


Figure 1. Percentage of cites to papers published in ISI and non-ISI journals

The disciplinary profile of the institution is made up of some 53 subject categories (Table 2). From the analysis of the cocitation of these disciplines we came up with two graphic representations, resulting from MDS and cluster analysis, which provide an initial overall view of the intellectual structure of the institution.

In the map of subject cocitation, Figure 2, the categories are represented as bubbles whose size is directly proportional to the citation frequency. In the dendrogram from cluster analysis in Figure 3 the same groups are shown in form of hierarchic tree.

Table 2. Total number of cites and cocites for 53 subject categories

Category	Cites	Cocites
Plant Sciences	1061	8046
Zoology	631	6221
Genetics & Heredity	487	8219
Paleontology	467	4272
Multidisciplinary Science	418	6045
Biochemistry & Molecular Biology	403	6855
Geosciences, Interdisciplinary	386	5512
Marine & Freshwater Biology	355	4320
Oncology	322	5865
Environmental Sciences	312	5063
Ecology	306	4112
Biology	255	4014
Geochemistry & Geophysics	219	3430
Geology	206	3071
Parasitology	187	502
Biotechnology & Appl Microbiology	162	3001
Hematology	158	2881
Entomology	157	1199
Agriculture	156	1436
Oceanography	153	2861
Cell Biology	131	2759
Limnology	106	1344
Microbiology	103	1422
Biology, Miscellaneous	88	1666
Toxicology	72	1803
Engineering, Environmental	70	1400
Mycology	68	296
Pathology	62	1511
Anthropology	59	599
Meteorology & Atmospheric Sci.	59	982
Agriculture, Soil Science	57	492
Mineralogy	45	423
Biophysics	39	682
Virology	39	533
Immunology	38	1012
Chemistry	36	736
Fisheries	35	478
Chemistry, Analytical	33	227
Physiology	31	445
Forestry	30	263
Astronomy & Astrophysics	28	701
Public, Environ. & Occup Health	28	283
Food Science & Technology	27	373
Medicine, Res. & Experimental	27	742
Agriculture, Dairy & Animal Sci.	24	308
Biochemical Research Methods	24	487
Geography	24	746
Medicine, General & Internal	24	533
Water Resources	24	368
Anatomy & Morphology	21	145
Veterinary Sciences	18	290
Chemistry, Applied	17	221
Engineering, Civil	16	225

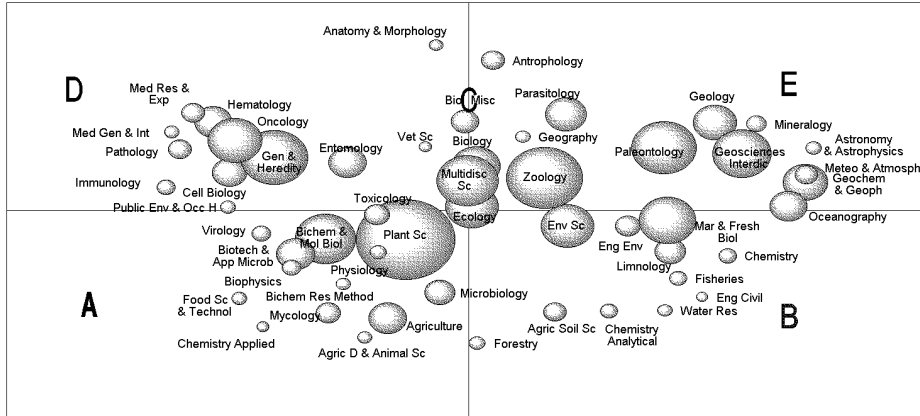


Figure 2. Map of subject cocitation (MDS) (accounted for 42 percent of all cites)

