IMPROVING SCIMAGO JOURNAL & COUNTRY RANK (SJR) SUBJECT CLASSIFICATION THROUGH REFERENCE ANALYSIS

Gómez-Núñez, Antonio Jesus (1-5); Vargas-Quesada, Benjamín (2-5); Moya-Anegón, Félix (3-5); Glänzel, Wolfgang (4)

(1) CSIC – Institute of Public Goods and Policies (IPP), Madrid (Spain). E-mail: antoniojesus.gomez@cchs.csic.es
(2) University of Granada – Faculty of Communication and Documentation
(3) CSIC – Institute of Public Goods and Policies (IPP), Madrid (Spain)
(4) Centre for R&D Monitoring (ECOOM) and Dept. MSI, K.U. Leuven, Leuven (Belgium)
(5) SCImago Research Group Associated Unit, Granada (Spain)

Abstract: In order to re-categorize the SCImago Journal & Country Rank (SJR) journals based on Scopus, as well as improve the SJR subject classification scheme, an iterative process built upon reference analysis of citing journals was designed. The first step entailed construction of a matrix containing citing journals and cited categories obtained through the aggregation of cited journals. Assuming that the most representative categories in each journal would be represented by the highest citation values regarding categories, the matrix vectors were reduced using a threshold to discern and discard the weakest relations. The process was refined on the basis of different parameters of a heuristic nature, including 1) the development of *several tests applying different thresholds*, 2) the designation of a *cutoff*, 3) the *number of iterations* to execute, and 4) a *manual review* operation of a certain amount of multi-categorized journals. Despite certain shortcomings related with journal classification, the method showed a solid performance in grouping journals at a level higher than categories —that is, aggregating journals into subject areas. It also enabled us to redesign the SJR classification scheme, providing for a more cohesive one that covers a good proportion of re-categorized journals.

Keywords: Reference Analysis, Journal Classification, Subject Categorization, Multidisciplinary Databases, SCImago Journal & Country Rank.

Introduction

Problems related to the classification of scientific knowledge have been widely discussed by scholars and researchers from different disciplines throughout history. In the limelight of the Library and Information Science field stand contributions by figures such as Dewey, Otlet, Ranganathan or Hjørland, for instance. According to Glänzel (Glänzel & Schubert 2003) "classification of science into a disciplinary structure is at least as old as science itself". However, the facet of human knowledge that Chen described as a "complex and dynamic network" (Chen 2008) likewise complicates the development of any reliable and representative disciplinary classification scheme that might allow to effectively delimit different subjects or disciplines configuring this highly complex network.

Knowledge takes place as a result of the curiosity and interest of human beings focused on explaining the surrounding environment and the phenomena taking place. To this end, it is necessary to conduct research-related processes that may be considered inherent to human being and essential to reach knowledge. Nowadays, research is influenced by factors such as its strong relationship with society, the ultra-specialization of areas and disciplines of knowledge, the competitiveness exercised by increasing practitioners, groups and research institutions, or the dynamism resulting from new trends and fashions. All this gives rise to a considerable growth of literature, as well as a constant restructuring and redefinition of the areas and disciplines of scientific knowledge, which ultimately interferes with our ability to design and implement classification systems that represent scientific knowledge.

Databases, regarded as great repositories responsible for storing the results of scientific research, require the use of efficient classification schemes. This is a vital need not only when searching and retrieving information, but also for preparing cohesive and reliable bibliometric analyses. Isabel Gomez (Gómez et al. 1996), in the context of growing interdisciplinary research, underlined the meteoric changes involving disciplines and journals (titles changing, journals merging, etc.), or the establishment of classification systems targeted to the specific interests of each particular database as common problems regarding the organization of recorded scientific knowledge. Although a number of models can be adopted to classify the contents of the various scientific databases, two major multidisciplinary databases — the ISI Web of Science (WOS) (Thomson Reuters 2010) and Scopus (Elsevier 2002)— both opted to use a similar model of classification that relies upon one hierarchical scheme encompassing a number of areas (first level) and subject categories (second level). All the source journals collected by these databases are placed in one or more area and category on the basis of criteria such as title, scope or citation patterns. Thus, in contrast to databases with a more specialized coverage, such as Medline or INSPEC, where papers are directly assigned to categories, under the WOS or Scopus classification model, journals are classified into categories, while the papers covered by them are assigned to source categories through indirect assignation.

Both WOS and Scopus have become key tools for the development of bibliometric surveys which, in the face of science evaluation, aid decision-making on the part of scientific administrators and politicians concerned with funding and the efficient assignment of resources. As far back as 1963, Weinberg (Weinberg 1963) claimed that the extensive growth of science required more resources in a society of limited resources, meaning it was necessary to choose among different areas or fields of science (*scientific choice*) and between the various institutions receiving government assistance (*institutional choice*). He therefore put forth a number of useful criteria for prioritizing when selecting, divided into *internally generated within the scientific field* itself, and *externally generated out of the field*, which included aspects such as technological, scientific and social merit. It is clear that bibliometric and scientometric analysis as developed from data covered by major scientific databases must be considered essential instruments in the evaluation process and in selecting the best ones within a system mainly based on merit. However, in order to ensure that surveys have high credibility and precision, it is necessary to define and delimit in a reliable manner each one of the subject fields and subfields of knowledge generated through research. For this reason the design of a flexible and adaptable classification model for categorizing scientific literature is held to be an essential matter.

