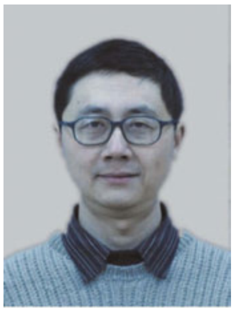
Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories

Hui Fang

State Key Laboratory of Analytical Chemistry for Life Science, School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China, <fanghui@nju.edu.cn>



Hui Fang received BSc in radio engineering (in 1990) and MSc degrees in signal processing (in 1993) from Southeast University in Nanjing, China, and the PhD degree in Electroanalytical Chemistry from Nanjing University in Nanjing, China, in 1998. He is now an associate professor at the School of Electronic Science and Engineering, Nanjing University. His research interests include information processing, data mining, artificial intelligence, instruments and instrumentation, and bibliometrics.

Fang, Hui. **Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories.** *Knowledge Organization. 42(3)*, 139-153. 35 references.

Abstract: In the Thomson Reuters Web of Science database, the subject categories of a journal are applied to all articles in the journal. However, many articles in multidisciplinary Sciences journals may only be represented by a small number of subject categories. To provide more accurate information on the research areas of articles in such journals, we can classify articles in these journals into subject categories as defined by Web of Science based on their references. For an article in a multidisciplinary sciences journal, the method counts the subject categories in all of the article's references indexed by Web of Science, and uses the most numerous subject categories of the references to determine the most appropriate classification of the article. We used articles in an issue of *Proceedings of the National Academy of Sciences* (*PNAS*) to validate the correctness of the method by comparing the obtained results with the categories of the articles as defined by *PNAS* and their content. This study shows that the method provides more precise search results for the subject category of interest in bibliometric investigations through recognition of articles in multidisciplinary sciences journals whose work relates to a particular subject category.

Received: 24 February 2015; Revised 23 April 2015; Accepted 1 May 2015

Keywords: subject categories, multidisciplinary sciences journals, classifying, *PNAS*, WoS

1.0 Introduction

Web of Science (WoS) from Thomson Reuters is a convenient platform for researchers to search for and retrieve articles related to their fields of interest. In addition to the article title, abstract and keywords, it provides other information about the articles it has indexed, such as the authors and their affiliations, publication name, DOI number, year published, financial support, research area, and WoS category. Based on these parameters, bibliometric researchers have investigated research activities (for example, Ardanuy et al. 2009; de la Moneda Corrochano et al. 2013; Diem and Wolter 2013; Moppett and Hardman 2011; Pinto et al. 2013; Wang et al. 2013), analyzed the research trends in a specific field (for example, Chiu and Ho 2007; Grossi et al. 2003; Krampen et al. 2011), examined knowledge diffusion (Chen et al. 2009), and studied paradigm shifts (Marx and Bornmann 2010).

In WoS, a research article is represented by the subject categories of the journal in which it is published. An article in a peer-reviewed journal is classified into the subject categories of the journal by several experts, including its author(s), who select the journal for submission, and reviewers and the editorial board of the journal, who accept the article for publication. WoS labels each journal with a maximum of six subject categories. Those journals that publish articles in more than six subject categories are labeled and categorized as multidisciplinary sciences journals; these include *Science*, *Nature*, and *Proceedings of the National Academy of Sciences* (*PNAS*), for example. The work behind each article in these journals usually relates to a limited number of subject categories, and does not encompass many disciplines. Thus, labelling articles in these journals as multidisciplinary sciences is not accurate.

Bibliometric studies often investigate the distribution of a certain topic in each related subject category to determine that category's contribution to the topic (for example, Aleixandre et al. 2013; Chen et al. 2014; Chiu and Ho 2007; Liu et al. 2012; Naqvi 2014; Porter and Youtie 2009; Yang et al. 2013; Wang et al. 2013; Zhuang et al. 2013). Those subject categories which contain many articles published in multidisciplinary sciences journals will be underrepresented in such searches. For research areas that attract considerable attention among scientists all over the world, such as "Immunology," many important articles are published in multidisciplinary sciences journals because of the high influence of such journals.

This problem related to multidisciplinary sciences journals can be overcome by classifying individual articles published in such journals into subject categories. Glänzel et al. (1999) used reference analysis to classify the subjects of papers published in multidisciplinary and general journals, and this method was also utilized to improve SCImago Journal & Country Rank subject classification (Gómez-Núñez et al. 2011). These previous studies have demonstrated the feasibility of the method. It remains a task to test the correctness of the method. That was the aim of present investigation, using a case study of articles in an issue of *PNAS*, which provides categories of the articles on its website.

The remainder of this article is organized as follows. Section 2 provides the background of this study. Section 3 presents the data used in this study and a detailed description of the methodology for classifying articles in multidisciplinary sciences journals into subject categories according to their references. Section 4 presents and discusses the results of this case study of the method under investigation. Conclusions and limitations are presented in the final sections.

2.0 Background

Classifying articles in multidisciplinary sciences journals can also be achieved by clustering articles (this procedure is not limited to articles in such journals) based on citation relations (Griffith et al. 1974; Small and Griffith 1974; Small and Sweeney 1985; Small et al. 1985; Small 1998; Lewison 1999; Gouvea Meireles et al. 2014). Using this method, Klavans and Boyack (2010) assigned more than 5.5 million publications to over 84,000 research areas. Waltman and Van Eck (2012) recently clustered about 10 million articles from the period 2001–2010. The clustering method can create a classification system other than the subject categories of WoS. Clustering articles into subject categories may also adopt combinations of co-citation and co-word analysis (Braam et al. 1991). Incorporating natural language processing techniques allows co-word analysis to utilize further linguistic relations as the basis for clustering (Ibekwe-SanJuan et al. 2002).

To apply the clustering method for classifying articles into research areas, it is necessary to download information related to a large number of articles from WoS to establish their co-occurrence relations. However, most WoS users can download only up to 500 records at a time, which makes it very difficult to download information on a large number of articles. A more practical method for classifying articles in multidisciplinary sciences journals into subject categories is needed before the results of clustering a large number of articles can be readily available for ordinary users who cannot conveniently download information related to a large number of articles.

Ordinary WoS users can categorize individual articles in multidisciplinary sciences journals indexed by WoS according to the articles' references (Glänzel et al. 1999). This approach is effective in three main ways. First, the content of an article is related to the content of its references; the references of the article introduce the background or area of applicability of the article or the tool or principle adopted by the article. Thus, an article has the same, similar, or related subject categories as its references (Glänzel et al. 1999; Gabel 2006). Second, the intersection of the subject categories of the references can reflect the subject categories of the article in which they are cited. One reference may cover several aspects, which correspond to different subject categories. However, the article in which that reference is cited relates to a subset of those aspects: they belong to a subset of the subject categories to which the research behind the citing article corresponds. For example, the journal *Neurobiology of Aging*, which WoS labels with the subject categories "Geriatrics & Gerontology" and "Neurosciences," publishes the results of studies in which the primary emphasis involves the mechanisms of changes in the nervous system associated with age or age-related diseases. An article in the category "neurosciences" may cite articles (as its references) in *Neurobiology of Aging* because the former refers to the latter's research on the mechanisms of the nervous system; an article in the category "geriatrics & gerontology" may cite articles (as its references) in *Neurobiology of Aging* because the former refers to the latter's research into diseases associated with age. Suppose an article belonging to only one subject category has two references: one reference is labelled with subject categories A, B, and C by the database; the other is labelled with A and D. It can then be inferred that the article belongs to subject category A. The subject categories B, C, and D of the two references may not be related to the subject matter of the citing article. Third, this method utilizes reliable classification of the references into the subject categories of publishing journals by experts (authors of the references, reviewers, and journal editorial boards), which is mentioned in the Introduction.

