ERACEP – Emerging Research Areas and their Coverage by ERC-supported Projects

Grant Agreement No. 240777 – ERACEP

Final Report

APRIL 2013

European Research Council

Fraunhofer ISI
Karlsruhe (Germany)

Thomas Reiss
Etienne Vignola-Gagné
Piret Kukk

Breslauer Straße 48
D-76139 Karlsruhe
Germany

ECOOM
KU Leuven (Belgium)

Wolfgang Glänzel
Bart Thijs

Waaistraat 6
B-3000 Leuven
Belgium

Executive Summary

THE ERC MISSION AND THE ROLE OF EMERGING RESEARCH TOPICS

The mission of the ERC is to encourage the highest quality research in Europe and to support frontier research across all fields of research on the basis of scientific excellence. This mission implies that ERC funding and operation should be able to address frontier research and emerging research topics. A main challenge for the implementation of this mission is that suitable assessment, evaluation, and monitoring frameworks of truly creative research need to be available. Accordingly, the ERC also started to build up a portfolio of projects and studies, aiming at providing intelligence for supporting its operational and strategic activities. The ERACEP (Emerging Research Areas and Their Coverage by ERC-Supported Projects) support action is one of these activities. It was carried out by a team of two partner institutions, namely the Fraunhofer Institute for Systems and Innovation Research (Karlsruhe, Germany) as coordinator of ERACEP and ECOOM of the Katholieke Universiteit Leuven (Belgium) (between October 1st, 2009 and February 28, 2013.

The main objectives of ERACEP are the identification of topically emerging research areas and the analysis to what extent activities supported by the ERC cover and contribute to these research areas.

THE ERACEP APPROACH

The ERACEP approach is based on the combination of two perspectives. The first perspective concerns the identification of topically emerging research areas, assuming that excellent truly creative research is particularly concerned with such emerging areas. The second perspective takes the view of ERC funding and explores how ERC-funded research themes map to the identified emerging areas.

For the translation of these two perspectives into a research agenda new methodological approaches had to be developed. For detecting emerging research topics a number of mainly expert-based foresight methods have been suggested and applied in different contexts. Frequently, standard future topics are identified by such methods. ERACEP adopted a bibliometric approach for identifying emerging topics starting with an open investigation in all fields of science, social sciences, arts and humanities. A bibliometric approach has the specific advantage that it refers to

objective data in contrast to expert-based approaches which inevitably imply some subjectivity. The open investigation implemented within ERACEP reflects as much as possible the open, flexible, bottom-up approach of the ERC funding schemes.

Also the implementation of the second perspective of ERACEP, the matching of ERC-funded activities to emerging research topics, required the development of new methods since no suitable and validated methodology was available at the start of ERACEP.

Accordingly, due to the inherent nature of its main objectives, ERACEP has an explorative and experimental character. This also implies that during the advancement of research within ERACEP the work plan had to be adapted to accommodate for new empirical findings. An important modification in this line was the decision to narrow-down the scope of the analysis by focusing more in-depth on the evolution of certain clusters of emerging topics instead of a broad mapping distributed over all areas of science.

THE IDENTIFICATION OF EMERGING RESEARCH TOPICS

The bibliometric approach for the detection of emerging research topics developed within ERACEP consists of two building blocks. The first building block comprises the identification of dynamic research fields based on publication activities. The rationale behind this approach is the notion that dynamic growth –in terms of publications- of a specific field reflects on the one hand an increasing interest of scientists in this field, so that more research groups are doing research in the respective area. On the other hand, dynamics could also indicate that a constant number of scientists are increasing their research activities. The second building block designed for the detection of emerging topics within dynamic fields consists of three bibliometric components: a cluster analysis on the fields, a fine-grained representation of clusters based on core documents, and a diachronic analysis of the evolution of links among clusters and topics over time.

For the identification of the (most) dynamic fields a Sharpe Ratio was calculated for all the subject categories included in the Science Citation Index Expanded, the Social Sciences Citation Index, and the Arts and Humanities Citation Index of Thomson Reuters. The Sharpe Ratio considers the development of a specific field in relation to the growth of all fields. Thereby, it adjusts for differences in absolute field size. As a result of this dynamic analysis thirteen fields were identified from the sciences, four from the social sciences, and three from the arts and humanities. Among the thirteen most dynamic fields from the sciences five fields can be categorised within the area of medicine. These include “nursing”, “orthopaedics”, “obstetrics and gynaecology”, “medical ethics”, and “oncology”. While “orthopaedics”, “obstetrics, gynaecology”, and “oncology” are closely related and concerned with the clinical and therapeutical dimension of medicine, “nursing” and “medical ethics” represent healthcare and social dimension. Three of the dynamic fields belong to the engineering domain. They cover rather small and specialised fields like “robotics” and “characterisation

and testing of materials" as well as the larger field of "biomedical engineering". Two fields represent the biological sciences. The first one concerns "biotechnology and applied microbiology", the second one deals with "behavioural sciences". The three remaining fields from the sciences can be considered as stand-alone fields. They include "operations research and management sciences", "environmental sciences", and "energy and fuels". Within the four most dynamic fields of the social sciences two also belong to the medicine domain. These are "public, environmental and occupational health", and "experimental psychology". The two remaining fields are "transportation" and "geography". Within the arts and humanities three quite different dynamic fields have been detected: "archaeology", "religion", and "architecture".

Overall, the dynamic analysis using the Sharpe Ratio proved to be a powerful indicator of growth and an appropriate criterion for the selection of the most dynamic fields, resulting in a portfolio of quite diverse dynamic fields covering all categories of science.

For the first component of the detection of the emerging topics within the retained sample of dynamic fields, the cluster analysis, a hybrid approach was applied which forms a combination of two traditional cluster techniques: a citation-based approach and an approach based on textual links. The citation-based component uses bibliographic coupling, which has clear advantages compared to co-citation analyses, in particular for the identification of emerging topics. The most important advantage is that all papers have references which can be used for bibliographic coupling and that no response time is needed for citing literature which is crucial for a citation-based approach. The textual component is based on term frequencies, where terms extracted from titles, abstracts, and publication keywords are analysed. Both components are combined to a hybrid measure of similarity between documents which forms the basis for clustering. For the labelling and representation of clusters, core documents are used. These are defined as documents which are strongly linked – i.e. referred to - by many other documents, and thus represent the most interconnected part of the network, thereby also representing the main (or central) content of the network. The third component of the methodology comprised a diachronic analysis of the links between clusters over time. Cluster analyses were carried out for two not overlapping time periods. To determine links between the structures of the two periods, citation links between core documents in one period i.e. 2004–2008 to all publications in the clusters of the other period i.e. 1999–2003 were used.