Review

This work introduces a proposal to improve the categorization of Scopus database journals included at the SCImago Journal and Country Rank (SJR) portal (SCImago Lab 2007) using journal reference analysis, one of the many techniques applied in the vast arena of scientific literature for the classification, categorization and delimitation of subject fields. Narin (Narin 1976) was a pioneer in proposing that papers be classified by allocating them to the category of journals to which they belonged. He held that citation recount was useful not only for bibliometric purposes, but also for the classification of publications. In earlier work (Narin, Carpenter, & Nancy 1972), using references and citation analysis, different graphic models were developed to represent the relations established between a set of journals and the disciplines they pertained to. In papers published with Pinski (Narin, Pinski, & Gee 1976) (Pinski & Narin 1976), he used the analysis of bibliographic references to aggregate journals into different groups and subject categories.

Glänzel employed reference analysis to develop an item-by-item classification model applicable at the level of items (papers) rather than at the source level. Firstly, he analyzed paper reference lists of the Science Citation Index (SCI) *multidisciplinary* and *general journals* (Glänzel, Schubert, & Czerwon 1999). He then applied a methodology similar to that used with papers published in journals covered by the *Social Science Citation Index* (SSCI) (Glänzel and others 1999). Finally, in a further contribution, Glänzel (Glänzel & Schubert 2003) devised a new classification scheme applicable to all areas of scientific knowledge (science, social sciences, and arts & humanities) with scientometric evaluation purposes. The three-step building process included, at step 3, the classification of papers appearing in journals with ambiguous or poorly defined categories (i.e. multidisciplinary), on the basis of reference analysis.

Searching for a way to upgrade and restructure classification of journals, Leydesdorff (Leydesdorff 2002) developed a proposal to define shifts in the classification schemes of databases due to the inclusion of new journals or modifications (merges or title changes) affecting them. This proposal focused on transactions and relationships among journals involving citation. Its main goal was to define a posteriori the changes in categories represented by different sets of journals, giving rise to a dynamic and evolutionary update of the classification schemes used in databases.

More recently, in order to define and delimit the ISI categories of *Oncology* and *Cardiac & Cardiovascular System*, Lopez-Illescas (López-Illescas and others 2009) put forth an approach combining_the use of WOS specialized journal categories together with reference analysis. Under this approach, it is assumed that scientific journal articles are well categorized within a gjven subfield of the source journal. Therefore, a subfield could be properly delimited by a group of papers from specialized journals in a particular subfield (*subfield's specialist journals*) and another group of papers belonging to non-specialist journals (*additional journals*) that cite journal papers from a previously established citation threshold.

Improving and updating the categorization of Scopus database journals included in the SCImago Journal and Country Rank (SIR) website calls for some reallocation and delimitation of subject areas and categories in order to restructure the scientific knowledge encompassed by SJR journals. It is thus intended to represent a consistent and congruent new disciplinary structure founded upon a set of well defined subject categories. Once the new classification scheme is defined, it is necessary to re-categorize journals, assigning the subject categories considered under the new scheme. This process largely entails reference analysis. In the case of journals with an insufficient number of bibliographic references, e.g. social science or arts & humanities journals, it will be necessary to tackle other methodological procedures in order to categorize them.

The final goal of our proposal is therefore to redefine the subject areas and categories of SJR journals through reference analysis. Narin (Narin 1976) established the importance of citations among papers to define the structure of scientific literature. At a macro level, he found citation analysis useful for representing and relating areas and subject categories by mapping journals. By further exploring this idea, we intend to more soundly define categories or disciplines representing the knowledge covered by the scientific literature of SJR journals.

Methodology

The SJR two-level hierarchical classification scheme, consisting of 27 areas and 308 subject categories, was used in this study. It had been previously defined by SCImago group members on an empirical basis, taking into account characteristic and discriminative journal features such as title or scope, and expert opinions. The starting point was the editors' journal categorization based on their scope statements. Many authors hold a priori classification schemes developed by experts (relying on their scholarship, knowledge and experience in specific fields) to be useful not only for information retrieval, but also for bibliometric and scientometric purposes. Glänzel (Glänzel & Schubert 2003) judged this proposal as sensible and pragmatic. Earlier on, Schubert (Schubert & Braun 1996) affirmed that in reference standardization processes, essential for the development of scientometric indicators and their subsequent comparison, "comparative assessments based on prior classification schemes are usually easier to comprehend and accept".

We then submitted a query to retrieve Scopus data from SJR in order to derive a neighbor list containing citing-journals, cited-journals and values representing the relationships established among them. The data set covered a 6-year period, from 2003 to 2008, with references going back as far as 1996 (to 2008), and including a total of 17158 journals. For this process, journal self-citation values were discarded. By using this list, an asymmetric journal-category citation matrix was constructed whose values display the amount of citing-journal references linking to SJR categories, reached via aggregation of the cited-journal categories. Therefore, improved final categorization of journals was achieved on the basis of SJR categorization (previously assigned) of cited-journals. The relationship values established among journals and categories were later transformed into percentages. Finally, categories labeled as Miscellaneous and Multidisciplinary were removed from the analysis.

Observation of journal-category vectors derived from the journal-category matrix, revealed the existence of a large amount of residual values in each one. These values reflect weaker relations established between the journals and certain categories (Figure 1). We assumed that, for each journal vector, the most representative categories were reflected by higher percentage values. So as to avoid the weakest links representing the categories with less influence on journal topic, a threshold was established. This allowed us to transform original vectors by keeping only values or aggregate values (cumulative sum) equal to or higher than the threshold defined, while values below it were isolated (Figure 2). Thus, the method works by stressing the generality of journals in order to define their definitive categories.