3.0 Data and methods

3.1 Data

Data were obtained from WoS. Articles in the first issue of *PNAS* in 2014 were used as examples for classification into the subject categories defined by WoS. *PNAS* classifies its articles into three research fields: "physical sciences," "social sciences," and "biological sciences." Each field has several subsidiary groups. For example, the subjects in "physical sciences" include: "applied mathematics," "applied physical sciences," "chemistry," "earth, atmospheric, and planetary sciences," "engineering," "environmental sciences," and "physics." Articles are listed below each subject heading on the *PNAS* website. The described method could be validated by comparing the obtained results with the categories of the articles as defined by *PNAS* (*PNAS* subject categories). As with other multidisciplinary sciences journals that provide subject information about articles, such as *Science* and *PLoS One*, some *PNAS* subject categories have no counterparts in WoS. This means that the subject categories declared by authors may not be in accordance with those of WoS. For an article in such *PNAS* subject categories, we judged the WoS subject categories to which its content relates by reading the article. Except for "*PNAS* subject category," each "subject category" in the sections below represents that defined by WoS. For convenience, we defined the article subject category as the WoS subject category to which the article should be assigned according to its content, and the recognized subject category as that to which the article was assigned by the method.

3.2 Classifying articles in multidisciplinary sciences journals into subject categories

WoS provides information on the references of every article it indexes. The references are also classified into subject categories by WoS. If one reference is labelled by n (from 1 to 6) subject categories, then the reference is equally assigned to the n subject categories by $1/n$ (Waltman 2012) because it is unclear which subject category it belongs to without further information (Bornmann 2014). Suppose an article has L references which are represented by N (L and N are positive integers) subject categories by WoS. Here we use a matrix $\mathbf{S}_{L\times N}$ to represent the assignment of references to each subject category:

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1N} \\ s_{21} & s_{22} & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{L1} & s_{L2} & \cdots & s_{LN} \end{bmatrix}, \qquad (1)$$

where

$$s_{ij} = \begin{cases} 0 & \text{the } i\text{-th reference is not labelled with the } j\text{-th subject category} \\ \frac{1}{n_i} & \text{the } i\text{-th reference is labelled with the } j\text{-th subject category} \end{cases},$$

and n_i is the number of subject categories with which the i-th reference is labelled. s_{ij} can be regarded as the score of the j-th subject category given to the article from its i-th reference. The subject categories of each reference are taken as the subject categories of the journal in which it was published and can be obtained from Journal Citation Reports provided by WoS. The information on the references of the inspected article, such as the journals publishing the references, can be extracted from the full record (including cited references) of the WoS file for users to download.

Vector **M** was defined as representing the scores of each subject category obtained from all the references of an article:

$$\mathbf{M} = (m_1, m_2, \ldots m_N)^{\mathrm{T}}, \qquad (2)$$

where

$$m_j = \sum_{i=1}^{L} s_{ij}, \ (j = 1, 2, \cdots, N).$$

The order of m_j is that of the subject categories in Equation 1. The larger the m_j, the higher is the possibility of the article belonging to the j-th subject category. Suppose,

$$j_{\max} = \arg\max_j m_j \qquad (j = 1, 2, \cdots N) \qquad (3)$$

And

$$m_{\max} = \max_j m_j \qquad (j = 1, 2, \cdots N). \qquad (3')$$

Then, the article can be classified into the $j_{\max}$-th subject category.

Some articles may be related to more than one subject category. If m_{j_1} of any j_1 reaches or exceeds a preset threshold (m_{th}), then the article can also be classified to the j_1-th subject category. The threshold is defined as:

$$m_{th} = m_{max} / d \, , \qquad (4)$$

where d is a parameter to adjust threshold and thus the number of recognized subject categories, defined as the threshold factor.

An alternative method to determine the threshold is to set the threshold at some proportion of the total number of references (López-Illescas et al. 2009). Both methods assign an article to the subject categories in which it has more references than any other subject categories. Evaluating the relative merits of these two methods is beyond the scope of this case study: it demands further investigation.

4.0 Results and discussion

There were 114 articles in the first issue of *PNAS* in 2014. Seven were categorized with two subject headings in the journal. Those articles involved the subjects of physics, chemistry, life sciences and social sciences. Table 1 lists the titles of the 114 articles and their number which was used in Table 2. Table 2 presents the *PNAS* subject categories ($\text{Subject}_{PNAS}1$ and $\text{Subject}_{PNAS}2$) and the recognized subject categories ($\text{SC}_{\text{recog1}}$ to $\text{SC}_{\text{recog6}}$, ranked in descending order of score in Equation 2) for the 114 articles analysed. In the tabulated results, non-italics represent the recognized subject categories when the threshold factor was 3. When the threshold factor was 4, additional results were obtained and are represented by italics. The maximal score m_{max} in Equation 3, which is used to calculate the threshold m_{th} in Equation 4, is selected from **M** (Equation 2), except for the category of "Multidisciplinary Sciences," which does not provide useful information about the subject of the references. If the highest score of an article indicated "Multidisciplinary Sciences," then m_{max} was taken as the second highest category.

The *PNAS* subject categories and the WoS subject categories were not exactly the same. In Table 2, the *PNAS* subject categories "applied biological sciences," "biochemistry," "cell biology," "ecology," "immunology," "microbiology," "neuroscience," "physiology," "plant biology," and "psychological and cognitive sciences" had corresponding or similar WoS subject categories, and the articles related to these subjects were assigned to the article subject categories based on the subject categories of the references. The *PNAS* subject categories "applied physical sciences," "chemistry," "physics," "earth, atmospheric, and planetary sciences," "medical sciences," "social sciences" each encompasses several WoS subject categories. The articles of these *PNAS* subject categories were classified appropriately to the article subject categories by the method.

The recognized subject categories of some articles were not the same as the *PNAS* subject categories. The 15th article analyzed the interaction between plants and parasites. The recognized subject categories were "plant sciences," "zoology," "biochemistry & molecular biology" and "multidisciplinary sciences." The object of the study was bananas, so it was appropriately labelled as "agricultural sciences" by *PNAS*, which is its area of applicability. Most *PNAS* "anthropology" articles were correctly categorized by our method. The exception was the 18th article which discussed the evolution of the human hand. It was reasonable that it was labelled as "anthropology" by *PNAS* or classified as "evolutionary biology" by our method. The *PNAS* subject category "genetics" was similar to the WoS subject category "genetics & heredity." Three of five *PNAS* "genetics" articles were correctly assigned to the WoS counterpart. The other two were the 57th and 110th articles. The former discussed DNA molecules, so it is not surprising that it was assigned to the WoS subject category "biochemistry & molecular biology." The 110th article investigated neuronal translational control and microRNA function. Our method classified it as "neurosciences" which was consistent with the object of the study.