By this fine-mapping analysis three cases of cluster evolution were identified which can indicate new emerging topics. The first case are existing clusters with an exceptional growth over time. The second case are completely new clusters with roots in other clusters of a previous period. The third type of emerging topics represents existing clusters with a topic shift. The fine-mapping methodology was applied to six selected dynamic fields and resulted in the detection of ten emerging topics. Within the dynamic field "biomedical engineering" two emerging topics are detected. The first topic is labelled "brain-computer interface" and represents the first type of emerging topics, namely a topic which is already present in the first period but is characterised

by an extensive growth over time. The second topic is a completely new topic labelled "kinematics". It emerged with links to the clusters "imaging", "cartilage" and "bone cement" and is concerned mainly with joint and muscle kinematics during motion and corresponding models.

Within the dynamic field "energy and fuels" again two emerging topics are detected. The first one is labelled "biodiesel" and related to the topic about diesel and other petrol-derived fuels from the first period. The topic is characterised by an extensive growth and also by a topic shift towards non-petrol fuels. The second emerging topic is a new type of topic labelled "fuel cells". The seed for this topic lies with the topics solar cells and renewable energy from the first period.

When analysing emerging topics within the dynamic field of environmental sciences, it is important to consider that the whole field is characterised by an extensive growth over the last decade. The first of the two detected emerging topics is labelled "radiation". It represents a completely new topic with no close links to any of the topics from the first period. The second topic is named "nano pollution" and emerged from the topic on water from the first period. This topic is not only characterised by a topic shift towards nano materials but also by a considerable growth.

Within the dynamic field "obstetrics and gynaecology", one emerging topic labelled as "prenatal diagnosis" was detected. The topic is already present in the first period but experienced a clear topic shift over time. In particular, it is influenced by technological advancements in three-dimensional ultrasound imaging technologies, which allow to focus on new diagnostical approaches using, for example, fetal bone images.

Two emerging topics could be identified within the dynamic field "public, environmental and occupational health". The first topic, "environmental issues", deals with specific health issues related to the environment or location where people live. The second topic is labelled "AIDS in Africa". It emerged from HIV research present in several topics in the first period considered. But the focus of this topic is completely different. It is mainly devoted to social, socio-political and regional aspects related to the HIV infections with a regional focus on AIDS in Africa.

Finally, within the dynamic field "geography" one emerging topic labelled "state and region" was detected. It arose via a combination of the topics on "nationality and region" from the first period. However, it is not just a merger of the two topics but a real shift in content can be observed. The focus now is on the region as defining identity in a globalised world.

An important element of the ERACEP approach was not only to rely on advanced bibliometric and text-mining methods for the identification of emerging topics, but also to combine this view with a qualitative expert-based view. Accordingly, qualitative assessments of the fields and topics were carried out via interviews with experts from the respective disciplines. All in all, a considerable share of the emerging topics as identified by the different bibliometric analyses could be confirmed by this expert validation. However, also examples were identified where the bibliometric methodology alone seems not to be sufficient for identifying the most emerging topics within certain disciplines. We observe an interesting and notable correlation

between validation of fields and the ISI subject categories. Most discrepancies are present in the Arts and Humanities where experts had most difficulties to confirm the bibliometric results. A main reason for this deviation is that in fields like "architecture" and "religion" unlike in natural sciences a substantial amount of articles is published in non-English journals that could not be included in the bibliometric analyses. In addition, many regions in Europe have traditionally quite different research focuses in the dynamic areas identified via bibliometric analyses. This reflects, for example, geographical and cultural influences on fields like "architecture" and certainly "archaeology".

Taken together, the bibliometric cluster analysis has proven to be a powerful methodology for identifying emerging fields of science. However, it also became clear that there is a need to complement the quantitative statistical analysis by qualitative expert assessment in order to obtain robust and validated results.

INTERNATIONAL POSITIONING

ERACEP created a large set of publications reflecting emerging research topics worldwide. The availability of such a dataset allows to carry out additional country-specific analyses which had not been considered as feasible at the beginning of ERACEP. In particular, the following questions related to an international positioning of research activities were tackled:

- Which countries are the most active players in the emerging topic?
- Do papers from European scientists differ from American papers?
- To what extent do countries active in these topics collaborate with each other?

The following examples illustrate how the results of ERACEP can be used to put the emerging topics and trends observed into a geo-political context.

In general, we observe substantial differences with respect to national contributions and international collaborations amongst the distinct dynamic fields and emerging topics. The USA is contributing to about 60 per cent of all papers in the emerging topic labelled "AIDS in Africa", while its overall share in the field "energy and fuels" is just 17.6 per cent with for the emerging topics roughly 15 per cent for "biodiesel" and 22 per cent for "fuel cells". An opposite relationship can be detected for the People's Republic of China. Although Africa is of high economic importance to China, they contribute only to 2.5 per cent of all papers concerned with "AIDS in Africa", while their share in papers on "biodiesel" (17.3%) exceeds the share of the United States. Also, evidence show that the US take a clear lead in topics concerned with medical issues, while this leadership is challenged in fields related to environmental issues and also to the environmental dimension in public health.

The collaboration analysis is based on the construction of networks for dynamic fields as well as for specific emerging topics. We observe a dense network in environmental sciences where almost all European countries have strong cooperation

links with each other. We also observe a remarkable peripheral position of the United States with respect to the European environmental network. A similar observation can be made in the emerging topic "environmental issues within public health". A completely different cooperation pattern is obtained in the emerging topic "brain-computer interface" where the US plays a central role. The field "energy and fuels" represents an interesting cooperation structure in that we observe a loose cooperation network. This might be explained by the fact that countries are well aware of technological and economic potential that can be realised from science developments in these fields. Therefore, countries' degree of openness change according to emerging fields. In the previous example, this could reflect a stronger economic interest in "energy and fuels" as compared to the medical area where the effort to contribute to global health might dominate.