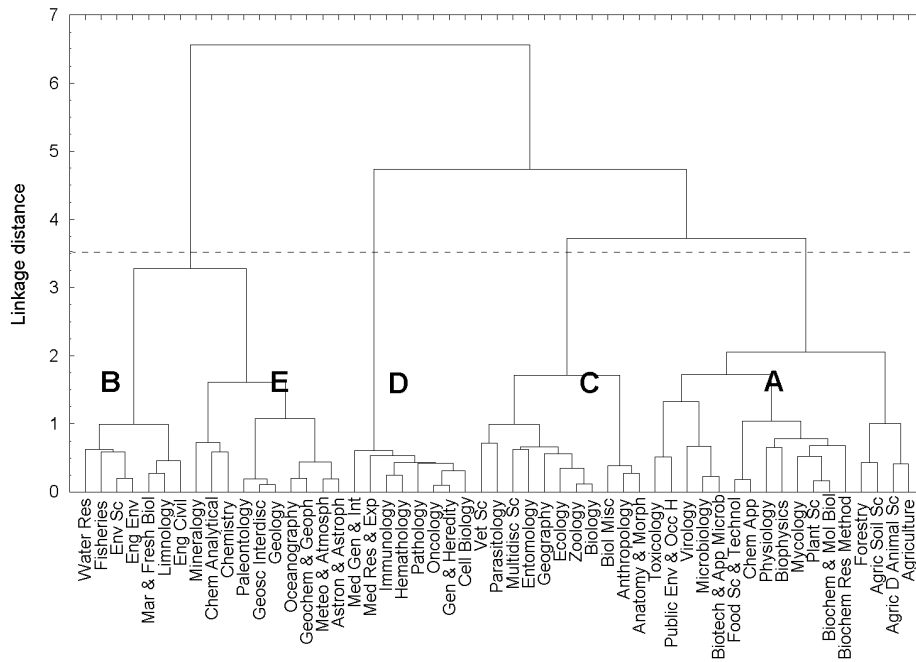


Figure 3. Dendrogram of subject cocitation (Cluster analysis) Ward's method 1 – Pearson's r

These graphs show a heterogeneous disciplinary structure reflecting a network of interdisciplinary relationships. The clusters by disciplines and sub disciplines represented in the dendrogram with a distance of relation of $3 < y < 4$ reveal a structure made up of five main thematic groups that are identified in the map of Figure 2 as A, B, C, D and E.

Group A, situated in the lower left corner of the map, represents the field of Botany, Plant and Fungal Biochemistry and Agriculture. Group B, in the lower right section, reflects research in Freshwater Biology. Group C, in the center of the map, is concerned with a wide array of studies, including Zoology, Parasitology, Entomology, and Ecology. Group D, in the upper left-hand section, represents research in Cell Biology and Biomedicine with a strong presence in the disciplines of Genetics and Heredity, and Oncology. Group E, in the upper right part of the map, shows research in Paleontology and Geosciences.

In the case of group C, the central position it occupies on the map reflects the important degree of interrelation of its component disciplines with others, particularly the ones from groups A, B and E. The presence of Ecology in the center of the map is not only due to its ties to Botany (group A), to Biology and Ecology of aquatic media (group B) and to Paleo-environmental studies (group E), but it more importantly reveals the essential nucleus of the institution's research. The intermediate position of Zoology with respect to groups B and E stems from its connections with the studies of freshwater Biology and of Paleontology. The situation of Entomology halfway between group C and group D is mainly due to the presence of studies about insect genetics. Finally, the presence of the category Multidisciplinary Science in the center of the map evidences that the studies published in multidisciplinary journals are cited in the bibliography of the scientific contributions of the bulk of the disciplines and sub disciplines that make up the structures.

In order to identify the composition of each thematic group, the journals and other sources most frequently cited in each one of them were analyzed. Cocitation analysis of the entities included in each group led us to identify the main research fronts.

Group A (Figure 4) contains four research fronts tied to the area of Botany and related disciplines: Taxonomy and Plant Phylogeny (front 1); Mycology (front 2) Plant Physiology (front 3) and Paleobotany and Palynology (front 4). Some journals near the center of the map would be common to the different fronts. Such are the cases of the *Proceedings of the National Academy of Science of the United States* (P Natl Acad Sci USA), which is multidisciplinary in nature, the *Canadian Journal of Botany* (Can J Bot), cited in the contributions of all four fronts, *Annals of Botany* (Ann Bot London) and *Science*, cited by fronts 1, 2 and 3, respectively situated in the left, lower and right sectors.