No.	Title of Article
1	High-resolution photoacoustic tomography of resting-state functional connectivity in the mouse brain
2	Theory of epithelial sheet morphology in three dimensions
3	Rationale and mechanism for the low photoinactivation rate of bacteria in plasma
4	Initial stages of calcium uptake and mineral deposition in sea urchin embryos
5	Platinum supported on titanium –ruthenium oxide is a remarkably stable electrocatayst for hydrogen fuel cell vehicles
6	Re-Os geochronology and coupled Os-Sr isotope constraints on the Sturtian snowball Earth
7	Amphitheater-headed canyons formed by megaflooding at Malad Gorge, Idaho
8	Fe-vacancy order and superconductivity in tetragonal β-$Fe_{1-x}Se$
9	Kondo conductance across the smallest spin 1/2 radical molecule
10	Avalanches mediate crystallization in a hard-sphere glass
11	Evidence supporting an intentional Neandertal burial at La Chapelle-aux-Saints
12	Critical slowing down as early warning for the onset and termination of depression
13	Media's role in broadcasting acute stress following the Boston Marathon bombings

14	Dynamic pricing of network goods with boundedly rational consumers
15	Phenalenone-type phytoalexins mediate resistance of banana plants (Musa spp.) to the burrowing nematode Radopholus similis
16	Cultural assemblages show nested structure in humans and chimpanzees but not orangutans
17	Earliest evidence for commensal processes of cat domestication
18	Early Pleistocene third metacarpal from Kenya and the evolution of modern human-like hand morphology
19	Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines
20	Rewiring yeast sugar transporter preference through modifying a conserved protein motif
21	Structure of a eukaryotic thiaminase I
22	Reaction-based fluorescent sensor for investigating mobile Zn^{2+} in mitochondria of healthy versus cancerous prostate cells
23	Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism
24	Regulation of PTEN inhibition by the pleckstrin homology domain of P-REX2 during insulin signaling and glucose homeostasis
25	Biological role of prolyl 3-hydroxylation in type IV collagen
26	Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver
27	Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance
28	Large effect of membrane tension on the fluid–solid phase transitions of two–component phosphatidylcholine vesicles
29	Transmembrane allosteric coupling of the gates in a potassium channel
30	Designed amyloid fibers as materials for selective carbon dioxide capture
31	Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS
32	Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE)
33	Morphological optimization for access to dual oxidants in biofilms
34	Structure of Est3 reveals a bimodal surface with differential roles in telomere replication
35	Measuring membrane protein stability under native conditions
36	Direct observation of a transient ternary complex during IκBα-mediated dissociation of NF-κB from DNA
37	Heteromerization of PIP aquaporins affects their intrinsic permeability
38	Protein structural ensembles are revealed by redefining X-ray electron density noise
39	Nucleolin is important for Epstein-Barr virus nuclear antigen 1-mediated episome binding, maintenance, and transcription
40	Celastrol increases glucocerebrosidase activity in Gaucher disease by modulating molecular chaperones
41	Identification of cancer initiating cells in *K-Ras* driven lung adenocarcinoma
42	Enhanced stability of Mcl1, a prosurvival Bcl2 relative, blunts stress-induced apoptosis, causes male sterility, and promotes tumorigenesis
43	A mechanism for retromer endosomal coat complex assembly with cargo
44	Evaluation of intramitochondrial ATP levels identifies G0/G1 switch gene 2 as a positive regulator of oxidative phosphorylation
45	JMJD5 regulates PKM2 nuclear translocation and reprograms HIF-1α-mediated glucose metabolism
46	Tumor suppressor and deubiquitinase BAP1 promotes DNA double-strand break repair
47	miR-218 opposes a critical RTK-HIF pathway in mesenchymal glioblastoma
48	Emerging predictable features of replicated biological invasion fronts
49	Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion, and infection biology
50	Interannual variation in land-use intensity enhances grassland multidiversity
51	Drastic neofunctionalization associated with evolution of the timezyme AANAT 500 Mya
52	Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts
53	Policing of reproduction by hidden threats in a cooperative mammal
54	A unique covalent bond in basement membrane is a primordial innovation for tissue evolution
55	PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells
56	Scan statistic-based analysis of exome sequencing data identifies *FAN1* at 15q13.3 as a susceptibility gene for schizophrenia and autism
57	Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors
58	Contribution of phenotypic heterogeneity to adaptive antibiotic resistance
59	Ohnologs are overrepresented in pathogenic copy number mutations
60	IL-25 and type 2 innate lymphoid cells induce pulmonary fibrosis
61	Altered inactivation of commensal LPS due to acyloxyacyl hydrolase deficiency in colonic dendritic cells impairs mucosal Th17 immunity
62	Dynamic control of β1 integrin adhesion by the plexinD1-sema3E axis
63	Salmonella exploits NLRP12-dependent innate immune signaling to suppress host defenses during infection
64	β-Catenin induces T-cell transformation by promoting genomic instability
65	An amphioxus RAG1-like DNA fragment encodes a functional central domain of vertebrate core RAG1
66	Alloreactive cytotoxic T cells provide means to decipher the immunopeptidome and reveal a plethora of tumor-associated self-epitopes
67	mTOR target NDRG1 confers MGMT-dependent resistance to alkylating chemotherapy

68	Dual-modality gene reporter for in vivo imaging
69	Epstein-Barr Virus Nuclear Antigen 3C binds to BATF/IRF4 or SPI1/IRF4 composite sites and recruits Sin3A to repress CDKN2A
70	Neisseria meningitidis NalP cleaves human complement C3, facilitating degradation of C3b and survival in human serum
71	Identification of secreted bacterial proteins by noncanonical amino acid tagging
72	Mathematical modeling of primary succession of murine intestinal microbiota
73	A common solution to group 2 influenza virus neutralization
74	Human herpesvirus 6 (HHV-6) alters E2F1/Rb pathways and utilizes the E2F1 transcription factor to express viral genes
75	The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping
76	Local domains of motor cortical activity revealed by fiber-optic calcium recordings in behaving nonhuman primates
77	Protein kinase LKB1 regulates polarized dendrite formation of adult hippocampal newborn neurons
78	Comparison of explicit and incidental learning strategies in memory-impaired patients
79	Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex
80	Presynaptic mitochondrial morphology in monkey prefrontal cortex correlates with working memory and is improved with estrogen treatment
81	Bidirectional homeostatic plasticity induced by interneuron cell death and transplantation in vivo
82	Mechanisms underlying subunit independence in pyramidal neuron dendrites
83	Tonic $GABA_A$ conductance bidirectionally controls interneuron firing pattern and synchronization in the CA3 hippocampal network
84	Neurofibrillary tangle-bearing neurons are functionally integrated in cortical circuits in vivo
85	Cumulative latency advance underlies fast visual processing in desynchronized brain state
86	Optical control of trimeric P2X receptors and acid-sensing ion channels
87	Arabidopsis EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns
88	Overexpression of plasma membrane H+-ATPase in guard cells promotes light-induced stomatal opening and enhances plant growth
89	Data-poor management of African lion hunting using a relative index of abundance
90	Growth feedback as a basis for persister bistability
91	Identification of key regulators for the migration and invasion of rheumatoid synoviocytes through a systems approach
92	Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity
93	Distinct kinetics of synaptic structural plasticity, memory formation, and memory decay in massed and spaced learning
94	Effective functional maturation of invariant natural killer T cells is constrained by negative selection and T-cell antigen receptor affinity
95	Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate
96	SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations
97	De novo selection of oncogenes
98	Pch2 is a hexameric ring ATPase that remodels the chromosome axis protein Hop1
99	BK channel opening involves side-chain reorientation of multiple deep-pore residues
100	Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony
101	Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis
102	Structural and biochemical basis for the inhibition of cell death by APIP, a methionine salvage enzyme
103	AFF1 is a ubiquitous P-TEFb partner to enable Tat extraction of P-TEFb from 7SK snRNP and formation of SECs for HIV transactivation
104	Dual role for Islet-1 in promoting striatonigral and repressing striatopallidal genetic programs to specify striatonigral cell identity
105	Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy
106	Parkinson-related LRRK2 mutation R1441C/G/H impairs PKA phosphorylation of LRRK2 and disrupts its interaction with 14-3-3
107	Ghrelin triggers the synaptic incorporation of AMPA receptors in the hippocampus
108	Origins of R_2* orientation dependence in gray and white matter
109	Cytoglobin modulates myogenic progenitor cell viability and muscle regeneration
110	FMRP and Ataxin-2 function together in long-term olfactory habituation and neuronal translational control
111	A Cdc42- and Rac-interactive binding (CRIB) domain mediates functions of coronin
112	Distinct cerebellar engrams in short-term and long-term motor learning
113	Interplay of mevalonate and Hippo pathways regulates RHAMM transcription via YAP to modulate breast cancer cell motility
114	Trapping of naive lymphocytes triggers rapid growth and remodeling of the fibroblast network in reactive murine lymph nodes