MAPPING ERC-FUNDED RESEARCH TO EMERGING RESEARCH TOPICS

For mapping ERC-funded research to the emerging research topics the balance between false-positive and false-negative hits plays a crucial role for the applicability of any method. As it is the aim to investigate to what extent the ERC-supported activities cover and contribute to the detected emerging topics, the avoidance of false-negative hits is more crucial to the procedure. Accordingly, our choice of methodological approach was accommodated to this requirement. As a data sample for the method a set of 932 applications to the 2009 Starting Grant Call of the ERC could be used. For the mapping to emerging topics the same set of application data was used as for the identification of emerging topics via the hybrid clustering approaches. All in all, a set of 164,220 unique documents was used for the matching procedure.

A full-text-matching approach was chosen. As a first step a database of publications and a database of applications are set up in parallel. The individual datasets are processed, so that for both, the publications on the one hand and the applications on the other hand, text fields are obtained. In a next step these fields are indexed in both datasets using the LUCENE text index. From both indexes a set of common terms is extracted. Document-by-term matrices are created containing the raw frequency of each term in the document. After applying a weight to term frequencies, the matrices are combined into a paper-by-application similarity matrix. Taking the average similarity over papers within a certain topic results in the final topic-to-application similarity matrix.

For the validation of the results of the mapping a manual qualitative procedure is used. The aim of the validation is firstly to estimate the appropriateness of the developed mapping methodology and make adjustments if necessary, secondly to justify decisions upon cut-off thresholds, and thirdly to support the removal of false-positive applications. The main approach towards validation was analysing the content of the proposals manually and comparing their content to the matched fields.

As a result of the matching, 885 applications could be used and all of them shared at least one term with the documents in the publication set. Accordingly, no document was omitted due to restrictions of terms to the set of common terms. After the calculation of average similarities between proposals and topics and the assignment based on thresholds, 289 applications were mapped with at least one topic, 173 applications had multiple assignments.

One of the objectives of ERACEP was to explore to which extent ERC funding procedures are able to address and contribute to emerging topics. The results of our matching approach, which was carried out for a selected number of thematic fields and emerging topics, indicate that ERC funding indeed is able to address emerging topics. However, we observe substantial differences across different topics.

In biomedical research we find only a very small number of applications relevant to the emerging topics "brain-computer interface" and "kinematics" and none of them got funded. In the field "energy and fuels" with the emerging topics "fuel cells" and "biodiesel" a substantial number of proposals, of which several were successful could be linked to these emerging topics. In "environmental sciences" we find a small number of proposals being mapped with the emerging topics "radiation" and "nano pollution", but in "radiation" four out of five applications got granted, while for "nano pollution" it was none. In the field "obstetrics and gynaecology" most of the 39 applications could be linked with the topic "cancer". All other topics present very little numbers of applications related to them. This includes the emerging topic "prenatal diagnosis". A similar situation is observed in "public health" research, where the emerging topics "living environment" and "AIDS in Africa" are not addressed well. On the other hand, "AIDS in Africa" presents a rather high success rate. In the field "geography" we find a strong focus of applications on the emerging topic "state-region" and also a high success rate of these proposals. These differences between thematic fields in terms of coverage and success rates of proposals indicate that is might be useful to explore in more details the reasons for such differences. This would allow a better insight into the operation of ERC procedures. Bibliometric approaches are certainly not suited for such analyses. Rather, qualitative expert-based approaches would be adequate.

In conclusion, the matching methodology developed within ERACEP has several advantages as the text-based link can be quantified and calculated as soon as an application is submitted. It is using the full body of the text and does not involve passing of the text to extract references that have to be matched with individual papers or journals. Any improvement to the body of the text will of course also be beneficial to the matching exercise.

IMPLICATIONS FOR ERC'S OPERATION

ERACEP developed methods and tools which allow the detection of emerging research topics and a matching of grant applications to different scientific fields including such emerging topics. Both approaches are conceptually independent from each other and can serve different purposes within the workflow of ERC. The identification of

emerging topics is more relevant to the internal organisation of the ERC. The mapping of proposals relates directly to the main objectives of ERC to support and encourage creative scientists to be advantageous and take more risks. In any case, we consider it as crucial for both approaches that they should not be considered and used as a replacement of any expert-based assessment. Rather they are suited to support the activities of human experts.

A first immediate use of the ability to identify emerging research topics is an initiation of internal discussions about the structure and scope of different evaluation panels. Descriptions of existing panels could refer explicitly to new topics, respective key words could be updated and existing panels revised or new panels created according to the landscape of emerging topics. Accordingly, the main advantage of using emerging topics would be that such adjustments and modifications would not be driven by external classifications, but rather by the inherent evolution of science. A second immediate use would be the support in identifying appropriate panel members with the right expertise in emerging topics needed for the evaluations.

The matching approach of ERACEP can be used in an ex-ante and in an ex-post way during the evaluation procedures. In the ex-ante mode the mapping of applications to certain topics could facilitate the assignment of applications to evaluation panels after submission. Applications could also be matched with a set of key papers which in turn would allow to better identify the most appropriate external experts which might be included in the evaluation procedures. Finally, a pre-selection of applications would be possible in a sense that applications can be mapped to thematic clusters labelled as "emerging", "established" or even "vanishing". Such a pre-selection would facilitate the identification of applications that would require special attention during the evaluation. Ex-post results of the mapping exercise can be used for a reflexive assessment of ERC procedures. In particular, similarities and differences between the outcomes of different panels across different fields could be analysed. Topics, where the reasons for selection or non-selection may need additional consideration could be better identified; finally the combined analysis of mapping, the results of the review process and the outcomes of the successful grants could provide an important feedback for adjusting the long-term strategy and policy of the ERC.

Contents

Chapter 1

General introduction

The European Research Council (ERC) was founded in 2007 as the first pan-European organisation with the clear aim to support excellent basic research. "The ERC's mission is to encourage the highest quality research in Europe through competitive funding and to support investigator-initiated frontier research across all fields of research on the basis of scientific excellence."[1]

This mission implies that ERC funding and operation should be able to address frontier research and emerging research areas. According to the High-Level Expert Group on "Maximising the wider benefits of competitive basic research funding at European level" (High-Level Expert Group, 2005), frontier research is a new concept replacing the traditional distinction between basic and applied research, arguing that frontier research can address both, the creation of new knowledge but also the generation of potentially useful knowledge at the same time. Following this line of argumentation the creation of new knowledge be it more basic or be it more "useful", is at the centre of frontier research. Frontier research can be characterised as a research activity aiming at the creation of novelty in terms of new emerging research fields. On the other hand, the process of doing frontier research involves risk taking and interdisciplinarity (High-Level Expert Group, 2005).

