



BIBLIOMETRIC METHODS FOR DETECTING AND ANALYSING EMERGING RESEARCH TOPICS



Wolfgang Glänzel



Wolfgang Glänzel is director of the *Centre for R&D Monitoring (ECOOM)* and professor at the *Katholieke Universiteit Leuven*. He is also senior scientist at the *Institute for Science Policy Research, Hungarian Academy of Sciences* in Budapest. He is co-editor of the journal *Scientometrics* and secretary/treasurer of the *International Society for Scientometrics and Infometrics (ISSI)*.

Katholieke Universiteit Leuven
Faculty of Business and Economics (FBE)
Waaistraat, 6 - box 3536. 3000 Leuven, Belgium
Wolfgang.Glanzel@econ.kuleuven.be

Abstract

This study gives an overview of the process of clustering scientific disciplines using hybrid methods, detecting and labelling emerging topics and analysing the results using bibliometrics methods. The hybrid clustering techniques are based on bibliographic coupling and text-mining and 'core documents', and cross-citation links are used to identify emerging fields. The collaboration network of those countries that proved to be most active in the underlying disciplines, in combination with a set of standard indicators, form the groundwork for the bibliometric analysis of the detected emerging research topics.

Keywords

Bibliometrics, Clustering, Hybrid clustering, Emerging topics, Emerging fields, Research topics, Core documents, Bibliographic coupling, Text-mining, Collaboration networks, Countries, International collaboration.

Título: Métodos bibliométricos para la detección y análisis de temas de investigación emergentes

Resumen

Se ofrece una visión general del proceso de agrupación de las disciplinas científicas con métodos híbridos, detectando y etiquetando temas emergentes y analizando los resultados utilizando métodos bibliométricos. Las técnicas de clustering híbrido se basan en acoplamiento bibliográfico y minería de texto, y "documentos básicos". Se usan enlaces entre las citas para identificar los campos emergentes. La red de colaboración de los países que han demostrado ser los más activos en las disciplinas fundamentales, en combinación con un conjunto de indicadores estándar, forman las bases para el análisis bibliométrico de los temas de investigación emergentes detectados.

Palabras clave

Bibliometría, Agrupación, Clustering híbrido, Temas emergentes, Campos emergentes, Investigación, Documentos básicos, Documentos nucleares, Acoplamiento bibliográfico, Minería de texto, Redes de colaboración, Análisis por países.

Glänzel, Wolfgang. "Bibliometric methods for detecting and analysing emerging research topics". *El profesional de la información*, 2012, marzo-abril, v. 21, n. 1, pp. 194-201.

<http://dx.doi.org/10.3145/epi.2012.mar.11>

This paper is scheduled to be presented at the *VI Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología "Gilberto Sotolongo Aguilar"*, coinciding with *INFO 2012*, La Habana, Cuba, 16-20 April 2012.

Introduction

One of the big challenges in scientometrics is the mapping of the epistemological structure of science. In particular, monitoring the dynamics and evolution of the found structures, notably of the emergence of new research topics, has become the focus of interest. Scientometricians have developed quite a range of methods and tools to map and study

the cognitive-epistemological structure of science. Citation-links between documents, keywords and textual descriptors extracted from the text body of scientific documents have become the cognitive basis of the favourite techniques in the mapping of science. The first cognitive maps, which were based on co-citation links (**Small**, 1973), were regularly published in *ISI's Atlas of Science* (see, for instance, **Garfield**, 1987). However, co-citation analysis has severe limitations as

found by **Diana Hicks** (1976), among others. One of the most critical issues is that the time needed to produce the highly cited publications on which the co-citation mapping is based might be too long for the detection of new research topics. By contrast, bibliographic coupling has obvious advantages compared with co-citation clustering (**King**, 1987, **Glänzel**; **Czerwon**, 1996). The most important one is that as soon as papers are published and indexed in the database, those papers that are closely related by bibliographic coupling can provide snapshots of early stages of a research topic's evolution. This property is shared by text-based techniques. Co-word analysis (**Callon et al.**, 1983) or techniques based on the text analysis of the natural language used in the document have, however, other limitations. While citation-based methods often result in extremely sparse matrices (**Janssens**, 2007) or a large number of singletons when clustering the document space (**Jarneving**, 2005), text-based methods tend to overestimate similarities between documents, most notably when the same vocabulary is used in local environments, in which new research topics also emerge. Clustering of a given document space can therefore be based on combinations of different bibliometric similarity measures. As early as 1980, **Bichteler** and **Eaton** suggested the combined use of bibliographic coupling and co-citation in the context of document retrieval. Later, in the early 1990s, **Braam et al.** (1991a,b) and **Zitt**; **Bassecouard** (1994) combined textual and citation-link methods to improve the efficiency of clustering and to reduce the effect of the known shortcomings of the individual components. In particular, the combination makes it possible to cluster documents whenever citation links are weak or missing. This feature is most important in social sciences and humanities fields where citations to journal articles play a less important role than in the sciences. The advantage of this combination has been confirmed by several recent studies (e.g., **Boyack**; **Klavans**, 2010; **Janssens et al.** 2008) as well. Furthermore, the textual component can be used to label the obtained clusters (**Glänzel et al.**, 2009). Following these studies, the present paper also uses a combined text- and citation-based hybrid approach to analyse the underlying structures within which emerging topics are expected.