Table 1. Articles in the first issue of *PNAS* in 2014.

There was no corresponding WoS subject category of the *PNAS* subject category "sustainability science." The 89th article was labelled as such by *PNAS*, and it discussed the sustainable management of terrestrial lion hunting. Our method assigned it to the WoS subject category "ecology" which is also suitable. There was also no WoS counterpart of the *PNAS* subject category "systems biology," to which the 91st and 92nd articles belonged. "Systems biology" adopts a holistic approach to the study of complex biological systems. The 91st article studied fibroblast-like synoviocytes of rheumatoid arthritis; and the 92nd article studied growth of bacteria. Our method categorized these as "rheumatology" and "microbiology," respectively, which was an accurate reflection of their content.

In Table 2, there are five articles (51st to 55th) assigned to the *PNAS* subject category "evolution." Only the 53rd article was recognized as the WoS subject category "evolutionary biology." The other four articles introduced their research background as "evolution," and used methods of "biochemistry & molecular biology," "cell biology," "developmental biology," "genetics & heredity," and "zoology" to investigate specific aspects of evolution. Our method correctly identified the related subject categories of these four articles from their references.

There were 12 articles solely labelled as "biophysics and computational biology" by *PNAS* (30th to 38th, 96th, 99th and 105th). Nine were identified as the WoS subject category "biochemistry & molecular biology," not exactly similar to the *PNAS* subject category. The reason may be that there are differences in the definitions of the subject categories between *PNAS* and WoS. Some aspects of physical chemistry or chemical physics related to biology are assigned to biophysics in *PNAS*, but to biochemistry in WoS. The 30th article was recognized as "chemistry, multidisciplinary" and multidisciplinary sciences as its references were published in journals covering many disciplines. The research area of the 32nd article was "immunology" as reflected by its references, while it used technology of "biophysics and computational biology." There was a similar case with the 33rd article's research area, which according to the references was "microbiology."

The method identified more subject categories than were labelled by *PNAS* for many articles. The additional recognized subject categories were relevant to the articles. For example, the *PNAS* subject category of article 66 was "immunology," which was recognized by the method. This article studied the immunotherapy of cancer, and involved detection of HLA-A2-bound peptides from two leukaemia-associated differentiation antigens. Thus, it was also appropriately recognized as being in the subject categories of "hematology" and "oncology." One other recognized subject category "biochemical research methods" correlated with the subject categories above. The other recognized subject category of multidisciplinary sciences indicated that it contained some references from multidisciplinary sciences journals.

Among the 114 articles examined in this case study, we regarded 78 articles as having been almost perfectly classified. Excluding "multidisciplinary sciences," their recognized subject category with the highest score was the same as, belonged to, or covered their *PNAS* subject category (or its counterpart). If one of those articles had two *PNAS* subject categories, both categories were identified. For $d = 3$, we identified *PNAS* subject categories for 22 articles but not as the first recognized subject category. We regard those articles as acceptably classified. For $d = 4$, 25 articles were acceptably classified. The acceptably classified articles included three articles (89th, 91st, and 92nd) whose *PNAS* subject category had no counterpart in WoS. Eleven articles had identified subject categories other than *PNAS* subject categories. Though those articles related to the identified subject categories, the adopted method failed to identify all aspects of the articles. All 11 problematically classified articles were interdisciplinary ones. However, other interdisciplinary articles in the almost perfectly and acceptably classified groups were successfully classified.

4.1 Results with the recognized subject category of references in multidisciplinary sciences journals

As shown in Table 2, multidisciplinary sciences was the recognized subject category of 86 articles because they cited many references in these types of journals. Some even had multidisciplinary sciences in the first two recognized subject categories in terms of score. However, the subject category multidisciplinary sciences provided no useful information on article content in most cases. To further use the information from the references, we can modify the classification method above using the following steps:

1. Recognize the subject categories of the references of articles which are published in multidisciplinary sciences journals with the method described in Equations 1 to 4 (this is feasible because references in multidisciplinary sciences journals are also articles);
2. Use the recognized subject categories of each reference in multidisciplinary sciences journals obtained in Step 1 to replace the subject category multidisciplinary sciences of the reference;
3. Classify the article inspected using the method described in Equations 1 to 4 with the subject categories of references modified in Step 2.

For example, we selected 10 articles with multidisciplinary sciences in the top two recognized subject categories, and further refined the results as shown in Table 3 (threshold factor of 3) and Table 4 (threshold factor of 4).

Comparing Tables 3 and 4 with Table 2, the recognized subject categories of more than half the articles by the modified method in this section were more focused than those determined by the method without modifica-