The mission of the ERC poses several challenges to its operational procedures. In particular it requires suitable evaluation and monitoring frameworks that allow the identification, assessment and support of truly creative research. In order to support its activities in this respect, the ERC has started to build up a portfolio of different projects and studies to provide intelligence for assisting its operational and strategic activities. In this line, a first call for Coordination and Support Actions was launched in 2008 and a second call (ERC, 2009) followed identifying four main topics:

- ERC science management and organisation,
- Research themes and scientific output,
- Researchers and host institutions, and
- Policy and structures.

[1] The mission statement of the ERC can be found on the ERC website, http://erc.europa.eu/mission.

Among these the "research themes and scientific output" topic asked for support activities developing tools and analyses which should facilitate the capturing and mapping of ERC-funded research within the landscape of science and draw implications for the emergence of new fields and Europe's competitive performance in such areas (ERC, 2009). The ERACEP (Emerging Research Areas and their Coverage by ERC-Supported Projects) support action was selected from this call.

The main objectives of ERACEP are the identification of topically emerging research areas, and the analysis to what extent the activities supported by the ERC cover and contribute to these research areas.

To meet these objectives, the ERACEP support action developed a conceptual approach which is combining two perspectives (Figure 1.1). The first perspective concerns the identification of topically emerging research areas with the assumption that excellent truly creative research can be found in particular in these emerging areas. The second perspective takes the view of ERC funding and explores how ERC-funded research themes map to the identified emerging areas.

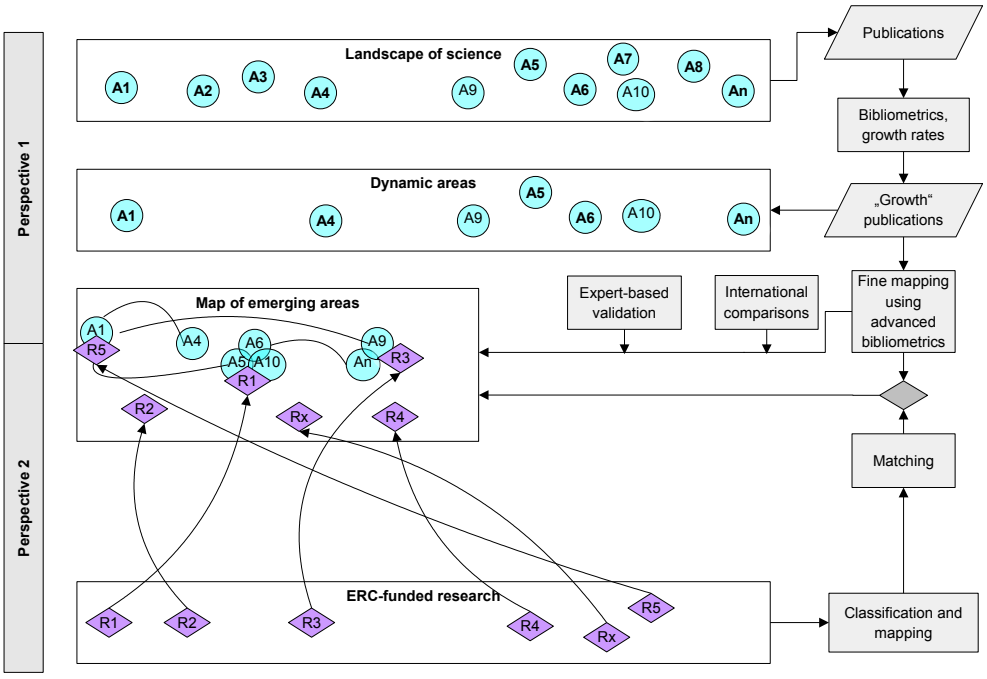

FIGURE 1.1. The ERACEP approach (A_x: Research area; R_x: ERC-funded research activity)

The notion of emerging research areas is at the centre of ERACEP. On the one hand, it is quite obvious that emerging research covers the novelty dimension of frontier research. On the other hand, the question comes up how emerging research can be defined and identified. According to Glänzel and Thijs (2012) it cannot be taken for granted that scientists themselves are really aware that their research field

is an emerging one. For example, new ideas underlying research activities or rapidly growing research output from certain areas might be indications of emerging research. However, this is certainly not a sufficient condition (Glänzel and Thijs, 2012). It is also required that a new emerging topic is characterised by some degree of coherence and also independence from its "mother topic" and other disciplines. This also implies that a certain sustainability of the topic and not just a sort of satellite character with respect to other disciplines should be given. These features of emerging research topics indicate that it is by far not easy to detect such topics.

Various methods have been suggested for identifying early research trends, in particular various foresight or forward-looking activities, such as Delphi surveys, expert-based brain storming activities, expert conferences, participatory workshops, world cafés, roadmappings, scenario approaches, and the like[2]. Although many such activities are currently undertaken in a lot of countries, there is an obvious scarcity of really new topics. Standard "future" themes frequently refer to very broad areas, such as biotechnology, nanotechnology, superconductivity, information and communication technologies and similar fields which are already relevant nowadays and it is expected that such fields have a future potential of relevant growth. Accordingly, research policies – be it in the regional or the national context – largely focus on supporting such fields. However, so far only few suggestions have been forwarded as to really new promising fields at a lower level of aggregation. Therefore, it is useful to think about new approaches to identify alternative new fields with relevance for the future development of technology and innovation.

The ERACEP support action adopted a bibliometric approach for the identification of emerging research fields. The specific advantage of a bibliometric search for emerging areas is its reference to objective data in contrast to the expert-based approaches mentioned above, which inevitably imply a high level of subjectivity. Most bibliometric analyses with regard to emerging areas are based on different citation methods (Takeda and Kajikawa, 2008, Shibata et al., 2008). However, these approaches are linked to specific fields and need concrete entry points as prerequisite and are not appropriate for broad open scans of all areas of science. Considering in particular the flexible bottom-up approach of ERC funding schemes, such open scans within the landscape of science are of particular importance for capturing and mapping ERC-funded research activities. Accordingly, within ERACEP we aimed at an open investigation in all fields of science, using advanced bibliometric tools.