Methods

In a first step, scientific disciplines such as *JCR Subject Categories* with striking growth patterns are selected for cluster analysis in different subsequent but not overlapping periods. The scheme of a complete bibliometric analysis of an emerging topic can be summarised as follows.

Identification and analysis of emerging topics

For the scientometric analysis of emerging topics six steps are suggested, which are briefly summarised below. Some of these steps are optional and not always required for the analysis.

1. Structural analysis of the discipline

The first step is the selection of the field or discipline in which the emergent topic is sought. The discipline might be pre-defined (e.g., a discipline of particular interest) or result from other bibliometric exercises (e.g., subjects with strik-

ing growth patterns). A larger period is then subdivided into non-overlapping sub-periods. A structural analysis is conducted separately in these time slides; the clustering should preferably be based on hybrid methods.

2. Dynamic analysis of the discipline

In order to detect emergence, the evolution of the underlying discipline and its internal structures is monitored. Link relations are established between the sub-structures (clusters) of each time slide obtained from the first step.

3. Identification of emerging topics

This step forms the true core of the analysis. The change in the cluster structures over time is evaluated and new, emerging or substantially changing clusters are identified. Different methods for labelling the topic are in use. Two of these are described along with some results in the following section.

4. Delineation of the topic (optional)

If the previous step results in the identification of an emerging topic, it might happen that the hybrid cluster algorithm provides a document set that contains noise, or does not cover all relevant publications. In other words, the cluster might represent the emerging topic in a way, but not cover the topic in an adequate manner. In order to improve topic coverage, advanced bibliometrics-aided search strategies can be developed to reduce noise and to add other relevant documents to the topic cluster. This optional step might also become necessary if the time period underlying the study has to be extended from that used for the topic detection. An example for the application of such a search strategy to bioinformatics in the period 2006-2009 was given by **Glänzel et al.** (2009).

5. Network analysis of the topic (optional)

This might concern the internal structure of the topic, its links to other subjects (environment) or both. An example was given in an on-going project on 'entrepreneurship research' with the *University Sussex* and *University of Missouri-Kansas City (UMKC)* (cf. **Meyer et al.**, 2011).

6. Bibliometric study of the topic

In this final step a bibliometric analysis is conducted of the topic detected in step 3. The study might comprise the identification of main actors (countries, regions, institutions and/or individuals), patterns of international collaboration and the analysis of publication activity and citation impact. Examples have been given for bioinformatics (**Glänzel et al.**, 2009) and for selected topics in the sciences and social sciences (**Glänzel**; **Thijs**, 2011a).

The following section will report the complete process of detecting and analysing emerging research topics according to the above scheme using examples from four disciplines in the sciences and social sciences. The optional steps have been omitted in the present study.

Data sources and results

The underlying non-overlapping 5-year periods are 1999-2003 and 2004-2008. The emerging topics have been sought

within four selected subject categories according to *Thomson Reuters' Journal Citation Reports (JCR)*. In previous studies (Glänzel; Thijs, 2011b, 2012; Glänzel, 2012) a combined text- and citation-based hybrid approach was used to analyse the cognitive structure of these disciplines and to detect emerging topics. First similarity measures to be used by the cluster algorithm were derived from text-mining of titles, abstracts and keywords of documents (term frequencies) as well as from citation-based links (bibliographic coupling) between the documents assigned to the discipline under study (cf. Glänzel; Thijs, 2011b). For the combination of different similarities, the following solution was found. The similarity measure of each component (i.e., bibliographic coupling and term frequency, respectively) was determined by the cosine between the corresponding two documents in a vector space model. In particular, the hybrid (cosine) similarity measure r is then defined as the cosine of the linear combination of the underlying angles, i.e.,

$$r = \cos x = \cos(\sum \lambda_i x_i) = \cos \sum \lambda_i \arccos(\xi_i), \quad \lambda_i \in [0,1], \sum \lambda_i = 1,$$

where ξ_1 is the citation-based and ξ_2 the textual similarity, λ_i denotes the free parameter for the linear combination and $\arccos(\xi_i)$ denotes the underlying angles according to the vector space model. *Thomson Reuters' Web of Science* was used as the data source. The number of clusters found is shown in table 1.

JCR Subject Category	Number of clusters	
	1 st time slide	2 nd time slide
Public, environmental & occupational health	6	7
Environmental sciences	6	8
Energy & fuels	7	8
Biomedical engineering	8	9

Table 1. Number of clusters in four selected disciplines in two disjointed time slides [data sourced from *Thomson Reuters Web of Knowledge*]

Figure 1 shows a graphical visualisation of possible cluster structures and their links in two disjointed time slides, t_1 and t_2 . The dark circle in the second slide might indicate a potential emerging topic that is linked to related topics in both time slides. It might have emerged from a topic (the leftmost cluster of slide t_1 in figure 1) to which it is linked together with its adjacent cluster (the leftmost cluster of slide t_2). The question arises of how to decide if this cluster can indeed be considered 'emerging' or not. In order to answer this question, we first introduce the notion of 'core documents' and then some criteria of emergence.

Core documents have first been introduced in the context of co-citation analysis as most (co-)cited papers forming a core (cf. Small, 1973). The term 'core documents' has been re-introduced by Glänzel and Czerwon (1996) in the context of bibliographic coupling and defined as follows.