In group B (Figure 5) we identified three research fronts. The first of these is tied to studies of Biogeochemistry in aquatic media, especially concerned with contamination.

The second front is dedicated to studies of Biology and Ecology of freshwater environments. The third is dedicated to Limnology. Fronts 2 and 3 share a considerable number of journals.

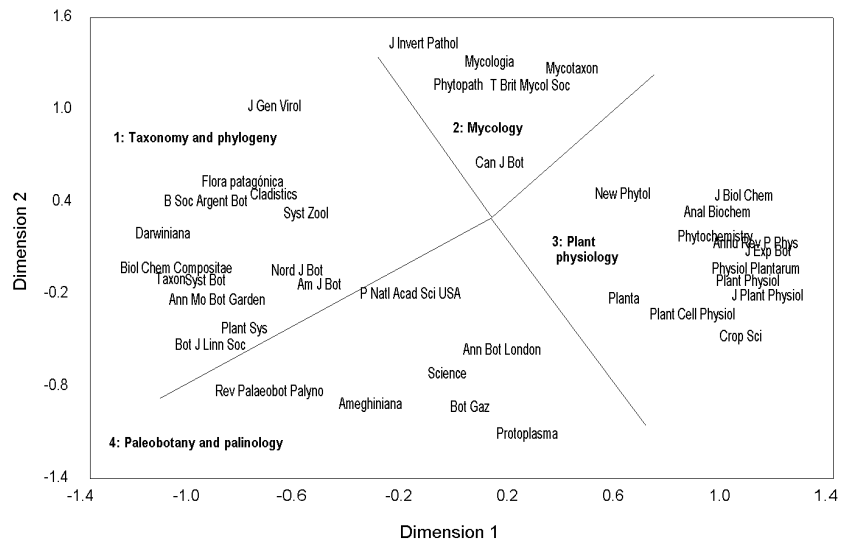


Figure 4. Map of journals cocitation (MDS) – Group A

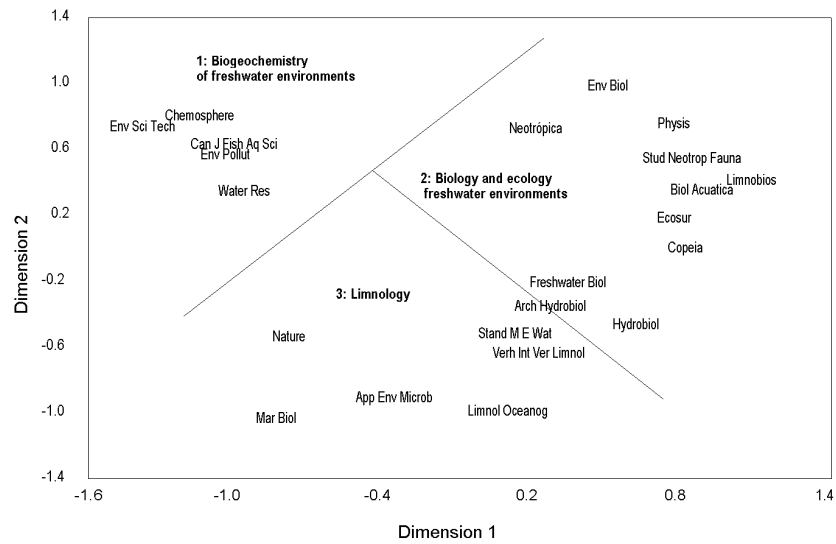


Figure 5. Map of journals cocitation (MDS) – Group B

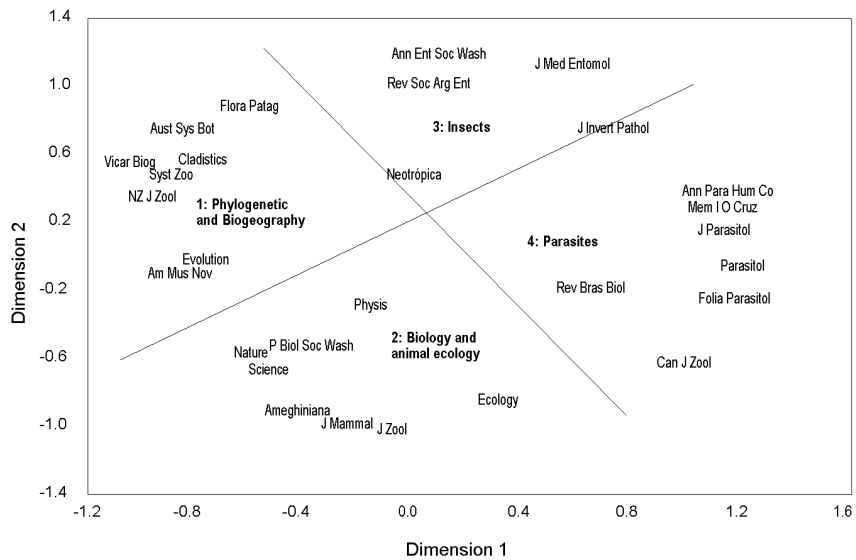


Figure 6. Map of journals cocitation (MDS) – Group C

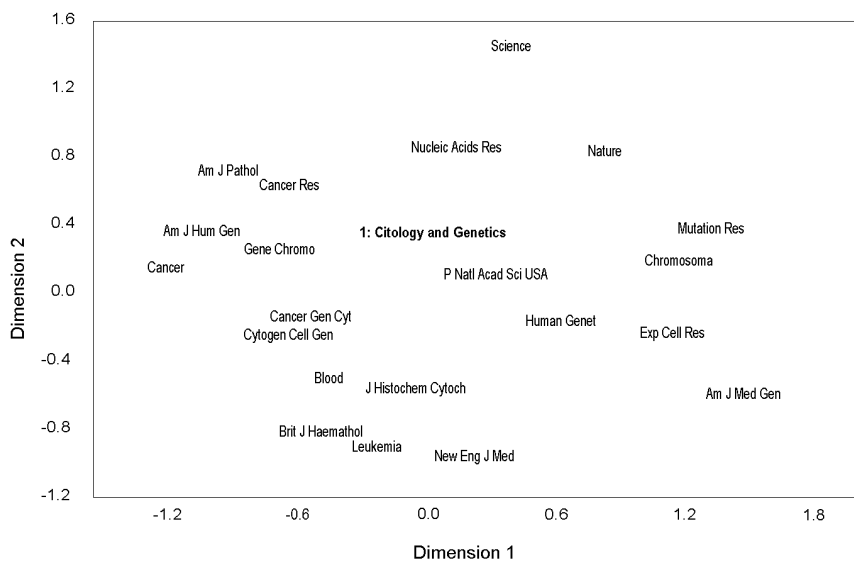


Figure 7. Map of journals cocitation (MDS) Group D

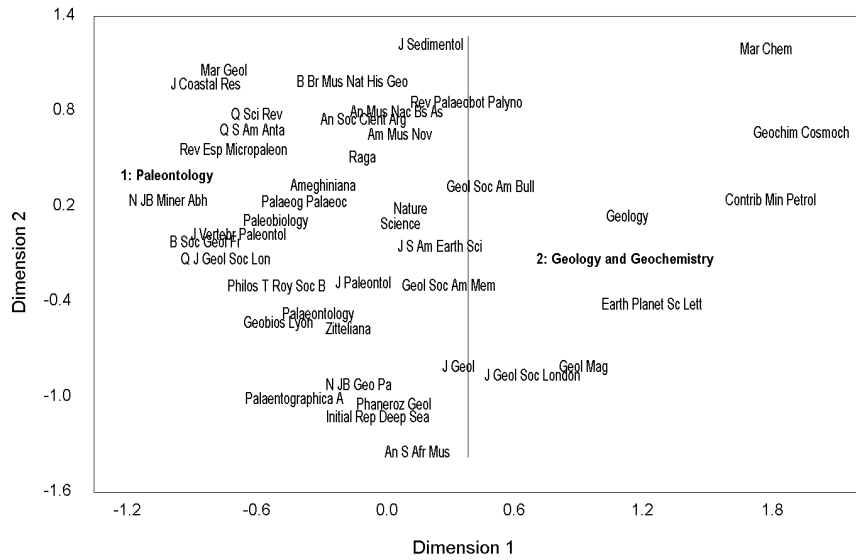


Figure 8. Map of journals cocitation (MDS) Group E

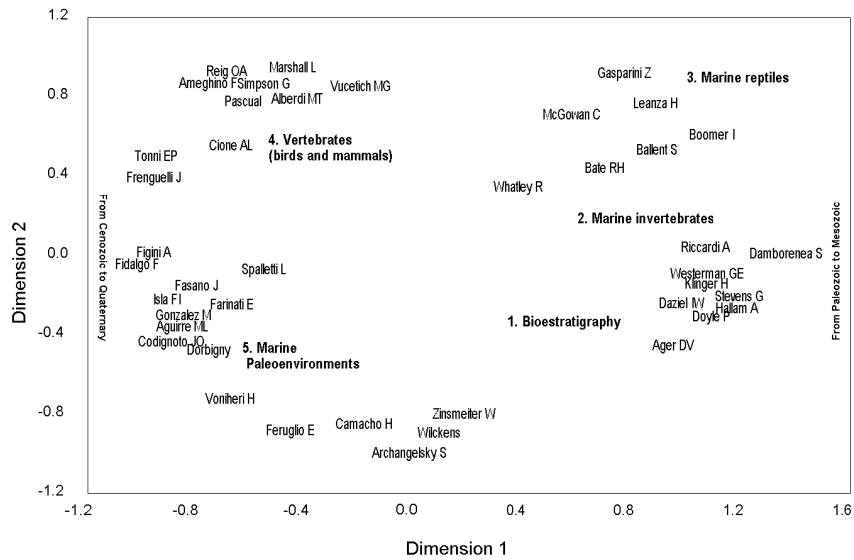


Figure 9. Map of authors cocitation (MDS) Group E