The result of the bibliometric analysis is a map of promising emerging areas, distributed over all areas of science, requiring excellent and creative research for their further development. The second perspective of the ERACEP concept takes the view of ERC-funded activities. The challenge here is to map these activities to the landscape of emerging topics as identified during the first part of ERACEP. We explored three different approaches for this matching activity. The first approach basically comprises

[2] An overview of current forward-looking activities at a European and also trans-European level can be found on the socio-economic sciences and humanities website of the European Commission, http://ec.europa.eu/research/social-sciences/forward-looking_en.html.

a qualitative assessment of grant applications based on titles, keywords and subject assignments. The second approach is based on scientific journals mentioned in grant applications, which could be mapped to dynamic emerging fields as identified within Part 1 of ERACEP. The third approach, which finally turned out to be the most useful one, comprises a text matching between proposals and thematic clusters within the landscape of science by using text-mining tools.

The ERACEP strategy adopts two basic principles:

Principle 1: Combine quantitative statistic approaches with qualitative expert-based approaches.

While advanced bibliometric and text-mining methods play a crucial role in the ERACEP approach in order to identify emerging topics, we also put a lot of emphasise on a validation of such results via additional qualitative expert assessments and literature analysis.

Principle 2: Exploration and concentration.

ERACEP per se has an explorative and experimental character in a sense that it developed and tested novel tools and methods, which had not yet been established in the scientific community. Accordingly, the aim of ERACEP was not to provide a complete coverage of all emerging topics within science and all ERC-supported activities. Rather, the goal of ERACEP was to provide and develop a new procedure in an experimental way and illustrate this approach for a selected sample of emerging topics, and for a selected number of grant applications.

These two principles are illustrated in the following Figure 1.2.

The conceptual approach of ERACEP is translated into the following operational objectives:

Objective 1: Identification of dynamic areas within the landscape of science

Objective 2: Identification of emerging topics within the identified dynamic areas of science.

Objective 3: Mapping of ERC-funded research to the emerging topics identified in the previous steps.

The structure of this report corresponds to these objectives in the following way: In Chapter 2 we will describe in detail how dynamic areas are identified within the landscape of science. Thereby objective 1 will be addressed. Chapter 3 corresponding to objective 2 elaborates the fine mapping of dynamic fields in order to identify emerging topics of science. Chapter 4 is concerned with the perspective of ERC funding activities and presents the matching of ERC proposals to emerging topics of science. Accordingly, in this section we will elaborate on objective 3. In Chapter 5 conclusions are drawn focussing on methodological developments, on content issues, and in particular on the user perspective in a sense that benefits, opportunities but also limitations of applying ERACEP insights within ERC procedures will be worked out.

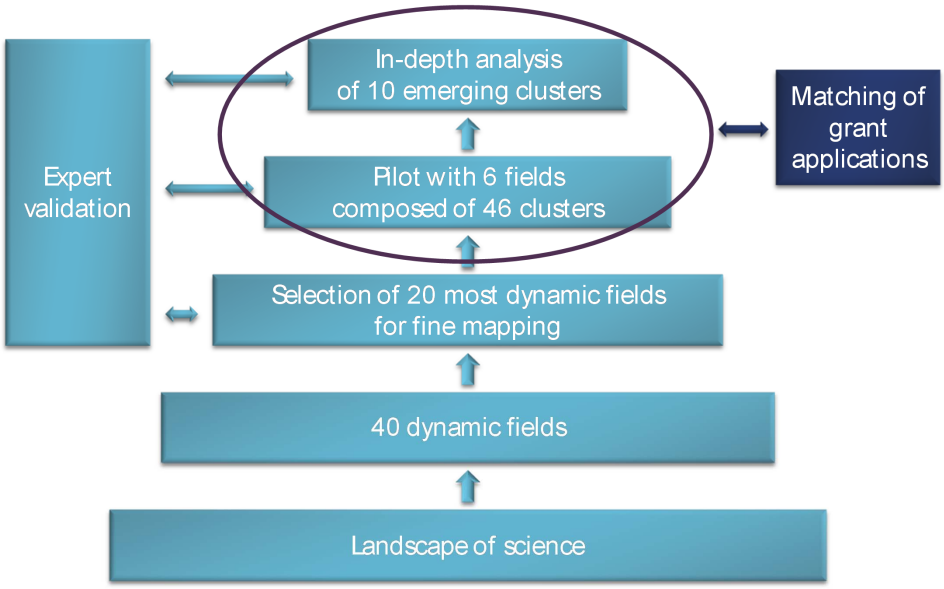

FIGURE 1.2. Two guiding principles of the ERACEP approach: 1. Bibliometrics and expert validation; 2. Exploration and concentration

Chapter 2

Dynamic fields

2.1 METHODOLOGICAL APPROACH

As a first step for the identification of emerging topics, dynamic fields within the landscape of science were identified based on publishing activities. The rational behind this approach is the notion that dynamic growth of a specific topic reflects on the one hand an increasing interest of scientists in this topic, so that more research groups are doing research in the respective area, resulting in an increasing number of scientific publications. On the other hand, dynamics could also indicate that a more or less constant number of scientists are increasing their research activities, resulting again in a growing publication output. Both phenomena point to the emergence of new fields in science (Schmoch et al., 2005, Reiss et al., 2007). In this chapter we will describe the methods and procedures applied for the identification of such dynamic fields. These fields will be presented and discussed with respect to their delineation, coverage and growth characteristics.

In order to identify dynamically growing research fields, mainly two distinct measures were applied: the overall growth ratio and the Sharpe ratio. The Sharpe ratio originally was developed in the context of financial sciences and is used for the observation and evaluation of stock markets (Fischer, 2001, p. 271). It considers the development of a specific field in relation to the growth of all fields. Thereby the Sharpe ratio allows for adjusting to differences in absolute field sizes. This index has a positive value if the relative growth of a field is above the growth of all fields, and a negative value if the relative growth of the field is below the overall growth of the stock of publications in all fields. In addition, the growth rate is normalised by the standard deviation of the annual growth rates of the field considered, so that the stability of growth is included as well. In general, we can expect for the smaller fields that the annual fluctuations of growth are higher than for larger ones. Thus, the Sharpe ratio offers an improved comparability of different fields as it compensates size effects to a certain extent.