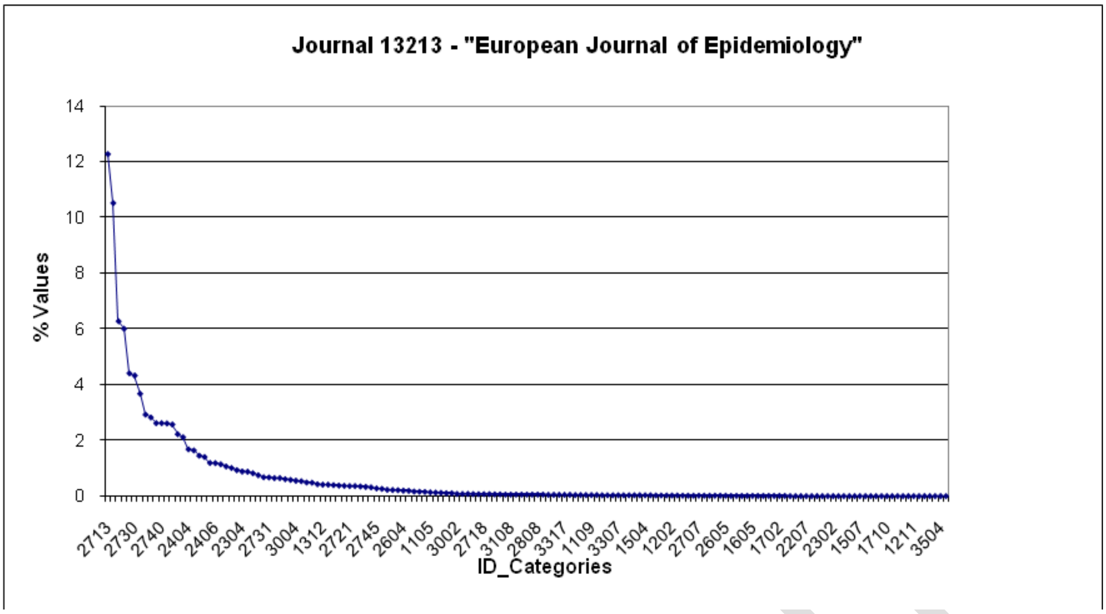

Figure 1: Distribution of Categories of Journal 13213 after Iteration 1

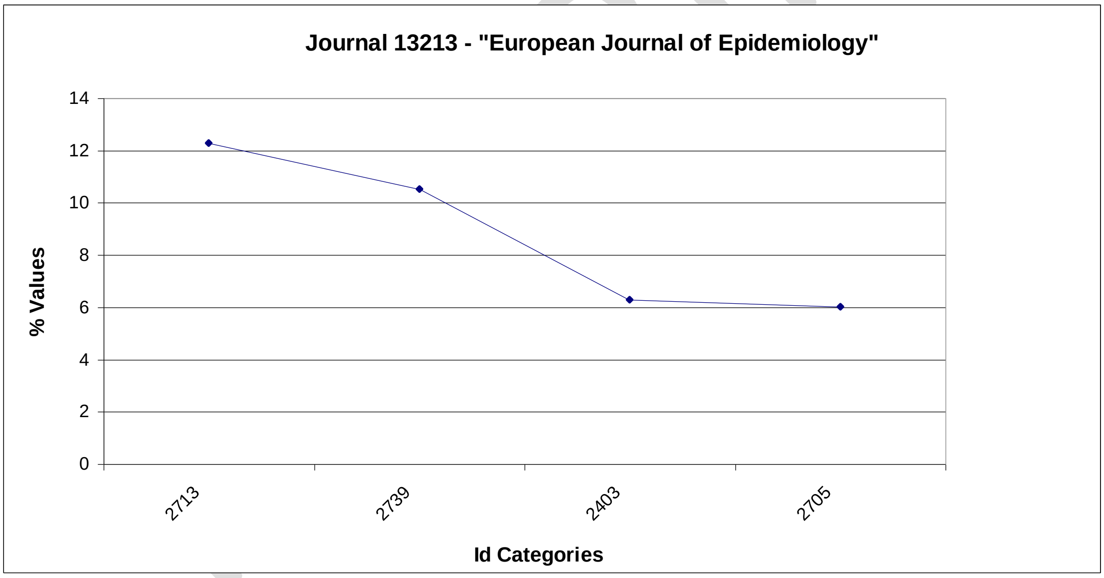

Figure 2: Distribution of Categories of Journal 13213 after Iteration 1 and Threshold 31%

As the next step, we ran an iterative categorization process during *n times*, so that classification of the journals constituting the citation network would be enhanced and gain in pertinence. That is, as the process advanced, relations among journals and categories would be seen to change; while journal-journal links remained identical, cited journal categories were altered through iterations by means of a feedback process. It was therefore necessary to clarify some important questions for the method's performance, such as: 1) *number of iterations* to execute; 2) *cutoff* or *stopping point* in order to reach a well-delimited and consistent subject categories scheme; and 3) *the threshold value* to apply to vectors.

On the one hand, through an heuristic approach based on observation of changes produced at distribution of categories per iteration —i.e., the number of categories keeping cited by journals after each replication (Figure 3)— we concluded that a total of 12 iterations was enough to obtain an overall view of process performance. On the other hand, it was noted that cutoff depends largely upon the threshold value established. To select the more appropriate threshold, we resorted to several empirical tests using values from 25% to 60%. These tests evidenced that best threshold was 31%. Thereby, once vectors were optimized, the matrix was reduced to approximately a 1/3 share of their values, retaining only the strongest relations between journals and categories. Then, by analyzing certain indicators achieved after adopting 31% threshold, iteration 2 was determined to be the best point to halt the process. At that point, changes resulting in journal categorization were the most balanced according to relevant indicators such as *Mean categories per journal* or *Number of cited categories per iteration*. Table 1, collecting these indicators, makes manifest that the steepest drop in the distribution of categories per iteration was between 1 and 3. Not only did this reinforced iteration 2 serve as an excellent cutoff, but also, journals with a high number of assigned categories could thereby be avoided. Distribution of journals with *n categories* can be seen in Figure 4.