Definition: Core documents are papers that have at least $n > 0$ links of at least a given strength $r > 0$ according to the cosine measure based on bibliographic coupling.

This definition has been extended to hybrid similarities (Glänzel; Thijs, 2011) by applying the same criteria to hy-

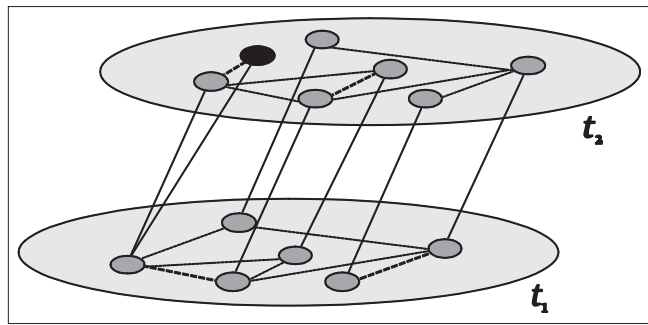


Figure 1. Sketch of a research field's changing topic structure over time (dotted lines represent internal structures, solid lines connect the time slides t_1 and t_2)

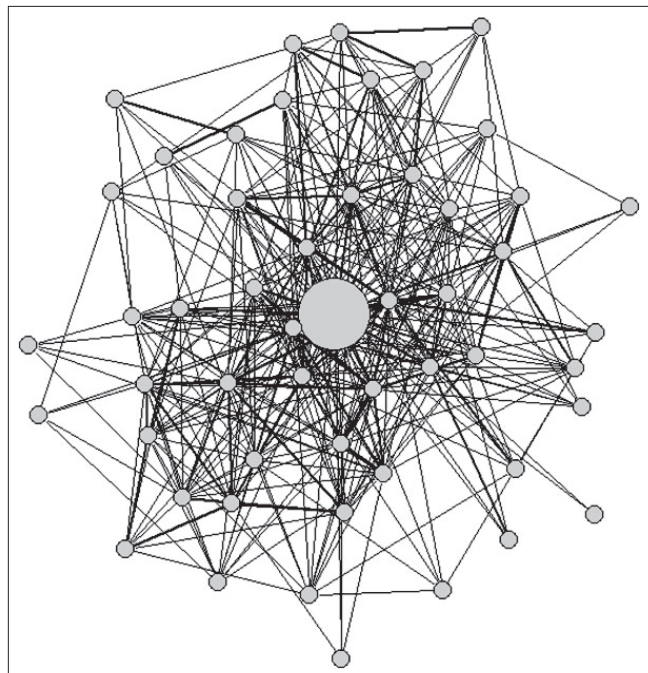


Figure 2. Visualisation of the link environment of a 'core document' [Data sourced from *Thomson Reuters Web of Knowledge*]

brid similarities. In verbal terms, core documents are papers that have strong (hybrid text-citation) links with many other papers. Core documents should ideally represent about 1% of the original document set (i.e., of the corresponding topic or discipline). The link environment of a typical core document is shown in figure 2.

Core documents have also been used to create links between clusters in the different time periods as visualised in figure 1. Citation links between core documents *in one period to all publications in the clusters of the other period* are used to determine links between the topic structures of the two time slides. This method helps reduce noise that might otherwise occur when cross-citation links are shown between all documents of the corresponding clusters. The noise is caused by less relevant documents located far from the medoids of the corresponding clusters (see Glänzel; Thijs, 2012). This part of the analysis forms the 3rd step according to the scheme in the previous section.

Before introducing the examples, I briefly summarise the main criteria for emerging research topics. First and foremost, emerging topics are expected to have already reached a certain critical mass, to form coherent clusters, and to still

have strong links to their “mother fields”. Then three types of clusters might be considered strong indications of emerging topics.

Type (1) Existing cluster with an exceptional growth with regard to the previous period,

Type (2) Completely new cluster with its root in other clusters in the previous period and

Type (3) Existing cluster with a strong topic shift in the new time period.

Once an emerging topic has been identified, the cognitive description of the topic can be based on the titles of the core documents and on the best keywords. Keywords alone are often not sufficient to provide an adequate description of the detected topic because of the common vocabulary. Four examples for emerging topics within the selected disciplines are shown in table 2.

JCR Subject Category	Emerging topic
Environmental sciences	nano-pollution
Public, environmental & occupational health	environmental factors
Energy & fuels	biofuels
Biomedical engineering	brain-machine-interface

Table 2. Examples for emerging topics in four selected subject categories [data sourced from *Thomson Reuters Web of Knowledge*]

Following the scheme proposed in the methodological section and skipping the two optional steps, a bibliometric analysis of the detected topics can be conducted. In the following, the most important results of a macro level analysis based on 16 selected countries and the European Union (15 members) as a supra-national region are summarised for the four emerging topics in table 2. Additional data and further details on these topics and the underlying subject categories can be found in the study by Glänzel and Thijs (2011a).

For the bibliometric analysis of the topics, the set of standard indicators introduced by Braun *et al.* (1985) has been used. Besides the Mean Observed Citation Rate (MOCR), its journal-based expectation (MECR) and the ratio of the two measures, the Relative Citation Rate (RCR), are applied. All citation indicators are based on three-year citation windows.