No.	Subject$_{pnas1}$ [a, b]	Subject$_{pnas2}$ [a, b]	Sc$_{recog1}$ [a, c]	Sc$_{recog2}$ [a, c]	Sc$_{recog3}$ [a, c]	Sc$_{recog4}$ [a, c]	Sc$_{recog5}$ [a, c]	Sc$_{recog6}$ [a, c]
1	AppPhySci	neuroscience	neurosciences (18.7)	rnmmi (8.87)	multisci (8)	*optics* (5)		
2	AppPhySci		multisci (10)	cell biology (8.5)	biophysics (3.5)	biomolbio (3)	*develbio* (2.5)	*phymul* (2.25)
3	Chemistry	biochemistry	hematology (7.5)	biomolbio (4.17)	biophysics (4.2)	chemmul (3)	microbiology (3)	chemphy (2.67)
4	Chemistry	biocombio	multisci (7)	develbio (4.5)	zoology (3)	chemmul (2.5)	cell biology (1.83)	matscimul (1.5)
5	Chemistry		electrochemistry (11.8)	chemphy (5.5)	*multisci* (3.5)	*matscicf* (3)		
6	EAPSci		geomul (27.3)	geogeo (15.5)	geology (9.5)	*multisci* (9)		
7	EAPSci		geomul (14.8)	geology (7)	*geogeo* (4)			
8	Physics		phymul (14)	phyconmat (9.12)	multisci (9)	*matscimul* (4.3)		
9	Physics		phymul (12)	phyconmat (11.1)	multisci (10)	chemphy (4.5)		
10	Physics		phymul (19.3)	phyconmat (7.17)	*pamc* (6)			
11	Anthropology		anthropology (7)	geomul (5.5)	multisci (4)	*evobio* (2.3)		
12	PsyCogSci		psychiatry (11.5)	multisci (8)	psyclin (5.5)	psychology (4.5)	*psyexp* (3)	*ecology* (3)
13	PsyCogSci		psychiatry (7.5)	medgenint (5)	psymul (3)	psysoc (2.5)	*psychology* (2.33)	*peoh* (2)
14	Social Sciences		economics (9.5)					
15	AgriSci		plant sciences (9.17)	zoology (5)	biomolbio (4)	*multisci* (3)		
16	Anthropology		evobio (6.67)	ecology (6)	multisci (5)	anthropology (5)	biology (4.33)	cell biology (3)
17	Anthropology		multisci (11)	geomul (5)	geogeo (2)	anthropology (2)	*biology* (1.33)	
18	Anthropology		evobio (19)	multisci (10)				
19	AppBioSci		multisci (7)	bioappmic (6.83)	biomolbio (3.3)	*medresexp* (1.8)		
20	AppBioSci		bioappmic (19.3)	microbiology (8.42)	biomolbio (8.1)			
21	Biochemistry		biomolbio (22.8)					
22	Biochemistry		chemmul (22)	biomolbio (8)	*oncology* (5.5)			
23	Biochemistry		chemmul (16)	biomolbio (9.67)	*multisci* (5)			
24	Biochemistry		biomolbio (8.5)	cell biology (7.5)	multisci (4)			
25	Biochemistry		biomolbio (7)	hematology (4.83)	pervasdis (3.3)	oncology (3)	*develbio* (2)	*genher* (2)
26	Biochemistry		biomolbio (12)	cell biology (10.8)	genher (6.8)	multisci (5)	*biology* (3.5)	
27	Biochemistry		biomolbio (10.4)	oncology (9.25)	*phampham* (3.2)	*chemmed* (3.2)		
28	AppPhySci	biocombio	biophysics (12.8)	biomolbio (6.33)	chemphy (6.1)			
29	Chemistry	biocombio	multisci (13)	biomolbio (11.8)	biophysics (4)	*spectroscopy* (3.7)	*chemmul* (3)	*cell biology* (3)
30	BioComBio		chemmul (12.8)	multisci (8)				
31	BioComBio		biomolbio (8)	multisci (7)	neurosciences (4)	genher (3)		

No.	$Subject_{pnas1}$ [a, b]	$Subject_{pnas2}$ [a, b]	Sc_{recog1} [a, c]	Sc_{recog2} [a, c]	Sc_{recog3} [a, c]	Sc_{recog4} [a, c]	Sc_{recog5} [a, c]	Sc_{recog6} [a, c]
32	BioComBio		immunology (8.5)					
33	BioComBio		multisci (13)	microbiology (12)	bioappmic (4.5)			
34	BioComBio		biomolbio (14)	multisci (6)	*genher* (3.5)			
35	BioComBio		biomolbio (9.33)	*multisci* (3)				
36	BioComBio		biomolbio (14.5)	multisci (10)	cell biology (8.8)			
37	BioComBio		plant sciences (20.3)	biomolbio (11.3)	multisci (8)	cell biology (7.8)		
38	BioComBio		biomolbio (9.67)	multisci (5)	crystallography (4.5)	bioresmeth (4.5)	biophysics (3.67)	*cell biology* (2.67)
39	Cell Biology		virology (11.7)	biomolbio (11)	cell biology (7.5)	*multisci* (3)		
40	Cell Biology		biomolbio (9.75)	multisci (9)	oncology (7.8)	cell biology (6.3)	genher (4)	hematology (3.5)
41	Cell Biology		oncology (5.2)	cell biology (4.03)	multisci (3)	*biomolbio* (1.5)	*gerher* (1.33)	
42	Cell Biology		multisci (11)	cell biology (9.5)	biomolbio (5.3)	*genher* (2.7)		
43	Cell Biology		cell biology (22.2)					
44	Cell Biology		biomolbio (9)	cell biology (7.5)	multisci (6)			
45	Cell Biology		multisci (6)	biomolbio (6)	cell biology (5)	*oncology* (1.5)		
46	Cell Biology		multisci (9)	cell biology (8.75)	biomolbio (7.8)	oncology (4.3)		
47	Cell Biology		oncology (15.8)	cell biology (8.92)	*biomolbio* (4.6)			
48	Physics	ecology	multisci (13)	ecology (8.67)	phymul (3)	*genher* (2.5)	*evobio* (2.5)	*biology* (2.33)
49	Ecology		ecology (23.4)	multisci (11)	evobio (8.6)			
50	Ecology		ecology (12.8)	multisci (10)	*biology* (3.3)			
51	Evolution		biomolbio (14.5)	genher (5.83)	*biology* (4)	*cell biology* (3.7)		
52	Evolution		multisci (10)	microbiology (5)	biomolbio (3.8)	entomolgy (2)	bioappmic (2)	biology (1.83)
53	Evolution		multisci (9)	ecology (9)	zoology (8.7)	behsci (7.2)	evobio (5.33)	biology (3.67)
54	Evolution		multisci (8)	biomolbio (7.83)	cell biology (7.2)	develbio (3.3)	medgenint (3)	bioresmeth (3)
55	Evolution		develbio (14.9)	cell biology (8.25)	biomolbio (7.4)	multisci (7)		
56	Genetics		genher (12.3)	neurosciences (6)	multisci (5)	biomolbio (4.8)		
57	Genetics		biomolbio (18)	multisci (7)				
58	Genetics		microbiology (26.2)	genher (9.83)				
59	Genetics		genher (33.8)	*multisci* (11)				
60	Immunology		immunology (21.5)					
61	Immunology		immunology (18.2)	multisci (9)				
62	Immunology		immunology (21.5)	multisci (11)	cell biology (7.5)			
63	Immunology		immunology (14.5)	multisci (8)	*biomolbio* (4)			
64	Immunology		multisci (11)	immunology (10)	biomolbio (8.5)	cell biology (6.7)	hematology (5.5)	oncology (4.5)
65	Immunology		biomolbio (9)	immunology (6.5)	multisci (5)	cell biology (3.3)		
66	Immunology		immunology (12)	hematology (9)	oncology (6)	*multisci* (3)	*bioresmeth* (3)	
67	MedSci		oncology (11.8)	biomolbio (4.08)				
68	MedSci		rnmmi (7)	bioresmeth (3)	*bioappmic* (2.3)	*biomolbio* (2.3)		