The Sharpe ratio (Sharpe, 1966) was calculated as follows:

$$\text{Sharpe ratio}_j = \frac{G_j - G}{D_j} \tag{2.1}$$

with

G_j = Absolute growth of field j in period 1998-2008 (2003-2008)
G = Absolute growth of all field in period 1998-2008 (2003-2008)
D_j = Standard deviation of annual growth in field j in period 1998-2008 (2003-2008)

The considered time period was from 1998 to 2008 with two investigation periods, a ten-year period from 1998 to 2008 and a five-year period from 2003 to 2008. The following indicators were calculated:

- Growth rates for each year from 1998 to 2008.
- Overall growth ratios, mean annual growth, standard deviation in growth ratios and Sharpe ratio.

In order to analyse the data, the different fields were sorted according to the overall growth rates and Sharpe ratio for both periods. Fields were selected according to Sharpe ratio rankings due to the advantages in terms of compensating for size effects as mentioned above.

In addition to these measures for dynamics of the fields the size of the fields was considered by indicating the total number of publications in the year 2008. The assumption behind the decision to include some larger fields which are not yet identified as highly dynamic fields was, that within these large fields dynamics in subfields might exist which would otherwise not be identified.

In order to cover all areas of science and to consider differences between the fields, separate analyses were conducted for the Science Citation Index Expanded (SCIE), the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI) of Thomson Reuters.

Publication data from the Science Citation Index (SCIE) was accessed via the commercial provider STN. Subject codes were derived from Thomson Reuters' homepage (http://www.thomsonscientific.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=K). Over all, there are currently 172 different subject categories in the SCIE which were included in our analysis. From the database the number of publications per subject category and year (1998-2008 separately) worldwide were retrieved. The search was restricted to articles only. As there were some changes in the database structure and content during the ten-years period considered (particluarly with respect to nanosciences and nanotechnology, and mathematical and computational biology which emerged as new categories), the five-years period 2003 to 2008 was used for the final evaluation of growth fields. For social sciences publication data from the Social Science Citation Index (SSCI) was accessed via the host Dialogue. Again, subject categories were derived from Thomson Reuters' homepage (http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/

jlsubcatg.cgi?PC=SS). Over all, there are currently 55 different subject categories in the SSCI.

During some first analyses we observed that some fields are rather small and also a high degree of overlap in the journals assigned to some subject categories became evident. In order to deal with these issues, we introduced the following five aggregate subject categories.

1. BUSINESS OR ECONOMICS OR (BUSINESS, FINANCE) OR MANAGEMENT OR PLANNING & DEVELOPMENT)
2. (INTERNATIONAL RELATIONS) OR (POLITICAL SCIENCE) OR (PUBLIC ADMINISTRATION)
3. COMMUNICATION OR (EDUCATION & EDUCATIONAL RESEARCH) OR (EDUCATION, SCIENTIFIC DISCIPLINES) OR (EDUCATION, SPECIAL) OR (INFORMATION SCIENCE & LIBRARY SCIENCE)
4. ANTHROPOLOGY OR (ETHNIC STUDIES) OR (FAMILY STUDIES) OR SOCIOLOGY OR (WOMEN'S STUDIES)
5. DEMOGRAPHY OR (SOCIAL ISSUES) OR (SOCIAL SCIENCES, INTERDISCIPLINARY) OR (SOCIAL SCIENCES, BIOMEDICAL) OR (SOCIAL WORK)

The number of publications per category and year (1998-2008) worldwide was extracted. Compared to the SCIE search, the search strategy within the SSCI was extended to include articles as well as book reviews as these are "typical" forms of communication in this area.

Since a detailed analysis of time series of different fields indicated that between 1998 and 2002 in some cases only very few publications were included we used the latest period 2003 to 2008 for the identification of dynamic fields (s. above).

Publications from the area of arts and humanities were extracted from the Arts & Humanities Citation Index (AHCI) via the host Dialogue. Again, subject codes were taken from Thomson Reuters' homepage (http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=H). In total, there are currently 26 different subject categories in the AHCI. The number of publications per category and year over the period 1998 to 2008 worldwide was retrieved.

In the arts and humanities different communication media are used compared to the science area and also compared to social sciences. Accordingly, we applied a broader search strategy including not only articles but also the categories book review, dance performance review, editorial, editorial material, film review, letter, music performance review, music score review, record review, review bibliography, software review, theatre review, tv review, radio review.

2.2 RESULTS

Using the growth measures as described in the previous section, 50 dynamic fields were identified from the Science Citation Index Expanded based on their Sharpe ratios over the period 2003 to 2008 (Table 2.1). The top three fields with the largest Sharpe ratio are