The first emerging topic, which was found in the JCR subject category ‘environmental sciences’, was labelled *nano-pollution*. The topic can be considered to be of Type (2) according to the above typology. It represents 12.6% (3533 documents) of the discipline. The 10 best TF-IDF (term frequency–inverse document frequency) keywords of this cluster are as follows (data sourced from *Thomson Reuters Web of Knowledge*).

pharmaceuticals; aquatic environment; porous media; constructed wetlands; tandem mass spectrometry; *escherichia coli*; mass spectrometry; light; performance liquid chromatography; sewage treatment plants

The keywords are quite general and do not really describe a coherent topic. The following five core documents, which

Country	ISO code	Papers	World share	MOCR	MECR	RCR
Belgium	BEL	50	1.4%	1.62	1.83	0.89
Brazil	BRA	57	1.6%	2.04	2.11	0.97
Denmark	DNK	53	1.5%	3.23	2.81	1.15
France	FRA	143	4.0%	2.35	2.39	0.98
Germany	DEU	229	6.5%	2.68	2.36	1.14
Greece	GRC	35	1.0%	1.17	1.90	0.62
India	IND	186	5.3%	1.33	1.69	0.79
Italy	ITA	143	4.0%	2.10	2.12	0.99
Japan	JPN	144	4.1%	1.71	2.26	0.76
Netherlands	NLD	63	1.8%	2.29	2.22	1.03
China	CHN	368	10.4%	2.15	2.35	0.92
Poland	POL	91	2.6%	0.69	0.98	0.71
Spain	ESP	178	5.0%	2.70	2.41	1.12
Sweden	SWE	63	1.8%	2.14	2.05	1.05
UK	GBR	193	5.5%	3.39	2.57	1.32
USA	USA	930	26.3%	2.81	2.58	1.09
EUR15	EUR	1109	31.4%	2.45	2.31	1.06

Table 3. Bibliometric macro data on the emerging topic ‘nano-pollution’ [Data sourced from *Thomson Reuters Web of Knowledge*]

were taken from the same source, provide a more specific description.

- On colloid retention in saturated porous media in the presence of energy barriers: The failure of alpha and opportunities to predict eta
- The significance of heterogeneity on mass flux from Dnapi source zones: An experimental investigation
- C-60 colloid formation in aqueous systems: Effects of preparation method on size, structure, and surface charge
- Individual and mixture effects of selected pharmaceuticals and personal care products on the marine phytoplankton species *Dunaliella tertiolecta*
- Nanomaterials as possible contaminants: the fullerene example

The topic’s mean observed citation rate amounts to 2.18. China is the second largest contributor behind the US. The citation impact of the US and EU is distinctly above this reference standard. The relatively high impact of China’s publications in this emerging topic is worth mentioning. China’s citation impact exceeds that of Japan, and is very close to the reference standard.

International collaboration in this emerging topic is quite intense. The map of international co-publications is presented in figure 3. For the visualisation Pajek (Kamada-Kawai layout) was used (Batagelj; Mrvar, 2003). The thickness of lines connecting country pairs is based on the strength of their co-publication links. This strength is measured by the Salton Index (i.e., the cosine similarity of countries based on their joint co-publications and their total publication output). In the map, the USA forms the most important node in the collaboration network and connects the European cluster with China and Japan.

The second emerging topic was found in the discipline ‘public, environmental & occupational health’. This topic, again,

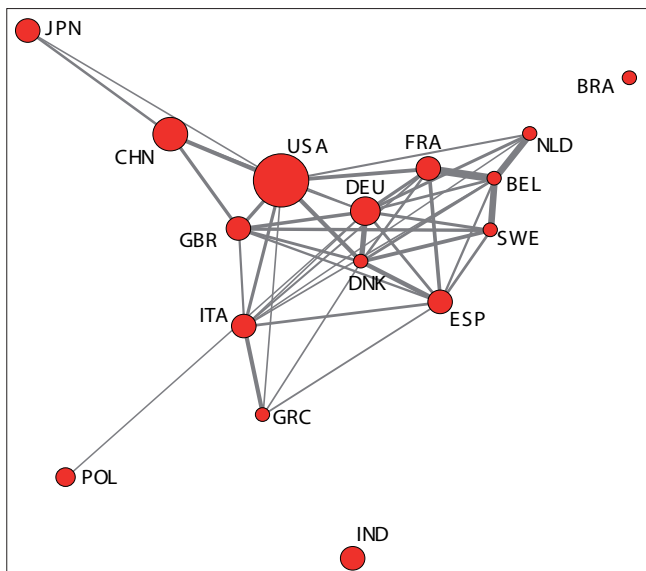


Figure 3. International collaboration in the emerging topic ‘nano pollution’ [Data sourced from Thomson Reuters Web of Knowledge, figure sourced from Glänzel and Thijs, 2011a]

represents Type (2) according to our typology. It represents 20.5% (5945 documents) of the discipline. The 10 best TF-IDF keywords of this cluster are as follows (data sourced from Thomson Reuters Web of Knowledge).

malaria; pcr; outbreak; identification; tests; antibodies; plasmodium falciparum; resistance; regression; infections

The following five core documents taken from the same source provide complementary information.