Group C (Figure 6) mainly comprises four fronts. In the upper left quadrant are the studies about Phylogeny and Biogeography (front 1); in the lower left quadrant are

animal Biology and Ecology (front 2), in the upper central right part of the map are studies about insects (front 3), and the right sector of the map reflects studies about parasites (front 4). The journals closest to the center of the map, such as *Neotrópica* and *Physis*, are largely shared by fronts 1, 2 and 3. Also appearing in nuclei are multidisciplinary journals such as *Science* and *Nature*, cited in the contributions of fronts 1 and 2. The presence of the *Journal of Invertebrate Pathology* (J Invert Pathol) amid fronts 3 and 4 is due to studies about vectors and parasitic insects.

In group D (Figure 7) we identified a research front concerned with Cytology and Human Genetics. Although they cannot be considered to constitute a second front, there is a noteworthy presence of cytogenetic studies of non-human species also apparent in the map.

Group E (Figure 8) is strongly represented by the studies of Paleontology situated in the central to left part of the map. Also appearing are studies of Geology and Geochemistry. To detect the fronts in this grouping, it was necessary to identify the authors most frequently cited and perform author cocitation analysis. In the map resulting from MDS (Figure 9), five research fronts can be visualized. Those of the right sector of the map reflect, from bottom to top, the evolution over time from the Paleozoic to the Mesozoic. The first is represented by biostratigraphical studies of the Paleozoic (especially the Ordovician) and the Mesozoic; the second is linked to marine invertebrates, and the third to the study of marine reptiles. In the left sector of the map we have, from top to bottom, the geological and paleontological studies from the Cenozoic to the late Quaternary. Here two other fronts appear: vertebrates, with a strong presence in research in birds and mammals (front 4) and marine paleoenvironment, evolution of coastlines and marine deposits (front 5).