TABLE 2.1. Top 50 dynamic SCIE fields sorted by Sharpe ratio, period 2003-200

No	SCIE Subject Code	Factor (03-08)	Mean (03-08)	Std dev (03-08)	Sharpe ratio (03-08)
1	NANOSCIENCE & NANOTECHNOLOGY	49.41	161.55	188.80	0.26
2	MATHEMATICAL & COMPUTATIONAL BIOLOGY	30.60	126.33	124.04	0.24
3	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	0.64	10.36	2.33	0.16
4	NURSING	1.71	22.38	9.46	0.15
5	ORTHOPEDICS	0.59	9.78	2.34	0.14
6	ENVIRONMENTAL SCIENCES	0.55	9.20	2.60	0.11
7	OPERATIONS RESEARCH & MANAGEMENT SCIENCE	0.70	11.25	4.08	0.11
8	OBSTETRICS & GYNECOLOGY	0.40	7.02	1.40	0.10
9	MEDICAL ETHICS	1.10	16.33	9.17	0.09
10	ROBOTICS	0.57	9.44	3.44	0.09
11	ENGINEERING, BIOMEDICAL	0.60	9.84	3.84	0.09
12	ENERGY & FUELS	0.62	10.18	4.14	0.09
13	ONCOLOGY	0.34	5.96	0.82	0.08
14	ENGINEERING, MULTIDISCIPLINARY	0.75	12.01	5.84	0.08
15	BIOTECHNOLOGY & APPLIED MICROBIOLOGY	0.45	7.75	2.26	0.08
16	HEALTH CARE SCIENCES & SERVICES	0.58	9.64	3.89	0.08
17	MATERIALS SCIENCE, CHARACTERIZATION & TESTING	0.54	9.10	3.69	0.07
18	MATERIALS SCIENCE, MULTIDISCIPLINARY	0.53	8.92	3.73	0.07
19	MECHANICS	0.48	8.14	3.12	0.07
20	STATISTICS & PROBABILITY	0.54	9.04	4.04	0.07
21	MATERIALS SCIENCE, BIOMATERIALS	0.73	11.80	7.09	0.07
22	ZOOLOGY	0.46	7.91	2.96	0.07
23	ELECTROCHEMISTRY	0.80	12.67	8.11	0.07
24	BEHAVIORAL SCIENCES	0.59	9.76	5.07	0.06
25	MATHEMATICS, INTERDISCIPLINARY APPLICATIONS	1.21	18.03	15.11	0.06
26	CHEMISTRY, MULTIDISCIPLINARY	0.44	7.60	2.86	0.06
27	GERIATRICS & GERONTOLOGY	0.56	9.42	4.90	0.06
28	EDUCATION, SCIENTIFIC DISCIPLINES	0.59	9.86	5.48	0.06
29	REHABILITATION	0.57	9.53	5.19	0.06
30	CHEMISTRY, PHYSICAL	0.51	8.71	4.34	0.06
31	ENGINEERING, CIVIL	0.68	11.20	7.69	0.05
32	OPTICS	0.54	9.17	5.11	0.05
33	CHEMISTRY, MEDICINAL	0.54	9.09	5.05	0.05
34	MICROBIOLOGY	0.35	6.13	1.51	0.05
35	EVOLUTIONARY BIOLOGY	0.48	8.17	4.06	0.05
36	AGRICULTURE, MULTIDISCIPLINARY	0.73	11.87	8.98	0.05
37	AGRICULTURAL ENGINEERING	1.55	22.35	24.89	0.05
38	PUBLIC, ENVIRONMENTAL & OCCUPTIONAL HEALTH	0.47	8.13	4.29	0.05
39	CHEMISTRY, APPLIED	0.44	7.63	3.71	0.05
40	GEOLOGY	0.52	8.80	5.60	0.04
41	INTEGRATIVE & COMPLEMENTARY MEDICINE	0.82	13.31	12.57	0.04
42	EMERGENCY MEDICINE	0.62	10.46	9.03	0.04
43	REMOTE SENSING	0.57	9.69	7.82	0.04
44	PARASITOLOGY	0.53	8.99	6.72	0.04
45	PHYSICS, APPLIED	0.48	7.02	5.49	0.04
46	FOOD SCIENCE & TECHNOLOGY	0.51	8.78	6.46	0.04
47	TELECOMMUNICATIONS	0.57	9.61	8.02	0.04
48	PSYCHOLOGY	0.42	7.39	4.28	0.04
49	CLINICAL NEUROLOGY	0.30	5.43	0.97	0.04
50	MATHEMATICS	0.38	6.67	3.10	0.04

- Nanoscience and nanotechnology
- Mathematical and computational biology
- Computer science and interdisciplinary applications

The field size of these three areas as indicated by the number of publications in 2008, is quite diverse, ranging from 3,919 publications in the case of mathematical and computational biology, to 15,577 publications in the case of nanoscience and nanotechnology.

Out of this total list of 50 dynamic fields only a selected number of fields could be used for the further analysis in order to identify emerging topics. During the selection procedure some specific features of individual fields had to be considered. In particular it turned out that all fields described as "multidisciplinary" or "interdisciplinary" are quite broad and display a high level of overlap with other more specialised fields.

In these cases, the more narrowly defined fields were selected instead of the broader ones since it could be expected that such a more precise demarcation of fields would facilitate the following bibliometric fine analysis for the identification of emerging topics. Due to the high level of overlap in articles assigned to both fields, no information is assumed to be lost due to this procedure. An example for such a selection are the fields no. 21 (materials science and biomaterials) and no. 18 (materials science, multidisciplinary). In this case the more narrowly defined field "materials science and biomaterials" was preferred. In order to avoid further overlap, also the combination of certain fields was considered. In particular, the fields no. 29 (rehabilitation) and no. 16 (health care sciences and services) were combined with field no. 4 (nursing). Following these modifications, finally 27 fields were selected from the Science Citation Index and ended into the top 40 list of dynamic fields as shown in Table 2.4 below.

Applying a similar procedure, the most dynamic publications as measured by Sharpe ratio over the period 2003 to 2008 were selected from the social sciences (Table 2.2). The most dynamic fields are

- Public, environmental and occupational health
- Nursing and
- Transportation.

Among these, "public, environmental and occupational health" represents the largest field with 7,862 publications in 2008, "nursing" corresponds to 4,826 publications, while in the field "transportation" 1,063 publications were counted for the year 2008.

Out of these fields the first eight categories were selected for the top 40 list of dynamic fields summarised in Table 2.4.

When comparing the results of the Social Science Citation Index Expanded with the results from the Science Citation Index (Table 2.1 and Table 2.2), it becomes obvious that the subject code "nursing" is appearing in both domains indicating a large overlap between these fields. Accordingly, for the following analyses the field "nursing" is treated only once. A similar situation holds true for the field "linguistic" which is also covered by the Arts & Humanities Citation Index (see below), and will also be treated only once.

As described in section 2.1, the Arts& Humanities Citation Index contains only 26 subject categories. The top 20 dynamic AHCI fields sorted by Sharpe ratio over the period 2003 to 2008 are summarised in Table 2.3.

The following fields are the most dynamic ones within the AHCI:

- Archaeology
- Religion
- Architecture

Among these, "religion" with 7,876 publications in 2008 is the largest one. "Architecture" is characterised by 3,421 publications in 2008 followed by "archaeology"

TABLE 2.2. Top 39 dynamic SSCI fields sorted by to Sharpe ratio, period 2003-2008.