- Mortality in 13 French cities during the August 2003 heat wave
- Ambient carbon monoxide may influence heart rate variability in subjects with coronary artery disease

Country	ISO code	Papers	World share	MOCR	MECR	RCR
Belgium	BEL	106	1.8%	4.27	3.39	1.26
Brazil	BRA	282	4.7%	2.20	2.49	0.88
Denmark	DNK	107	1.8%	5.58	3.82	1.46
France	FRA	344	5.8%	4.07	3.51	1.16
Germany	DEU	265	4.5%	4.69	3.56	1.32
Greece	GRC	50	0.8%	6.14	3.74	1.64
India	IND	148	2.5%	2.20	2.71	0.81
Italy	ITA	209	3.5%	3.55	3.26	1.09
Japan	JPN	186	3.1%	3.09	3.16	0.98
Netherlands	NLD	245	4.1%	4.44	4.00	1.11
China	CHN	193	3.2%	2.85	3.19	0.89
Poland	POL	53	0.9%	2.96	3.01	0.98
Spain	ESP	193	3.2%	3.28	2.92	1.13
Sweden	SWE	163	2.7%	4.11	3.45	1.19
UK	GBR	697	11.7%	4.65	3.70	1.26
USA	USA	2279	38.3%	4.89	4.09	1.19
EUR15	EUR	2126	35.8%	4.01	3.47	1.16

Table 4. Bibliometric macro data on the emerging topic ‘environmental factors’ [Data sourced from Thomson Reuters Web of Knowledge]

- Temperature and mortality among the elderly in the United States - A comparison of epidemiologic methods
- Effects of air pollution on heart rate variability: The VA Normative Aging Study
- Association of air pollution with increased incidence of ventricular tachyarrhythmias recorded by implanted cardioverter defibrillators

While the frequency of keywords points to epidemics and infections, core documents rather highlight different aspects, namely ‘heat wave’ and ‘air pollution’. That is why the topic was labelled more generally *environmental factors*.

Table 4 shows the bibliometric indicators for the most active countries. The US is the most important contributor although the EU has a similar large share in this emerging topic. UK ranks second behind the US in terms of activity, followed by France and Brazil. This is remarkable since China’s contribution is low although SARS (severe acute respiratory syndrome) was one of the issues studied in the literature of this cluster. The citation impact of the emerging topic ‘environmental factors’ amounts to 3.82. The high standard of the US and most members of the EU is not unexpected. The citation impact of Brazil and the Asian countries remains distinctly below the reference standard of the complete topic.

‘Environmental factors’ is apparently a global issue. The collaboration network is shown in figure 4. There is an intense international collaboration and the network has two important nodes, namely the US (globally) and the UK (locally) within Europe. Most countries have a high share of internationally co-authored papers. This ranges between 60% (and even higher for Greece, Belgium and Denmark) and somewhat less than 25% (India) in the countries under study. The median of 51.2% for this topic is pronouncedly high.

The third topic refers to the discipline ‘energy & fuels’ and was labelled *biofuel*. The topic represents Type (1) according to the above typology. It represents 24.2% (7059 papers) of the discipline. The 10 best TF-IDF keywords of this cluster

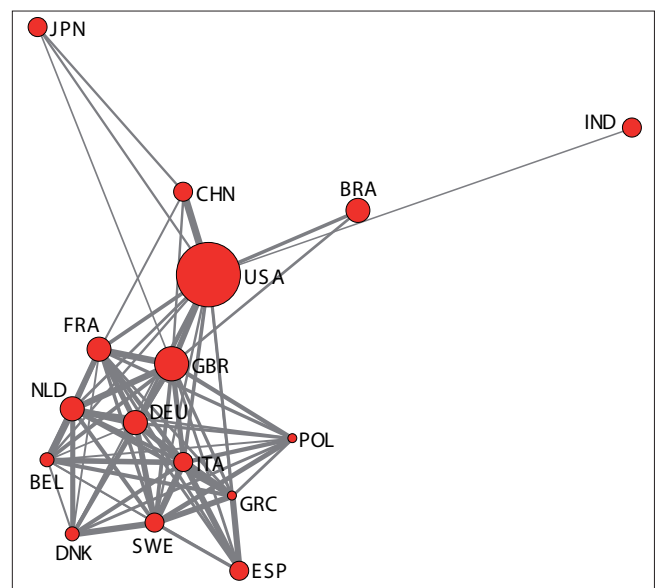


Figure 4. International collaboration in the emerging topic ‘environmental factors’ [Data sourced from Thomson Reuters Web of Knowledge, figure sourced from Glänzel and Thijs, 2011a]

are as follows (data sourced from *Thomson Reuters Web of Knowledge*).

ammonia; biogas; start up; anaerobic digestion; recovery; mechanical properties; carbon dioxide; *saccharomyces cerevisiae*; biodegradation; synthesis gas

The following five core documents were taken from the same source.