In addition to the analysis of the composition of each segment of the structure, the overall vision of these maps reveals that each group not reflects a total representation of all the disciplines that are included in the structure of Figures 2 and 3. This is because these disciplines do not have a sufficient presence in terms of citations of journals or authors.

Another aspect that is relevant in this analysis is the presence of the same journals cited and cocited in different groups. In some cases this is due to the fact that the segmentations carried out in the structures – in order to analyze the disciplinary composition and the fronts of research – produced fragmentations of the interdisciplinary relationships. One example of this can be seen in fronts: 1- Taxonomy and Phylogeny of group A; and 1- Phylogenetic studies and Biogeography of group C. The two “share” the journals *Cladistics*, *Systematic Zoology* and *Flora patagónica*; and so the rest of the journals that make up the front give a sharper definition of the profile of each. In other cases, there are journals cited by more diverse research teams. Such is the case of the journal *Ameghiniana*, with presence in front 4: Paleobotany and Palinology of group A; in front 4: Biology and animal ecology of group C; and in the studies of Paleontology of group E. Also deserving mention is that this journal receives

the greatest number of citations by the FCNyM. Similarly, *Neotrópica* is present in front 2 (Biology and Ecology of freshwater environments in group B), and tied mainly with front 3 (insects, group C). Yet in this case the reason lies in the general character of this journal within the field of natural sciences. Finally, the journals classified as being from multidisciplinary sciences (*Science*, *Nature* among others) are present in all the disciplinary groups of the structures.

Conclusions

The results of this study show that the Faculty of Natural Sciences and Museum (FCNyM) of the National University of La Plata (UNLP), Argentina, has a characteristically heterogeneous intellectual structure, made up of a network of interdisciplinary relationships and several research fronts in the area of Natural Sciences, especially tied to studies in Botany, Biology, Ecology, Cytogenetic and Paleontology. Ecology is seen to be the discipline that embraces most of the fronts, revealing its outstanding importance in the disciplinary profile of the institution.

To confront these results with those of previous studies is somewhat difficult, as one institutional domain can hardly be equal to another. Furthermore, most studies are dedicated to the analysis of specific disciplinary domains. Unlike the disciplinary domain studies in which one of the main difficulties reported resides in the selection of the entities (authors, journals, etc.) to include in the analysis [6, 14, 26], in institutional domains these entities are indeed given. For institutions with a multidisciplinary profile, the difficulties lie both in determining the boundaries between subject groupings that arise in the disciplinary structure, and in establishing the cutoff point for citation and cocitation frequency in order to guarantee the validity of the representations obtained.

Although revealing the intellectual structure of multidisciplinary institutional domains such as the case of the Faculty studied here is no simple task, the results, as presented here, show that the identification of multilevel structures by techniques of cocitation analysis is a valid alternative for giving an initial approach of their disciplinary profiles.

References

1. H. SMALL, Cocitation in the scientific literature: a new measure of the relationship between two documents, *Journal of the American Society for Information Science*, 24 (4) (1973) : 265–269.
2. I. V. MARSHAKOVA, System of document connection based on references, *Nauchno-Tekhnicheskaya Informatsiya*, Series II, (6) (1973) : 3–8.
3. H. D. WHITE, B. C. GRIFFITH, Author cocitation: a literature measure of intellectual structure, *Journal of the American Society for Information Science*, 32 (3) (1981) : 163–171.
4. H. D. WHITE, K. W. MCCAIN, Visualization of literatures, *Annual Review of Information Science and Technology*, 32 (1997) : 99–168.