No.	Subject$_{pnas1}$ [a, b]	Subject$_{pnas2}$ [a, b]	Sc$_{recog1}$ [a, c]	Sc$_{recog2}$ [a, c]	Sc$_{recog3}$ [a, c]	Sc$_{recog4}$ [a, c]	Sc$_{recog5}$ [a, c]	Sc$_{recog6}$ [a, c]
69	Microbiology		virology (17.3)	multisci (10)				
70	Microbiology		microbiology (11)	immunology (9.83)	biomolbio (5.8)	infecdis (3.8)	*multisci* (3)	
71	Microbiology		microbiology (15.8)	biomolbio (8.75)	bioresmeth (6)	*multisci* (4)		
72	Microbiology		multisci (8)	microbiology (4)	biomolbio (4)	microbiology (3.7)	bioappmic (3.67)	ecology (3)
73	Microbiology		multisci (15)	virology (5.67)	immunology (4.8)	biomolbio (3.3)	medresexp (2.83)	*microbiology* (1.5)
74	Microbiology		virology (9.83)	multisci (4)	cell biology (3.8)	biomolbio (3.8)		
75	Microbiology		microbiology (15)	biomolbio (13.6)	multisci (6)			
76	Neuroscience		neurosciences (15)	multisci (8)	*physiology* (4)			
77	Neuroscience		cell biology (13)	multisci (13)	biomolbio (9)	neurosciences (8)		
78	Neuroscience		neurosciences (4.83)	multisci (4)	psychology (3)	*behsci* (1.5)		
79	Neuroscience		neurosciences (28.3)					
80	Neuroscience		neurosciences (25.2)	*multisci* (7)				
81	Neuroscience		neurosciences (35.5)	multisci (15)				
82	Neuroscience		neurosciences (42.5)	multisci (17)				
83	Neuroscience		neurosciences (17.5)	*multisci* (5)				
84	Neuroscience		neurosciences (16.3)	multisci (6)				
85	Neuroscience		neurosciences (30)	multisci (10)				
86	Physiology		multisci (14)	neurosciences (10.5)	biomolbio (7.5)	chemmul (5)	physiology (4.5)	
87	Plant Biology		multisci (11)	cell biology (8.33)	genher (6.5)	biomolbio (6.3)	plant sciences (6)	
88	Plant Biology		plant sciences (16.3)	multisci (13)	*cell biology* (4.3)			
89	SusSci		ecology (4.67)	multisci (4)	fisheries (3.7)	biocon (2.2)	envsci (1.67)	*evobio* (1.5)
90	Physics	sysbio	microbiology (10)	multisci (5)				
91	SysBio		rheumatology (12)	*genher* (3)				
92	SysBio		microbiology (12.7)	multisci (7)				
93	Neuroscience		neurosciences (33)					
94	Immunology		immunology (32)					
95	Microbiology		microbiology (11.3)	multisci (11)	bioappmic (9.3)	fscitec (8.2)	plant sciences (4)	*bioresmeth* (3.67)
96	BioComBio		biomolbio (8.67)	multisci (6)	*biophysics* (2.7)			
97	Biochemistry		biomolbio (12.5)	virology (9)	multisci (9)			
98	Biochemistry		biomolbio (21.8)	cell biology (15.3)	genher (14)	*multisci* (7)		
99	BioComBio		physiology (22)	multisci (15)	neurosciences (13)	*biomolbio* (6)		
100	Neuroscience		neurosciences (27.5)	multisci (18)				
101	Plant Biology		plant sciences (18.8)	biomolbio (16.3)	multisci (13)			
102	Biochemistry		biomolbio (10.5)	oncology (3.75)	bioresmeth (3.7)	*multisci* (3)		
103	Biochemistry		biomolbio (13.5)	cell biology (6.83)	multisci (6)			
104	Neuroscience		neuroscience s (21.5)					
105	BioComBio		immunology (12)	multisci (11)	biomolbio (7.7)	virology (5.3)	*cell biology* (3.67)	

No.	$Subject_{pnas1}$ [a, b]	$Subject_{pnas2}$ [a, b]	Sc_{recog1} [a, c]	Sc_{recog2} [a, c]	Sc_{recog3} [a, c]	Sc_{recog4} [a, c]	Sc_{recog5} [a, c]	Sc_{recog6} [a, c]
106	Biochemistry		biomolbio (21.7)	multisci (9)	*cell biology* (6.2)	*neurosciences* (5.5)		
107	Neuroscience		neurosciences (31.3)	multisci (11)				
108	Neuroscience		rnmmi (14.1)	*neurosciences* (4.17)				
109	MedSci		biomolbio (19.8)	cell biology (15.3)	multisci (9)	*physiology* (6.2)		
110	Genetics		neurosciences (33.1)	multisci (17)	cell biology (16)	biomolbio (12)		
111	Biochemistry		cell biology (20.5)	biomolbio (9.5)				
112	Neuroscience		neurosciences (24.8)					
113	Cell Biology		oncology (13.5)	cell biology (8.17)	biomolbio (7.5)	multisci (6)		
114	Immunology		immunology (38.5)					

Table 2. Comparison of the *PNAS* subject categories and the recognized subject categories of articles in Table 1. In the recognized results, non-italics represent the results when $d = 3$ in Equation 4. When $d = 4$, additional results were obtained and are represented by italics.

Note:
a. $Subject_{pnas1}$ and $Subject_{pnas2}$ are the subjects of the article labelled by "*PNAS*". SC_{recog1}, SC_{recog2}, SC_{recog3}, SC_{recog4}, SC_{recog5}, and SC_{recog6} are the recognized subject categories of the article.
b. Abbreviation of some *PNAS* subject categories:

AgriSci: agricultural sciences;
AppBioSci: applied biological sciences;
AppPhySci: applied physical sciences;
BioComBio: biophysics and computational biology;
EAPSci: earth, atmospheric, and planetary sciences;
MedSci: medical sciences;
PsyCogSci: psychological and cognitive sciences;
SysBio: systems biology;
SusSci: sustainability science.
c. abbreviation of some subject categories labelled by wos:
BehSci: behavioral sciences;
BioAppMic: biotechnology & applied microbiology;
BioCon: biodiversity conservation;
BioMolBio: biochemistry & molecular biology;
BioResMeth: biochemical research methods;
ChemMed: chemistry, medicinal;
ChemMul: chemistry, multidisciplinary;
ChemPhy: chemistry, physical;
DevelBio: developmental biology;
EnvSci: environmental sciences;
EvoBio: evolutionary biology;
FSciTec: food science & technology;
GenHer: genetics & heredity;
GeoGeo: geochemistry & geophysics;
GeoMul: geosciences, multidisciplinary;
InfecDis: infectious diseases;
MatSciMul: materials science, multidisciplinary;
MedGenInt: medicine, general & internal;
MedResExp: medicine, research & experimental;
MatSciCF: materials science, coatings & films;
MultiSci: multidisciplinary sciences;
PAMC: physics, atomic, molecular & chemical;
PhyConMat: physics, condensed matter;
PsyClin: psychology, clinical;
PEOH: public, environmental & occupational health;
PhyMul: physics, multidisciplinary;
PhamPham: pharmacology & pharmacy;
PsyExp: psychology, experimental;
PsyMul: psychology, multidisciplinary;
PsySoc: psychology, social;
PerVasDis: peripheral vascular disease;
RNMMI: radiology, nuclear medicine & medical imaging.