- Process optimization for biodiesel production from mahua (*Madhuca indica*) oil using response surface methodology
- Continuous production of biodiesel via transesterification from vegetable oils in supercritical methanol
- Temperature effects on biohydrogen production in a granular sludge bed induced by activated carbon carriers
- Biohydrogen generation from jackfruit peel using anaerobic contact filter
- Biohydrogen-production from beer lees biomass by cow dung compost

Country	ISO code	Papers	World share	MOCR	MECR	RCR
Belgium	BEL	50	0.7%	4.14	3.82	1.08
Brazil	BRA	140	2.0%	3.71	3.76	0.99
Denmark	DNK	75	1.1%	10.01	3.44	2.91
France	FRA	328	4.6%	3.80	3.80	1.00
Germany	DEU	326	4.6%	4.63	3.80	1.22
Greece	GRC	120	1.7%	3.59	3.13	1.15
India	IND	421	6.0%	4.53	3.88	1.17
Italy	ITA	223	3.2%	3.89	3.65	1.06
Japan	JPN	499	7.1%	4.29	4.14	1.04
Netherlands	NLD	94	1.3%	4.65	3.89	1.19
China	CHN	1218	17.3%	4.12	3.78	1.09
Poland	POL	97	1.4%	3.66	3.66	1.00
Spain	ESP	262	3.7%	3.62	3.70	0.98
Sweden	SWE	93	1.3%	3.66	3.64	1.00
UK	GBR	307	4.3%	3.78	3.09	1.22
USA	USA	1036	14.7%	4.11	3.60	1.14
EUR15	EUR	1806	25.6%	4.02	3.58	1.12

Table 5. Bibliometric macro data on the emerging topic 'biofuel' [Data sourced from *Thomson Reuters Web of Knowledge*]

The scientometric indicators for this topic are presented in table 5. Unlike in the previous case, we find some unexpected patterns. The share of the EU in the world total is relatively low (about 25%) and that of the US (less than 15%) is markedly low. The US ranks second after China, which has become the most active country in this topic. Japan and India rank third and fourth, respectively, behind China and the US. According to the expectation, the observed citation impact of the EU and the US clearly exceeds the reference standard of 3.79. The citation impact of the individual members of the European Union reflects a quite differentiated picture with Denmark, Germany and the Netherlands at the high end and Greece, Italy, Spain and France with indicator values somewhat above or even below the world standard. Japan, China and India are very active in this subject and

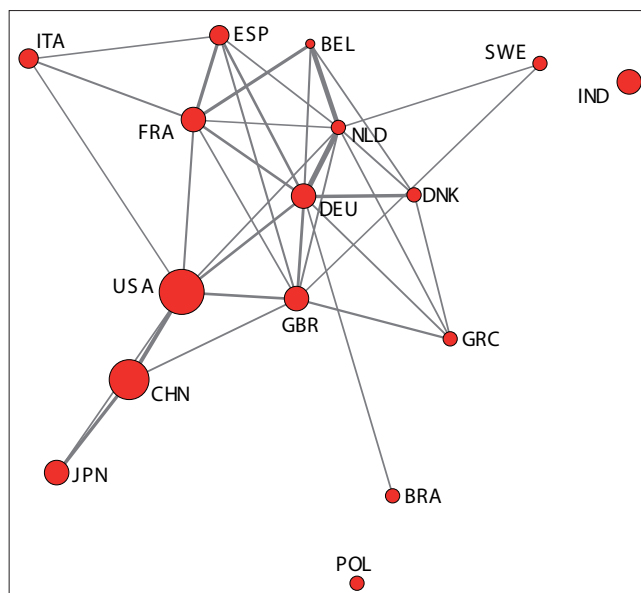


Figure 5. International collaboration in the emerging topic 'biofuel' [Data sourced from *Thomson Reuters Web of Knowledge*, figure sourced from Glänzel and Thijs, 2011a]

their research proved efficient as well. They outperform even several European countries in terms of citation impact. Biofuel-related research has apparently become a strategic subject in these countries.

International collaboration is less intense than in the previous cases (cf. figure 5). Only Belgium has more than 50% internationally co-authored papers in its publication output. The network is less dense than was, for instance, that of 'nano pollution'. The European cluster is also less coherent, more scattered and forms a kind of agglomerate of local clusters. The USA forms the bridge between the Far East and Europe.

The last emerging topic, which was found in the subject category 'biomedical engineering', was labelled *brain-machine interface*. The topic is of Type (1) according to the classification in the methodological section. It covers 19.4% (5632 documents) of all papers in the discipline. The 10 best TF-IDF keywords of this cluster are as follows (data sourced from *Thomson Reuters Web of Knowledge*).

signals; blood flow; classification; eeg; neural networks; pattern recognition; independent component analysis; hemodynamics; patterns; elasticity

Unfortunately, these keywords are not very specific. However, the core documents help specify the coverage of the topic. Five of them, taken from the same source as above, are presented below.

- Virtual keyboard controlled by spontaneous EEG activity
- Planar gradiometer for magnetic induction tomography (MIT): theoretical and experimental sensitivity maps for a low-contrast phantom
- Adaptive BCI based on variational Bayesian Kalman filtering: An empirical evaluation
- Model-based neural decoding of reaching movements: A maximum likelihood approach
- Ascertaining the importance of neurons to develop better brain-machine interfaces

Country	ISO code	Papers	World share	MOCR	MECR	RCR
Belgium	BEL	83	1.5%	2.96	2.67	1.11
Brazil	BRA	95	1.7%	2.00	2.68	0.75
Denmark	DNK	94	1.7%	2.51	2.98	0.84
France	FRA	241	4.3%	2.95	2.95	1.00
Germany	DEU	338	6.0%	2.69	2.67	1.01
Greece	GRC	98	1.7%	2.19	2.72	0.81
India	IND	48	0.9%	1.98	2.23	0.89
Italy	ITA	296	5.3%	3.05	2.88	1.06
Japan	JPN	221	3.9%	2.38	2.95	0.80
Netherlands	NLD	196	3.5%	3.53	3.17	1.11
China	CHN	256	4.5%	2.01	2.68	0.75
Poland	POL	71	1.3%	1.82	2.34	0.78
Spain	ESP	207	3.7%	3.13	2.91	1.08
Sweden	SWE	119	2.1%	2.63	2.83	0.93
UK	GBR	546	9.7%	3.00	3.04	0.98
USA	USA	1990	35.3%	3.09	3.36	0.92
EUR15	EUR	2227	39.5%	2.84	2.88	0.98