5. K. W. MCCAIN, Mapping authors in intellectual space: a technical overview, *Journal of the American Society for Information Science*, 41 (6) (1990) : 433–443.
6. F. D. MOYA-ANEGÓN, E. JIMÉNEZ, M. D. L. MONEDA, Research fronts in Library and Information Science in Spain (1985–1994), *Scientometrics*, 42 (2) (1998) : 229–246.
7. Y. DING, G. CHOWDHURY, S. FOO, Mapping the intellectual structure of information retrieval studies: an author cocitation analysis, 1987–1997, *Journal of Information Science*, 25 (1) (1999) : 67–78.
8. R. KARKI, Searching for bridges between disciplines: an author cocitation analysis on the research into scholarly communication, *Journal of Information Science*, 22 (5) (1996) : 323–334.
9. K. W. MCCAIN, Longitudinal author cocitation mapping: the changing structure of macroeconomics, *Journal of the American Society for Information Science*, 35 (6) (1984) : 351–359.
10. K. W. MCCAIN, The structure of biotechnology R & D, *Scientometrics*, 32 (2) (1995) : 153–175.
11. K. W. MCCAIN, Cocited author mapping as a valid representation of intellectual structure, *Journal of the American Society for Information Science*, 37 (3) (1986) : 111–122.
12. K. W. MCCAIN, Evaluating cocited author search performance in a collaborative specialty, *Journal of the American Society for Information Science*, 39 (6) (1988) : 428–431.
13. W. PAISLEY, An oasis where many trails cross: the improbable cocitation networks of a multidiscipline, *Journal of the American Society for Information Science*, 41 (6) (1990) : 459–468.
14. A. E. BAYER, J. C. SMART, G. W. MCLAUGHLIN, Mapping intellectual structure of a scientific through author cocitations, *Journal of the American Society for Information Science*, 41 (6) (1990) : 444–452.
15. M. J. CULNAN, C. A. O'REILLY, J. A. CHATMAN, Intellectual structure of research in organizational behavior, 1972–1984: a cocitation analysis, *Journal of the American Society for Information Science*, 41 (6) (1990) : 453–458.
16. S. B. EOM, Mapping the intellectual structure in research of decision support systems through author cocitation analysis (1971–1993), *Decision Support Systems*, 16 (1996) : 315–338.
17. S. B. EOM, R. S. FARRIS, The contributions of organizational science to the development of decision support systems research subspecialties, *Journal of the American Society for Information Science*, 47 (12) (1996) : 941–952.
18. H. KREUZMAN, A cocitation analysis of representative authors in philosophy: Examining the relations between epistemologists and philosophers of science, *Scientometrics*, 51 (3) (2001) : 525–539.
19. M. Y. TSAY, H. XU, C. W. WU, Author cocitation analysis of semiconductor literature, *Scientometrics*, 58 (3) (2003) : 529–545.
20. M. Y. TSAY, H. XU, C. W. WU, Journal cocitation analysis of semiconductor literature, *Scientometrics*, 57 (1) (2003) : 7–25.
21. J. E. ANDREWS, An author cocitation analysis of medical informatics, *Journal of the Medical Library Association (JMLA)*, 91 (1) (2003) : 47–56.
22. O. PERSSON, The intellectual base and research fronts of JASIS 1986–1990, *Journal of the American Society for Information Science*, 45 (1) (1994) : 31–38.
23. F. D. MOYA-ANEGÓN, V. HERRERO-SOLANA, Visibilidad internacional de la producción científica iberoamericana en Biblioteconomía y Documentación (1991–2000), *Ciencia da informacao*, 31 (3) (2002) : 54–65.
24. V. HERRERO-SOLANA, Producción científica de la Universidad Nacional de Mar del Plata: Análisis de dominio, *Nexos*, 8 (14) (2001) : 4–10.
25. M. E. TERUGGI, *Museo de La Plata 1888–1998: Una centuria de honra*, Fundación Museo de La Plata, La Plata, 1988.
26. H. D. WHITE, K. W. MCCAIN, Visualizing a discipline: an author cocitation analysis of information science, 1972–1995, *Journal of the American Society for Information Science*, 49 (4) (1998) : 327–355.