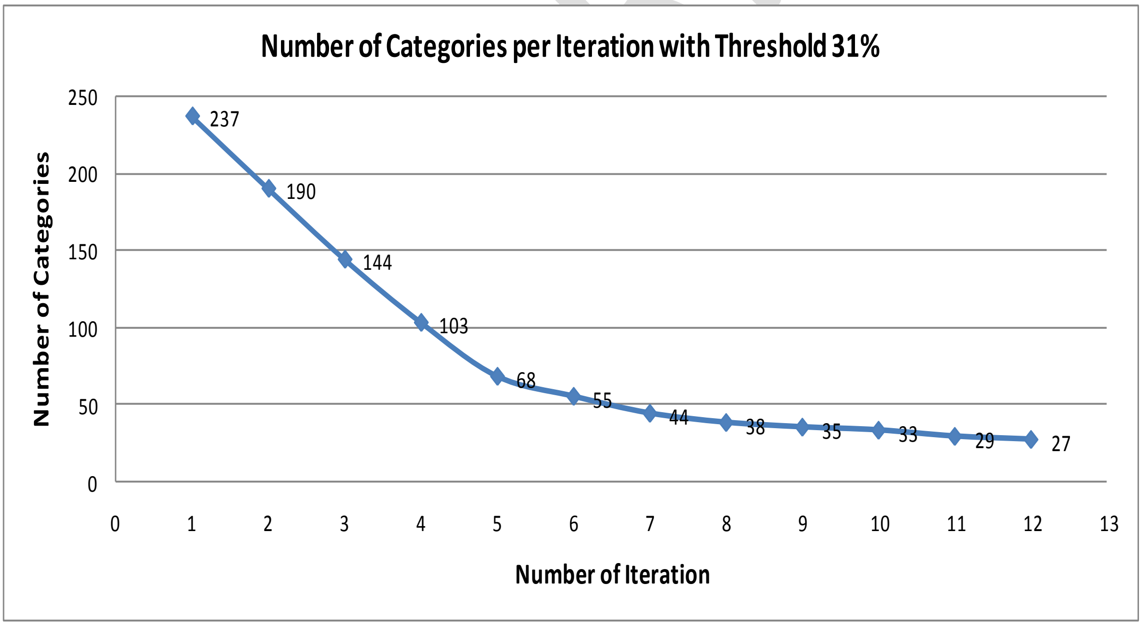

Figure 3: Distribution of Categories over Iterations using a Threshold of 31%

	Journals Categorized	Journals Not Categorized	Number of Records per Table	Mean of Categories per Journal	Num. of Cited Categories	Slope	Slope Percentage	Num. of Non-cited Categories
iteration 1	15584	1574	34046	2.19	237			66
iteration 2	**15595**	**1563**	**32317**	**2.07**	**190**	**47**	**22.4**	**113**
iteration 3	15595	1563	25606	1.64	144	46	21.9	159
iteration 4	15595	1563	20277	1.30	103	41	19.5	200
iteration 5	15595	1563	17514	1.12	68	35	16.7	235
iteration 6	15595	1563	16560	1.06	55	13	6.2	248
iteration 7	15595	1563	16209	1.04	44	11	5.2	259
iteration 8	15595	1563	16089	1.03	38	6	2.9	265
iteration 9	15595	1563	15982	1.03	35	3	1.4	268
iteration 10	15595	1563	15872	1.02	33	2	1.0	270
iteration 11	15595	1563	15815	1.01	29	4	1.9	274
iteration 12	15595	1563	15769	1.01	27	2	1.0	276

Table 1: Indicators obtained for a Threshold of 31%

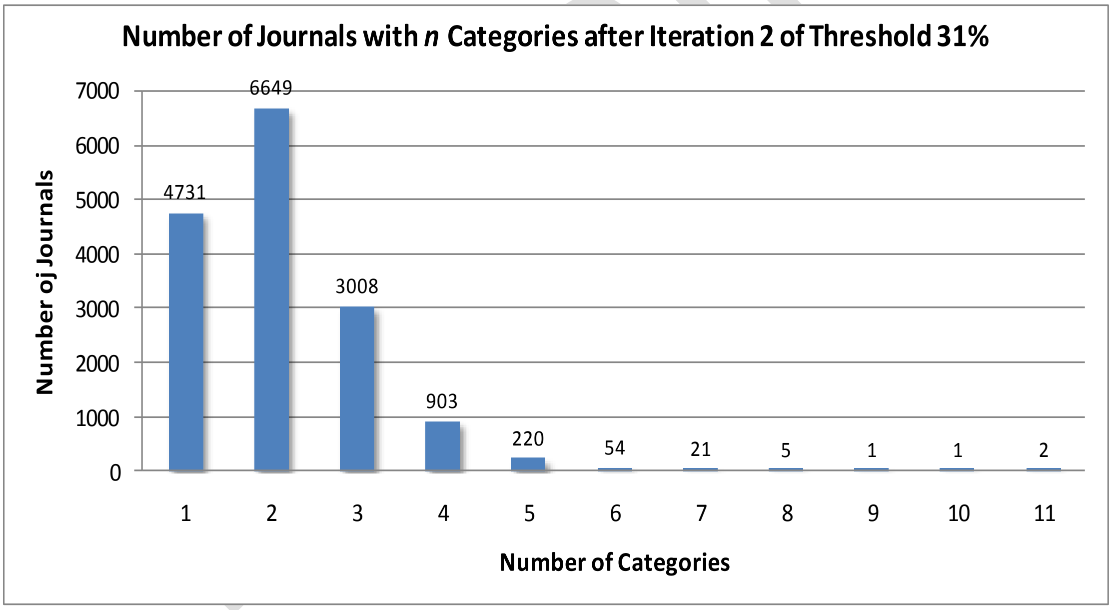

Figure 4: Distribution of Journals with *n* Categories after Iteration 2 and Threshold 31%

Having resolved the threshold, cutoff and number iterations to be applied, we considered that, in order to establish a stronger journal categorization and a more robust scheme, it was necessary that the categorized journals would satisfy the condition of having at least 30 items and 30 references pointing to database sources. Of course, some journals did not reach these values, and so the number of journals categorized decreased around 1500; but at the same time this ensured a good ratio of journals categorized and a well delimited set of subject categories.