No.	Subject$_{PNAS1}$	Subject$_{PNAS2}$	SC$_{recog1}$	SC$_{recog2}$	SC$_{recog3}$	SC$_{recog4}$	SC$_{recog5}$	S$_{Crecog6}$
2	appphysci		cell biology (12.32)	develbio (4.44)				
4	chemistry	biocombio	develbio (4.57)	zoology (3.30)	chemmul (3.27)	multisci (2.23)	cell biology (2.20)	matscimul (1.81)
17	anthropology		geomul (7.51)	multisci (4.66)	anthropology (2.85)			
32	biocombio		immunology (9.22)					
52	evolution		biomolbio (7.01)	microbiology (6.98)	entomolgy (2.86)			
53	evolution		ecology (10.57)	zoology (9.58)	behsci (7.88)	evobio (6.20)	biology (4.25)	
54	evolution		biomolbio (10.76)	cell biology (9.19)	develbio (4.19)			
77	neuroscience		neurosciences (14.24)	cell biology (14.24)	biomolbio (9.74)			
87	plant biology		cell biology (10.60)	biomolbio (9.25)	genher (8.12)	plant sciences (7.01)		
105	biocombio		immunology (15.17)	biomolbio (10.44)	virology (6.00)			

Table 3. Comparison of the *PNAS* subject categories and the recognized subject categories of 10 articles in Table 1. The subject categories of references in multidisciplinary sciences journals were replaced by their recognized subject categories ($d = 3$). Others are the same as in Table 2.

No.	Subject$_{PNAS1}$	Subject$_{PNAS2}$	SC$_{recog1}$	SC$_{recog2}$	SC$_{recog3}$	SC$_{recog4}$	SC$_{recog5}$	S$_{Crecog6}$
2	appphysci		cell biology (11.98)	develbio (4.40)	*biophysics* (3.93)	*biomolbio* (3.65)		
4	chemistry	biocombio	develbio (4.56)	zoology (3.33)	chemmul (3.27)	cell biology (2.20)	multisci (2. 08)	matscimul (1.75)
17	anthropology		geomul (7.48)	multisci (4.63)	anthropology (2.85)	*geogeo* (2.40)		
32	biocombio		immunology (9.16)					
52	evolution		microbiology (7.04)	biomolbio (6.86)	entomolgy (2.85)	*bioappmic* (2.15)	*biology* (1.94)	*multisci* (1.93)
53	evolution		ecology (10.67)	zoology (9.52)	behsci (7.96)	evobio (6.21)	biology (4.34)	
54	evolution		biomolbio (10.24)	cell biology (9.18)	develbio (4.24)	medgenint (3.4)	bioresmeth (3)	
77	neuroscience		cell biology (14.21)	neurosciences (13.68)	biomolbio (9.86)			
87	plant biology		cell biology (10.57)	biomolbio (9.21)	genher (8.18)	plant sciences (7.08)	*multisci* (3.02)	
105	biocombio		immunology (15.11)	biomolbio (10.44)	virology (5.88)	*cell biology* (4.86)		

Table 4. Comparison of the *PNAS* subject categories and the recognized subject categories of 10 articles in Table 1. The subject categories of references in multidisciplinary sciences journals were replaced by their recognized subject categories ($d = 4$). Others are the same as in Table 2.

tion. The number of articles in Tables 3 and 4 with recognized subject categories of multidisciplinary sciences was also reduced, especially with a threshold factor of 3. In addition, for articles recognized as multidisciplinary sciences in Tables 3 and 4, the ranking of multidisciplinary sciences in the recognized subject categories was lowered. For article 2, the ranking of "developmental biology," which is its area of application, climbed to 2 in Tables 3 and 4 from 5 in Table 2. The authors of the article used a mechanical model to study shapes of epithelial cells and the bending and buckling of epithelial sheets during embryo development, thus the recognized subject category was appropriate. Table 4 included "biophysics" for article 2 which reflected the methodology used in the study and showed that lowering the threshold in Equation 4 can reduce loss of an article subject category in the determination of the classification. With the recognized subject category of the references in multidisciplinary sciences journals of article 17, the rank of the subject category "anthropology," which is also the label used by *PNAS*, rose to be behind only "geosciences, multidisciplinary" and multidisciplinary sciences which cover many subject categories. Article 32 only had one subject category recognized in Table 2. The subject category with the second highest score was multidisciplinary sciences which is not listed in Table 2 for article 32 because its score was less than a quarter of the highest one. After classifying the references in the multidisciplinary sciences journals, the score of multidisciplinary sciences decreased to one third of its previous value in Table 2 (=2); about one third of the previous score of multidisciplinary sciences was distributed to the subject category with the highest score (Immunology), and the remaining third was distributed to the other 10 subject categories. *PNAS* assigned article 77 to "neurosciences," and in Table 2, the score of "neurosciences" was distinctly lower than the highest scoring category. However, in Tables 3 and 4, the score of "neurosciences" was highest or near the highest. The scores of the recognized subject categories of article 4 were distributed relatively evenly, even after its multidisciplinary sciences references were classified. The recognized subject categories of some articles, such as article 4 when the threshold factor was 4, were more than 6. These articles can be assigned to multidisciplinary sciences if the number of subject categories is limited to 6.

5.0 Conclusions

This study shows the possibility of more precise determination of subject categories of articles in multidisciplinary sciences journals indexed by a documentation database (such as WoS) according to information from the references.

The subject category of articles in multidisciplinary sciences journals, which often publish high-quality articles, is classified by simply counting the subject categories of the journals in which the references are published. For articles in the *PNAS* subject categories "applied biological sciences," "biochemistry," "cell biology," "ecology," "immunology," "microbiology," "neuroscience," "physiology," "plant biology," "psychological and cognitive sciences," "applied physical sciences," "chemistry," "physics," "earth, atmospheric, and planetary sciences," "medical sciences," and "social sciences" in the analyzed issue of *PNAS*, the method correctly recognized their subject categories. The recognized subject categories of some articles in *PNAS* differed from the *PNAS* subject categories because of recognition of areas of application and research methods used. The *PNAS* subject categories and the recognized subject categories for these articles reflected different aspects of the articles as a consequence of knowledge diffusion (Chen et al. 2009). The recognized subject categories determined by our method differed from the subjects of the articles labelled by *PNAS* as they recognized other aspects of the articles.

In this study, we adopted six subject categories to classify articles, as is done by WoS for journals. Among the 25 acceptably classified articles, 4 had *PNAS* subject categories that were identified as SC_{recog4}, 8 as SC_{recog3}, and 13 as SC_{recog2}. It would appear that using more than four subject categories to classify an article is meaningless. In future studies, we will apply this method to articles published in general journals (such as "chemistry, multidisciplinary" and "physics, multidisciplinary") and in journals of multiple subject categories.

The threshold factor adjusted the number of recognized subject categories. If it was small, the condition for a categorization as a subject other than the one with the highest score recognized was strict. This meant that the possibility of the article being assigned to recognized subject categories other than the article subject categories was small. However, some of the article subject categories may be lost in the result. If the threshold factor is large or the threshold is low, the possibility of the article subject categories being lost will be reduced, but some subject categories less relevant to the article through correlation with the article subject categories will be recognized.

The method can exclude irrelevant subject categories for most articles. For example, if one is interested in "neurosciences" articles in multidisciplinary sciences journals, he/she can ignore articles whose recognized subject categories do not include "neurosciences," such as articles on "physics, condensed matter." This case study also showed that if subject categories of the references in the multidisciplinary sciences journals are identified and used to replace the subject category multidisciplinary sciences of such references, the recognized subject categories of the articles citing them would be more focused.

6.0 Limitation

One key step of the method is to count the number of references indexed by the database for the inspected article. If the number of references indexed by the database is small, the outcome of the method may be unreliable because of a small sample effect. In view of this, the method can be used only when the number of references indexed by WoS is large enough, such as more than 20.

References

Aleixandre, Jose L., José L. Aleixandre-Tudó, Máxima Bolaños-Pizzaro and Rafael Aleixandre-Benavent. 2013. "Mapping the Scientific Research on Wine and Health (2001-2011)." *Journal of Agricultural and Food Chemistry* 61: 11871-80.