Table 6. Bibliometric macro data on the emerging topic ‘brain-machine interface’ [Data sourced from Thomson Reuters Web of Knowledge]

The scientometric indicators for the selected countries are shown in table 6. Nearly 40% of all papers have an author or co-author with an address in the EU. The US has a somewhat lower share in the world total. Within the European Union, we find by and large the usual balance of productivity observed in other disciplines as well. UK, Germany, France and Italy are the most active EU members, although in this case Italy is more active than France. The reference standard of citation impact in ‘brain-machine interface’ amounts to 2.67. As in the previous case, the observed citation impact of the EU and the US is above the reference standard and the US impact exceeds that of the EU. Interestingly, Spain, the Netherlands and Italy attract, on an average, more citations than Denmark, the UK and Germany.

International collaboration is relatively intense and somewhat global as well (cf. figure 6). The network differs from the previous ones as there is no distinct EU cluster. Europe, the US, Asia and Brazil form one large cluster. Only India remains somewhat isolated.

Conclusions

Based on the above review of the structural analysis of scientific disciplines, the following conclusions can be drawn. The hybrid approach proved successful even in a local environment, where the same or similar vocabulary is used by the scientific community. The option of using the weights of the citation-based and the textual component for fine-tuning according to the different citation cultures in the fundamental, applied and social sciences and the humanities proved extremely useful. The useful feature of core documents to represent important nodes in the hybrid citation/lexical network provided essential information about possible emergence through their cross-citations links over different time periods.

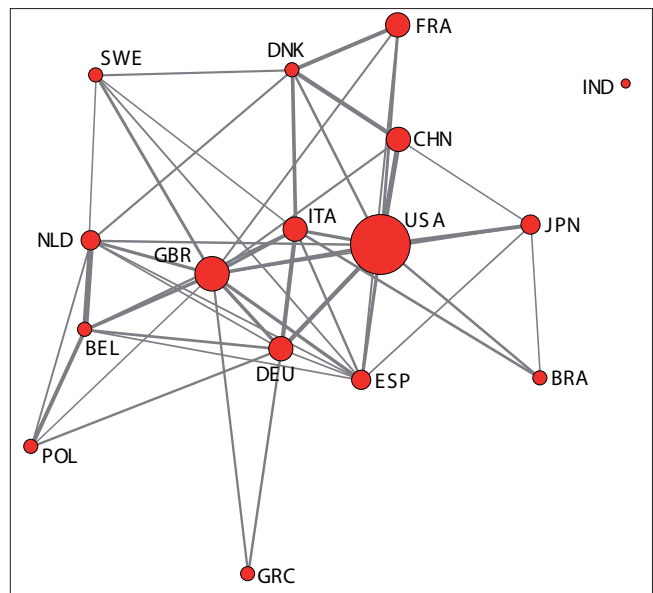


Figure 6. International collaboration in the emerging topic ‘brain-machine interface’ [Data sourced from Thomson Reuters Web of Knowledge, figure sourced from Glänzel and Thijs, 2011a]

Publication activity, citation impact and international collaboration reflect important characteristics of emerging topics. Several topics might require intense international collaboration like the ‘environmental factors’ in public health. In this topic regional aspects play an important part as well. The environmental factors might, for instance, differ in individual world regions but the fundamental phenomenon remains a global one. Other topics, like ‘brain-machine interface’ might be truly global issues so that particular sub-clusters or geographical polarisation could not be found. The loose networks in ‘biofuel’ shows that national issues might be in the foreground here. The differentiation in industry between sugarcane in Brazil, corn in the US and rice straw in Japan might illustrate the regional policies. Finally, the essential contribution of the emerging economies in Asia and South America to ‘nano pollution’ and ‘biofuel’ topics of research and their high citation impact is worth mentioning. Research in emerging topics is no longer a matter of the traditionally leading countries alone.

Acknowledgements

Methodology has partially been developed in the context of the Eracep project within the Coordination and Support Actions (CSAs) of the ERC work programme. The author wishes to acknowledge this support.

Figures 1 and 2 are reproduced from Glänzel and Thijs (2012) with permission of the publisher.

References

Batagelj, Vladimir; Mrvar, Andrej. "Pajek – analysis and visualization of large networks". In: M. Jünger, P. Mutzel (Eds.), *Graph drawing software*. Springer, 2003, pp. 77–103.