The last stage of the method involved the application of a manual review process of journals with more than 4 assigned categories so as to obtain a finer categorization and to prevent an extensive number of multi-categorized journals. This was done following a set of well defined rules, making it possible to readjust the final categorization of around 300 journals and allowing us to discard categories with low rates, to determine a maximum of 5 categories per journal, and to aggregate categories into homogenous sets of the same subject area (*miscellaneous categories*), or into heterogeneous ones of different areas (*multidisciplinary category*). By doing so, 8 new categories (*7 miscellaneous* and *1 multidisciplinary*) were obtained and added to the final set of categories forming part of the renewed classification scheme. Additionally, the multidisciplinary category was included in one new *multidisciplinary area*.

Results and Discussion

Table 2 collects some indicators related to different moments in the development and evolution of this categorization process, from SJR initial categorization to the final classification scheme obtained after applying thresholds, iterations and reviews. Comparison of the original SJR classification scheme and new one revealed various perceptible changes implying new aggregations of journals into different categories than before, and, most importantly, the disappearance of some categories that are not used by journals —in other words, categories that are not being linked by the current journal references. To a lesser degree, this affects areas as well, whose removal is fully tied to a total disuse of categories included into them. The total number of changes in journal categorization was over 12000, involving the addition, loss or new ranking (based on percentages) of categories assigned to journals.

	SJR	**Threshold 31% Iteration 1**	**Threshold 31% Iteration 2**	**Threshold 31% Iteration 2 Papers & Refs. to DB Sources >= 30**	**Threshold 31% Iteration 2 Review of Journals >4 Categories Assigned**
Categorized Journals	17158	15584	15595	14166	14166
Number of Areas	27	25	23	23	24
Number of Categories	308	237	190	186	198
Mean of Categories per Journal	1.54	2.19	2.07	2.11	2.06

Table 2: Differences among SJR original categorization and new one

The final categorization scheme (giving rise to a total of 14166 categorized journals) including subject areas, subject categories covered by them, journals citing each category, and their corresponding percentages over total journals categorized, can be found in the *Appendix* section of this paper. To calculate percentages, the overlap due to multi-categorization of journals was studied. At a higher level, Table 3 captures the final distribution of journals per area under the new SJR categorization scheme, collecting the number of journals covered by areas, percentages of this ratio, and the number of categories included in every given area together with their respective percentages.

AREA	Journals Covered per Areas	Percentage of Journals	Categories per Area	Percentage of Categories
Multidisciplinary	28	0.10	1	0.51
Agricultural and Biological Sciences	1543	5.28	11	5.56
Arts and Humanities	586	2.01	8	4.04
Biochemistry, Genetics and Molecular Biology	5213	17.85	13	6.57
Business, Management and Accounting	792	2.71	9	4.55
Chemical Engineering	232	0.79	6	3.03
Chemistry	938	3.21	7	3.54
Computer Science	629	2.15	12	6.06
Decision Sciences	110	0.38	1	0.51
Earth and Planetary Sciences	1056	3.62	11	5.56
Economics, Econometrics and Finance	840	2.88	2	1.01
Energy	172	0.59	5	2.53
Engineering	1968	6.74	13	6.57
Environmental Science	1169	4.00	8	4.04
Immunology and Microbirology	2023	6.93	5	2.53
Materials Science	619	2.12	8	4.04
Mathematics	739	2.53	9	4.55
Medicine	6931	23.73	35	17.68
Neuroscience	107	0.37	2	1.01
Pharmacology, Toxicology and Pharmaceutics	278	0.95	3	1.52
Physics and Astronomy	847	2.90	8	4.04
Psychology	278	0.95	5	2.53
Social Sciences	2037	6.97	14	7.07
Health Professions	72	0.25	2	1.01
Total	**29207**	**100**	**198**	**100**

Table 3: Final Categorization Scheme at Area Level

A simple glance at Table 3 suffices to discover a group of dense areas covering a high number of journals. In more specific terms, this means that a set of 21,940 journals, that is, approximately 75% of the total taking into account the overlap, cite categories covered by just 8 of the 24 areas constituting the final scheme. *Medicine* (23.73%) and *Biochemistry, Genetics and Molecular Biology* (17.85%) stand out quite clearly. As a general rule, the denser the area appears, the more categories it includes, although there are some exceptions, for instance, in *Computer Science*; *Immunology and Microbiology*; or *Economics, Econometrics and Finance*.

At the level of categories (see Appendix), we encountered one small group of very populous categories covering thousands of journals, a medium-size group of categories including hundreds of journals, and a great one formed by categories embracing fewer than 100 journals. To explore this finding, a ranking of categories based on the number of journals citing each category was constructed. These values were then transformed into percentages, and the cumulative percentages for this distribution were finally added as well. Similar to what happened with the areas, these findings (partially given in Table 4) evidenced a large aggregation of journals in several categories of the new classification scheme. It was moreover seen that only 15 of the 198 categories of the new classification scheme proved sufficient to categorize nearly 50% of the 14,416 journals conforming the final set (again, there was overlap in calculating percentages). Both the aggregations and the decreasing number of areas and categories most likely occurred because the method implies a flow of journals moving from certain categories to others as iterations proceed, as well as the final isolation of many categories.