Ardanuy, Jordi, Cristóbal Urbano and Lluís Quintana. 2009. "A Citation Analysis of Catalan Literary Studies (1974-2003): Towards a Bibliometrics of Humanities Studies in Minority Languages." *Scientometrics* 81: 347-66.

Bornmann, Lutz. 2014. "Assigning Publications to Multiple Subject Categories for Bibliometric Analysis, an Empirical Case Study Based on Percentiles." *Journal of Documentation* 70: 52-61.

Braam, Robert R., Henk F. Moed and Anthony F. J. Van Raan. 1991. "Mapping of Science by Combined Co-Citation and Co-Word Analysis I. Structural Aspects." *Journal of the American Society for Information Science* 42: 233-51.

Chen, Chaomei, Yue Chen, Mark Horowitz, Haiyan Hou, Zeyuan Liu and Donald Pellegrino. 2009. "Towards an Explanatory and Computational Theory of Scientific Discovery." *Journal of Informetrics* 3: 191-209.

Chen, Huaqi, Yuehua Wan, Shuian Jiang and Yanxia Cheng. 2014. "Alzheimer's Disease Research in the Future: Bibliometric Analysis of Cholinesterase Inhibitors from 1993 to 2012." *Scientometrics* 98: 1865-77.

Chiu, Wen-Ta and Yuh-Shan Ho. 2007. "Bibliometric Analysis of Tsunami Research." *Scientometrics* 73: 3-17.

De La Moneda Corrochano, Mercedes, Maria J. Lopez-Huertas and Evaristo Jimenez-Contreras. 2013. "Spanish Research in Knowledge Organization (2002-2010)." *Knowledge Organization* 40: 28-41.

Diem, Andrea and Stefan C. Wolter. 2013. "The Use of Bibliometrics to Measure Research Performance in Education Sciences." *Research in Higher Education* 54: 86-114.

Gabel, Jeff. 2006. "Improving Information Retrieval of Subjects Through Citation-Analysis." *Knowledge Organization* 33: 86-95.

Glänzel, W., A. Schubert and H. J. Czerwon. 1999. "An Item-By-Item Subject Classification of Papers Published in Multidisciplinary and General Journals Using Reference Analysis." *Scientometrics* 44: 427-39.

Gómez-Núñez, Antonio J., Benjamín Vargas-Quesada, Félix De Moya-Anegón and Wolfgang Glänzel. 2011. "Improving Scimago Journal & Country Rank (SJR) Subject Classification through Reference Analysis." *Scientometrics* 89: 741-58.

Gouvea Meireles, Magali Rezende, Valadares Cendón Cendon and Paulo Eduardo Maciel De Almeida. 2014. "Bibliometric Knowledge Organization: A Domain Analytic Method Using Artificial Neural Networks." *Knowledge Organization* 41: 145-59.

Griffith, Belver C., Henry. G. Small, Judith A. Stonehill and Sandra Dey. 1974. "The Structure of Scientific Literatures. II: Toward a Macro- and Microstructure for Science." *Science Studies* 4: 339-65.

Grossi, F., O. Belvedere and R. Rosso. 2003. "Geography of Clinical Cancer Research Publications from 1995 to 1999." *European Journal of Cancer* 39: 106-11.

Ibekwe-Sanjuan, Fidelia and Eric Sanjuan. 2002. "From Term Variants to Research Topics." *Knowledge Organization* 29: 181-97.

Klavans, Richard and Kevin Boyack. 2010. "Toward an Objective, Reliable and Accurate Method for Measuring Research Leadership." *Scientometrics* 82, no. 3: 539-53.

Krampen, Günter, Alexander Von Eye and Gabriel Schui. 2011. "Forecasting Trends of Development of Psychology from a Bibliometric Perspective." *Scientometrics* 87: 687-94.

Lewison, G. 1999. "The Definition and Calibration of Biomedical Subfields." *Scientometrics* 46: 529-37.

Liu Xingjian, F. Benjamin Zhan, Song Hong, Beibei Niu and Yaolin Liu. 2012. "A Bibliometric Study of Earthquake Research: 1900–2010." *Scientometrics* 92: 747-65.

López-Illescas, Carmen, Ed C. M. Noyons, Martijn S. Visser, Félix De Moya-Anegón and Henk F. Moed. 2009. "Expansion of Scientific Journal Categories Using Reference Analysis: How Can It Be Done and Does It Make a Difference?" *Scientometrics* 79: 473-90.

Marx, Werner and Lutz Bornmann. 2010. "How Accurately Does Thomas Kuhn's Model of Paradigm Change Describe the Transition from the Static View of the Universe to the Big Bang Theory in Cosmology? A Historical Reconstruction and Citation Analysis." *Scientometrics* 84: 441-64.

Moppett, I. K. and J. G. Hardman. 2011. "Bibliometrics of Anaesthesia Researchers in the UK." *British Journal of Anaesthesia* 107: 351-56.

Naqvi, Shehbaz Husain. 2014. "Polymer Science Research in India During 1999–2012: A Scientometric Study Based on Science Citation Index-Expanded." *Science, Technology and Society* 19: 95-108.

Pinto, María, María Isabel Escalona-Fernández and Antonio Pulgarín. 2013. "Information Literacy in Social Sciences and Health Sciences: A Bibliometric Study (1974–2011)." *Scientometrics* 95: 1071-94.

Porter, Alan L. and Jan Youtie. 2009. "How Interdisciplinary Is Nanotechnology?" *Journal of Nanoparticle Research* 11: 1023-41.

Small, Henry and Belver C. Griffith. 1974. "The Structure of Scientific Literatures. I: Identifying and Graphing Specialties." *Science Studies* 4: 17-40.

Small, Henry and E. Sweeney. 1985. "Clustering the Science Citation Index Using Co-Citations. I. A Comparison of Methods." *Scientometrics* 7: 391-409.

Small, Henry, E. Sweeney and E. Greenlee. 1985. "Clustering the Science Citation Index Using Co-Citations. II. Mapping Science." *Scientometrics* 8: 321-40.

Small, Henry. 1998. "A General Framework for Creating Large-Scale Maps of Science in Two Or Three Dimensions: The Sciviz System." *Scientometrics* 41: 125-33.

Waltman, Ludo and Nees Jan Van Eck. 2012. "A New Methodology for Constructing a Publication-Level Classification System of Science." *Journal of the American Society for Information Science and Technology* 63: 2378-92.

Waltman, Ludo. 2012. "An Empirical Analysis of the Use of Alphabetical Authorship in Scientific Publishing." *Journal of Informetrics* 6: 700-11.

Wang, Haijun, Minyan Liu, Song Hong and Yanhua Zhuang. 2013. "A Historical Review and Bibliometric Analysis of GPS Research from 1991–2010." *Scientometrics* 95: 35-44.

Yang, Lie, Zhulie Chen, Ting Liu, Zhe Gong, Yingjian Yu and Jia Wang. 2013. "Global Trends of Solid Waste Research from 1997 to 2011 by Using Bibliometric Analysis." *Scientometrics* 96:133-46.

Zhuang, Yanhua, Xingjia Liu, Thuminh Nguyen, Qingqing He and Song Hong. 2013. "Global Remote Sensing Research Trends During 1991–2010: A Bibliometric Analysis." *Scientometrics* 96: 203-19.