Bichteler, Julie; Eaton, Edward A. "The combined use of bibliographic coupling and co-citation for document-retrieval". *Jasis*, 1980, v. 31, n. 4, pp. 278–282. <http://dx.doi.org/10.1002/asi.4630310408>

- Boyack, Kevin W.; Klavans, Richard.** "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?" *Jasist*, 2010, v. 61, n. 12, pp. 2389–2404.
<http://dx.doi.org/10.1002/asi.21419>
- Braam, Robert R.; Moed, Henk F.; Van Raan, Anthony F. J.** "Mapping of science by combined co-citation and word analysis, Part 1: Structural aspects". *Jasis*, 1991a, v. 42, n. 4, pp. 233–251.
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<233::AID-ASI1>3.0.CO;2-I](http://dx.doi.org/10.1002/(SICI)1097-4571(199105)42:4<233::AID-ASI1>3.0.CO;2-I)
- Braam, Robert R.; Moed, Henk F.; Van Raan, Anthony F. J.** "Mapping of science by combined co-citation and word analysis, Part II: Dynamical aspects". *Jasis*, 1991b, v. 42, n. 4, pp. 252–266.
[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<252::AID-ASI2>3.0.CO;2-G](http://dx.doi.org/10.1002/(SICI)1097-4571(199105)42:4<252::AID-ASI2>3.0.CO;2-G)
- Braun, Tibor; Glänzel, Wolfgang; Schubert, Andras.** *Scientometric indicators. A 32 country comparison of publication productivity and citation impact*. World Scientific Publishing, Singapore – Philadelphia, 1985.
- Callon, Michel; Courtial, Jean P.; Turner, William A.; Bauin, Serge.** "From translations to problematic networks. An introduction to co-word analysis". *Social science information*, 1983, v. 22, n. 2, pp. 191–235.
<http://dx.doi.org/10.1177/053901883022002003>
- Garfield, Eugene.** "Launching the ISI Atlas of Science – for the new-year, a new generation of reviews". *Current contents*, 1987, v. 1, pp. 3–8.
- Glänzel, Wolfgang; Czerwon, Hans-Jürgen.** "A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level". *Scientometrics*, 1996, v. 37, n. 2, pp. 195–221.
<http://dx.doi.org/10.1007/BF02093621>
- Glänzel, Wolfgang.** "On the h-index. A mathematical approach to a new measure of publication activity and citation impact". *Scientometrics*, 2006, v. 67, n. 2, pp. 315–321.
<http://dx.doi.org/10.1007/s11192-006-0102-4>
- Glänzel, Wolfgang; Janssens, Frizo; Thijs, Bart.** "A comparative analysis of publication activity and citation impact based on the core literature in bioinformatics". *Scientometrics*, 2009, v. 79, n. 1, pp. 109–129.
<http://dx.doi.org/10.1007/s11192-009-0407-1>
- Glänzel, Wolfgang; Thijs, Bart.** "Research in emerging fields: who takes the lead?". *ISSI Newsletter*, 2011a, v. 7, n. 4, pp. 85–95.
- Glänzel, Wolfgang; Thijs, Bart.** "Using 'core documents' for the representation of clusters and topics". *Scientometrics*, 2011b, v. 88, n. 1, pp. 297–309.
<http://dx.doi.org/10.1007/s11192-011-0347-4>
- Glänzel, Wolfgang; Thijs, Bart.** "Using 'core documents' for detecting and labelling new emerging topics". *Scientometrics*, 2012, in press.
<http://dx.doi.org/10.1007/s11192-011-0591-7>
- Glänzel, Wolfgang.** "The role of core documents in bibliometric network analysis and their relation with h-type indices". *Scientometrics*, 2012, in press.
<http://dx.doi.org/10.1007/s11192-012-0639-3>
- Hirsch, Jorge E.** "An index to quantify an individual's scientific research output". *Procs. of the National Academy of Sciences of the United States of America*, 2005, v. 102, n. 46, pp. 16569–16572.
<http://dx.doi.org/10.1073/pnas.0507655102>
- Janssens, Frizo.** *Clustering of scientific fields by integrating text mining and bibliometrics*, Ph. D. Thesis, Faculty of Engineering, Katholieke Universiteit Leuven, 2007.
- Janssens, Frizo; Glänzel, Wolfgang; De Moor, Bart.** "A hybrid mapping of information science". *Scientometrics*, 2008, v. 75, n. 3, pp. 607–631.
<http://dx.doi.org/10.1007/s11192-007-2002-7>
- Jarneving, Bo.** *The combined application of bibliographic coupling and the complete link cluster method in bibliometric science mapping*. PhD Thesis, University College of Borås/Göteborg University, 2005.
- King, Jean.** "A review of bibliometric and other science indicators and their role in research evaluation". *Journal of information science*, 1987, v. 13, n. 5, pp. 261–276.
<http://dx.doi.org/10.1177/016555158701300501>
- Meyer, Martin; Libaers, Dirk; Thijs, Bart; Glänzel, Wolfgang.** "The emergence of entrepreneurship as a research field". In: E. Noyons, P. Ngulube, J. Leta (Eds), *Procs. of ISSI 2011 - The 13th Intl conf on scientometrics and informetrics*, Durban, South Africa, 2011, pp. 552–557.
- Small, Henry.** "Co-citation in scientific literature — New measure of relationship between 2 documents". *Jasis*, 1973, v. 24, n. 4, pp. 265–269.
<http://dx.doi.org/10.1002/asi.4630240406>
- Zitt, Michel; Bassecouard, Elise.** "Development of a method for detection and trend analysis of research fronts built by lexical or co-citation analysis". *Scientometrics*, 1994, v. 30, n. 1, pp. 333–351.
<http://dx.doi.org/10.1007/BF02017232>