Rank	CATEGORIES	Journals Citing Category	Percentage	Cumulative Percentage
1	Immunology	1740	5.96	5.96
2	Cell Biology	1688	5.78	11.74
3	Biochemistry	1483	5.08	16.81
4	Psychiatry and Mental Health	1380	4.72	21.54
5	Cardiology and Cardiovascular Medicine	1221	4.18	25.72
6	Public Health, Environmental and Occupational Health	1106	3.79	29.51
7	Sociology and Political Science	944	3.23	32.74
8	Economics and Econometrics	838	2.87	35.61
9	Electrical and Electronic Engineering	820	2.81	38.42
10	Oncology	683	2.34	40.75
11	Condensed Matter Physics	631	2.16	42.91
12	Ecology	616	2.11	45.02
13	Physical and Theoretical Chemistry	552	1.89	46.91
14	Cancer Research	536	1.84	48.75
15	Physiology	505	1.73	50.48

Table 4: Top 15 Ranked Journals per Category with New Categorization

A number of factors play some role in this phenomenon. Firstly, an implicit feature of the approach is that it keeps the most outstanding categories and discards the less representative ones per each journal. Thus, the method focuses on generality rather than specificity in its attempt to delineate and define a journal subject. Figures 1 and 2 (above) serve to illustrate this aspect of performance.

A second reason is the drawing power of certain categories, particularly from the area of pure sciences. The data provided in Table 4 reveal that, on the whole, categories ranked in foremost positions (*Immunology*; *Cell Biology*; *Biochemistry*; *Psyichiatry and Mental Health*; *Cardiology and Cardiovascular Medicine*, etc.) are encompassed in pure science areas such as *Immunology and Microbiology*; *Biochemistry, Genetics and Molecular Biology*; or *Medicine*. Only the categories of *Sociology and Political Science*;

and *Economics and Econometrics,* more connected to the area of the social sciences area, are an exception within the top 15 categories list. Therefore, since the method developed is based on journal reference analysis, we infer the existence of a substantial share of journals citing database pure science sources, despite the subject area or category where they are actually included. The disuse of different categories is a common issue for those categories with a low rate of journals citing them. Thus, categorized journals are finally attracted to more powerful categories as a consequence of the Matthew Effect. This phenomenon becomes more acute as more iterations are run.

Nevertheless, another possibility concerns the relatively new disciplines with a non-cohesive background. Normally these disciplines cite intellectual bases (Chen 2006) pertaining to other fields with very close boundaries, or which can find a "fertile ground in a neighboring field" (Small 1999) evoking inter-disciplinarily symptoms. Some examples of this are the categories *Gender Studies*; *Human Factors and Ergonomy*; *Nature and Landscape Conservation*, or a few from the area of *Nursing*.

Of course, all these explanations can be extrapolated to the application of the different thresholds used in the development and design of our method. One additional disadvantage is that, the higher the threshold, the higher the ratio of multi-categorized journals proved to be; and conversely, the lower the threshold, the lower the number of categories falling into the final categorization scheme.

Conclusions

The proposal featured permitted us to categorize 14,416 Scopus journals from an initial set of 17,158 as well as to restructure and redefine the SJR classification scheme at two levels of aggregation. Admittedly, while the method provided a consistent SJR classification scheme, we are mindful that it can not be considered as a definitive classification solution, since it does not provide a comprehensive and definitive placement of the journals assessed. For the time being, this approach should be supplemented with additional techniques, based either on citation or on text, in order to classify the whole set of covered journals.

A good performance of the method is closely linked to a good set-up of the main parameters, namely, total number of iterations to use, threshold to apply and cutoff fixed. Heuristic processes and empirical tests were determining factors for configuring it. The designation of 12 iterations was enough to make manifest that more iterations meant bigger aggregations of journals into a small set of categories. This fact may be useful in the case that one keeps running iterations until grouping journals into vast, basic areas of scientific knowledge. Regarding thresholds and cutoff, we noted they were very closely related. From the whole set of tests executed, the most balanced mix of these parameters, in terms of *number of categories cited by journals*, *mean of categories per journal*, and *number of multi-categorized journals*, took place at iteration 2 of the threshold 31%. Of course, the results of this combination were not the same in every test.

The method inevitably entails missed categories due to a large aggregation of journals into a reduced number of categories. Thus, an ever-increasing share of journals is seen to use an ever-decreasing share of categories. In other words, a small set of categories would suffice to categorize a vast set of journals, and we believe the method could offer better results by categorizing journals to a high level of aggregation, such as subject area. The category aggregation problem could be minimized by modifying the method, for instance using only the first iteration and discarding the remaining ones.

The citation flows between categories evidenced a clear attraction exerted by sources covering pure science. This happened among categories of different subject areas and also among categories of the same area. Some causes behind this might be related to database coverage, citation behavior, or the degree of consolidation of each particular discipline.

It is also interesting to highlight another positive aspect of the method, concerning the decreasing number of journals categorized under the Multidisciplinary category. Journals assigned to this category shifted to narrower categories later, mostly to Miscellaneous, which covers different categories inside the same subject area. In our study, the number of Multidisciplinary journals went from the 65 journals of the SJR original categorization to 28 journals under the new scheme.

Before closing, we underline that upcoming studies should provide a good framework to implement alternative techniques and to improve our method, so that a complete assignation of categories for each journal gathered and analyzed could be carried out. Forthcoming research efforts will thus be directed toward cluster analysis, examining the citation dimension through coupling and cross-citation. Later on, other possibilities may be explored, such as text dimension, using keywords or text parts extracted from journal articles.

Acknowledgment

The authors thank Jean Sanders for editing the text.

Bibliography

Chen, C. M. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology,* 57 (3), 359-377.

Chen, C. M. (2008). Classification of scientific networks using aggregated journal-journal citation relations in the Journal Citation Reports. *Journal of the American Society for Information Science and Technology,* 59 (14), 2296-2304.

Elsevier (2002). Scopus. http://www.scopus.com/home.url. Accessed 20-6-2011

Glänzel, W. and Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics,* 56 (3), 357-367.

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 (3), 427-439.

Glänzel, W., Schubert, A., Schoepflin, U., and Czerwon, H. J. (1999). An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis. *Scientometrics,* 46 (3), 431-441.

Gómez, I., Bordons, M., Fernández, M. T., and Méndez, A. (1996). Coping with the problem of subject classification diversity. *Scientometrics,* 35 (2), 223-235.

Leydesdorff, L. (2002). Dynamic and evolutionary updates of classificatory schemes in scientific journal structures. *Journal of the American Society for Information Science and Technology,* 53 (12), 987-994.

López-Illescas, C., Noyons, E. C. M., Visser, M. V., Moya-Anegón, F., and Moed, H. F. (2009). Expansion of scientific journal categories using reference analysis: How can it be done and does it make a difference? *Scientometrics,* 79 (3), 473-490.

Narin, F. *Evaluative bibliometrics : the use of publication and citation analysis in the evaluation of scientific activity*, Cherry Hill, N.J.: Computer Horizons, Inc., 1976.

Narin, F., Carpenter, M., and Nancy, C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science,* 23 (5), 323-331.

Narin, F., Pinski, G., and Gee, H. H. (1976). Structure of the Biomedical Literature. *Journal of the American Society for Information Science,* 27 (1), 25-45.

Pinski, G. and Narin, F. (1976). Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics. *Information Processing and Management,* 12 (5), 297-312.

Schubert, A. and Braun, T. (1996). Cross-field normalization of scientometric indicators. *Scientometrics,* 36 (3), 311-324.

Scimago Lab (2007). Scimago Journal & Country Rank. http://www.scimagojr.com/. Accessed 20-6-2011

Small, H. (1999). A passage through science: crossing disciplinary boundaries. *Library Trends,* 48 (1), 72-108.

Thomson Reuters (2010). ISI Web of Science. Thomson Reuters. http://apps.isiknowledge.com/WOS_GeneralSearch_input.do?preferencesSaved=&product=WOS&SID=Z1CB8C2BE1G8nk3f9gM&search_mode=GeneralSearch. Accessed 20-6-2011

Weinberg, A. (1963). Criteria for scientific choice. *Minerva,* 1 (2), 159-171.

Appendix

AREAS	CATEGORIES	Journal Citing Category	Percentage
Multidisciplinary	Multidisciplinary	28	0.096
Agricultural and Bioogical Sciences	Agricultural and Biological Sciences (miscellaneous)	2	0.007
	Agronomy and Crop Science	82	0.281
	Animal Science and Zoology	247	0.846
	Aquatic Science	272	0.931
	Ecology, Evolution, Behavior and Systematics	67	0.229
	Food Science	255	0.873
	Forestry	45	0.154
	Horticulture	1	0.003
	Insect Science	63	0.216
	Plant Science	427	1.462
	Soil Science	82	0.281
Arts and Humanities	History	250	0.856
	Language and Linguistics	159	0.544
	Classics	1	0.003
	Literature and Literary Theory	45	0.154
	Music	16	0.055
	Philosophy	87	0.298
	Religious Studies	19	0.065
	Visual Arts and Performing Arts	9	0.031
Biochemistry, Genetics and Molecular Biology	Biochemistry, Genetics and Molecular Biology (miscellaneous)	34	0.116
	Aging	15	0.051
	Biochemistry	1483	5.078
	Biophysics	3	0.010
	Biotechnology	69	0.236
	Cancer Research	536	1.835
	Cell Biology	1688	5.779
	Developmental Biology	13	0.045
	Endocrinology	292	1.000
	Genetics	431	1.476
	Molecular Biology	142	0.486
	Physiology	505	1.729
	Structural Biology	2	0.007
Business, Management and	Business, Management and Accounting (miscellaneous)	3	0.010
	Accounting	10	0.034

Accounting	Business and International Management	82	0.281
	Management Information Systems	9	0.031
	Management of Technology and Innovation	383	1.311
	Marketing	65	0.223
	Organizational Behavior and Human Resource Management	5	0.017
	Strategy and Management	227	0.777
	Tourism, Leisure and Hospitality Management	8	0.027
Chemical Engineering	Catalysis	61	0.209
	Chemical Health and Safety	1	0.003
	Colloid and Surface Chemistry	14	0.048
	Filtration and Separation	4	0.014
	Fluid Flow and Transfer Processes	105	0.360
	Process Chemistry and Technology	47	0.161
Chemistry	Chemistry (miscellaneous)	3	0.010
	Analytical Chemistry	121	0.414
	Electrochemistry	34	0.116
	Inorganic Chemistry	21	0.072
	Organic Chemistry	195	0.668
	Physical and Theoretical Chemistry	552	1.890
	Spectroscopy	12	0.041
Computer Science	Computer Science (miscellaneous)	7	0.024
	Artificial Intelligence	114	0.390
	Computational Theory and Mathematics	77	0.264
	Computer Graphics and Computer-Aided Design	39	0.134
	Computer Networks and Communications	3	0.010
	Computer Science Applications	8	0.027
	Computer Vision and Pattern Recognition	22	0.075
	Hardware and Architecture	43	0.147
	Human-Computer Interaction	7	0.024
	Information Systems	88	0.301
	Signal Processing	5	0.017
	Software	216	0.740
Decision Sciences	Management Science and Operations Research	110	0.377
Earth and Planetary Sciences	Earth and Planetary Sciences (miscellaneous)	4	0.014
	Atmospheric Science	191	0.654
	Computers in Earth Sciences	26	0.089
	Earth-Surface Processes	77	0.264
	Geochemistry and Petrology	351	1.202
	Geology	75	0.257
	Geophysics	111	0.380
	Geotechnical Engineering and Engineering Geology	73	0.250
	Oceanography	44	0.151
	Paleontology	54	0.185

	Space and Planetary Science	50	0.171
Economics, Econometrics and Finance	Economics and Econometrics	838	2.869
	Finance	2	0.007
Energy	Energy (miscellaneous)	1	0.003
	Energy Engineering and Power Technology	85	0.291
	Fuel Technology	17	0.058
	Nuclear Energy and Engineering	23	0.079
	Renewable Energy, Sustainability and the Environment	46	0.158
Engineering	Engineering (miscellaneous)	30	0.103
	Aerospace Engineering	15	0.051
	Biomedical Engineering	43	0.147
	Civil and Structural Engineering	136	0.466
	Computational Mechanics	140	0.479
	Control and Systems Engineering	232	0.794
	Electrical and Electronic Engineering	820	2.808
	Industrial and Manufacturing Engineering	57	0.195
	Mechanical Engineering	445	1.524
	Mechanics of Materials	1	0.003
	Ocean Engineering	17	0.058
	Safety, Risk, Reliability and Quality	13	0.045
	Building and Construction	19	0.065
Environmental Science	Environmental Science (miscellaneous)	6	0.021
	Ecology	616	2.109
	Environmental Chemistry	343	1.174
	Environmental Engineering	85	0.291
	Health, Toxicology and Mutagenesis	13	0.045
	Management, Monitoring, Policy and Law	8	0.027
	Waste Management and Disposal	1	0.003
	Water Science and Technology	97	0.332
Immunology and Microbiology	Applied Microbiology and Biotechnology	2	0.007
	Immunology	1740	5.957
	Microbiology	240	0.822
	Parasitology	40	0.137
	Virology	1	0.003
Materials Science	Materials Science (miscellaneous)	1	0.003
	Biomaterials	22	0.075
	Ceramics and Composites	48	0.164
	Electronic, Optical and Magnetic Materials	131	0.449
	Materials Chemistry	139	0.476
	Metals and Alloys	128	0.438
	Polymers and Plastics	130	0.445
	Surfaces, Coatings and Films	20	0.068
Mathematics	Algebra and Number Theory	134	0.459

	Analysis	15	0.051
	Applied Mathematics	370	1.267
	Computational Mathematics	20	0.068
	Discrete Mathematics and Combinatorics	18	0.062
	Logic	11	0.038
	Mathematical Physics	21	0.072
	Statistics and Probability	101	0.346
	Theoretical Computer Science	49	0.168
Medicine	Medicine (miscellaneous)	35	0.120
	Anesthesiology and Pain Medicine	76	0.260
	Cardiology and Cardiovascular Medicine	1221	4.181
	Critical Care and Intensive Care Medicine	8	0.027
	Complementary and Alternative Medicine	2	0.007
	Dermatology	85	0.291
	Emergency Medicine	14	0.048
	Endocrinology, Diabetes and Metabolism	67	0.229
	Epidemiology	30	0.103
	Gastroenterology	117	0.401
	Genetics (clinical)	3	0.010
	Geriatrics and Gerontology	48	0.164
	Health Informatics	4	0.014
	Health Policy	33	0.113
	Hematology	23	0.079
	Microbiology (medical)	1	0.003
	Nephrology	47	0.161
	Neurology (clinical)	406	1.390
	Obstetrics and Gynecology	137	0.469
	Oncology	683	2.338
	Ophthalmology	86	0.294
	Orthopedics and Sports Medicine	265	0.907
	Otorhinolaryngology	118	0.404
	Pathology and Forensic Medicine	102	0.349
	Pediatrics, Perinatology and Child Health	152	0.520
	Pharmacology (medical)	7	0.024
	Psychiatry and Mental Health	1380	4.725
	Public Health, Environmental and Occupational Health	1106	3.787
	Pulmonary and Respiratory Medicine	97	0.332
	Radiology, Nuclear Medicine and Imaging	164	0.562
	Rehabilitation	40	0.137
	Rheumatology	37	0.127
	Surgery	282	0.966
	Transplantation	1	0.003
	Urology	54	0.185

Neuroscience	Behavioral Neuroscience	10	0.034
	Cognitive Neuroscience	97	0.332
Pharmacology, Toxicology and Pharmaceutics	Pharmaceutical Science	38	0.130
	Pharmacology	180	0.616
	Toxicology	60	0.205
Physics and Astronomy	Acoustics and Ultrasonics	35	0.120
	Astronomy and Astrophysics	1	0.003
	Condensed Matter Physics	631	2.160
	Instrumentation	2	0.007
	Nuclear and High Energy Physics	80	0.274
	Atomic and Molecular Physics, and Optics	76	0.260
	Radiation	7	0.024
	Statistical and Nonlinear Physics	15	0.051
Psychology	Applied Psychology	4	0.014
	Clinical Psychology	4	0.014
	Developmental and Educational Psychology	125	0.428
	Experimental and Cognitive Psychology	143	0.490
	Social Psychology	2	0.007
Social Sciences	Archeology	67	0.229
	Development	2	0.007
	Education	328	1.123
	Geography, Planning and Development	318	1.089
	Health (social science)	8	0.027
	Law	105	0.360
	Library and Information Sciences	105	0.360
	Sociology and Political Science	944	3.232
	Transportation	37	0.127
	Anthropology	64	0.219
	Communication	39	0.134
	Cultural Studies	5	0.017
	Demography	5	0.017
	Urban Studies	10	0.034
Health Professions	Radiological and Ultrasound Technology	70	0.240
	Speech and Hearing	2	0.007