No	SSCI Subject Code	Field size (2008)	Factor	Mean 03-08	Std dev 03-08	Sharpe 03-08
1	PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH	7862	0.75	0.12	0.04	11.1586
2	NURSING	4826	1.56	0.21	0.11	10.8303
3	TRANSPORTATION	1063	1.20	0.17	0.10	8.1618
4	GEOGRAPHY	3191	0.81	0.13	0.06	7.8758
5	LINGUISTICS	3919	1.27	0.19	0.20	4.5669
6	PSYCHOLOGY, EXPERIMENTAL	4860	0.44	0.08	0.03	2.7242
7	BUSINESS OR ECONOMICS OR (BUSINESS, FINANCE) OR (MANAGEMENT OR PLANNING & DEVELOPMENT)	24892	0.45	0.08	0.07	1.6456
8	SOCIAL SCIENCES, MATHEMATICAL METHODS	1907	0.43	0.07	0.05	1.4588
9	HEALTH POLICY & SERVICES	3253	0.41	0.07	0.05	1.2374
10	ERGONOMICS	1013	0.40	0.07	0.05	0.9763
11	ENVIRONMENTAL STUDIES	3905	0.40	0.07	0.08	0.6753
12	PSYCHOLOGY, DEVELOPMENTAL	3445	0.38	0.07	0.08	0.4034
13	DEMOGRAPHY OR (SOCIAL ISSUES) OR (SOCIAL SCIENCES, INTERDISCIPLINARY) OR (SOCIAL SCIENCES, BIOMEDICAL) OR (SOCIAL WORK)	9341	0.35	0.06	0.02	0.3822
14	PSYCHOLOGY, MATHEMATICAL	558	0.36	0.07	0.09	0.1223
15	ETHICS	1695	0.35	0.06	0.07	0.0980
16	PSYCHOLOGY, APPLIED	2482	0.35	0.07	0.09	0.0840
17	PSYCHOLOGY, SOCIAL	2939	0.32	0.06	0.05	-0.5837
18	SUBSTANCE ABUSE	1537	0.28	0.06	0.10	-0.6004
19	PSYCHOLOGY, CLINICAL	5395	0.31	0.06	0.05	-0.6343
20	CRIMINOLOGY & PENOLOGY	1377	0.32	0.06	0.05	-0.6595
21	PSYCHOLOGY, EDUCATIONAL	1557	0.30	0.06	0.07	-0.7208
22	GERONTOLOGY	2059	0.28	0.05	0.09	-0.7265
23	AREA STUDIES	3533	0.30	0.06	0.06	-0.7868
24	REHABILITATION	2266	0.29	0.05	0.07	-0.7980
25	ANTHROPOLOGY OR (ETHNIC STUDIES) OR (FAMILY STUDIES) OR SOCIOLOGY OR (WOMEN'S STUDIES)	13242	0.29	0.05	0.05	-1.0688
26	PSYCHOLOGY, BIOLOGICAL	1049	0.30	0.05	0.04	-1.1497
27	PSYCHIATRY	7574	0.29	0.05	0.04	-1.4677
28	HISTORY	6981	0.24	0.05	0.07	-1.4760
29	(INTERNATIONAL RELATIONS) OR (POLITICAL SCIENCE) OR (PUBLIC ADMINISTRATION)	10523	0.28	0.05	0.04	-1.5524
30	HISTORY OF SOCIAL SCIENCES	1185	-0.05	0.00	0.17	-2.3432
31	SYCHOLOGY, MULTIDISCIPLINARY	5664	0.19	0.04	0.06	-2.7219
32	PSYCHOLOGY, MULTIDISCIPLINARY	5664	0.19	0.04	0.06	-2.7219
33	COMMUNICATION OR (EDUCATION & EDUCATIONAL RESEARCH) OR (EDUCATION, SCIENTIFIC DISCIPLINES) OR (EDUCATION, SPECIAL) OR (INFORMATION SCIENCE & LIBRARY SCIENCE)	18649	0.22	0.04	0.04	-3.0542
34	LAW	3850	0.16	0.03	0.06	-3.2117
35	HISTORY & PHILOSOPHY OF SCIENCE	2155	0.20	0.04	0.04	-4.3067
36	URBAN STUDIES	1427	0.10	0.02	0.06	-4.3091
37	INDUSTRIAL RELATIONS & LABOR	702	-0.05	-0.01	0.06	-6.5441
38	PSYCHOLOGY, PSYCHOANALYSIS	639	-0.13	-0.02	0.06	-7.7738
39	HOSPITALITY, LEISURE, SPORT & TOURISM	1023		13.55	3.75	

comprising 2,703 publications in 2008. The top 5 fields of the AHCI were selected for the further analysis and were entered into the top 40 list of all scientific fields. In addition, we also had a closer look at the field "language and linguistic", which does not appear in the table due to obvious database problems (in the period 2001-2004 only very few (below 10) publications are listed in the database). However, since we observe an extraordinary growth for this field over the period 2006 (1,209 publications) to 2008 (5,339 publications), this field was also considered in the more detailed fine-mapping analysis.

The list of the 40 most dynamic fields from the SCIE, the SSCI, and the AHCI is present-ed in Table 2.4. It consists of 27 fields from the Science Citation Index, 8 from social sciences and 5 fields from arts and humanities. Since "nursing" appears twice, the real number of this list indeed is 40 and not 41.

Table 2.3. Top 20 dynamic AHCI fields sorted by Sharpe ratio, period 2003-2008

No	AHCI Subject Code	Field size (2008)	Factor	Mean 03-08	Std dev 03-08	Sharpe 03-08
1	ARCHAEOLOGY	2703	0,56	0,10	0,09	48,595
2	RELIGION	7876	0,41	0,07	0,07	46,177
3	ARCHITECTURE	3421	0,38	0,07	0,06	43,024
4	MEDIEVAL & RENAISSANCE STUDIES	2299	0,61	0,12	0,23	21,860
5	PHILOSOPHY	9375	0,29	0,05	0,08	21,392
6	HISTORY & PHILOSOPHY OF SCIENCE	2747	0,24	0,04	0,06	19,964
7	LITERATURE, SLAVIC	563	0,33	0,07	0,18	12,673
8	POETRY	701	0,20	0,04	0,09	10,046
9	FILM, RADIO, TELEVISION	3670	0,42	0,11	0,34	0,9277
10	HUMANITIES, MULTIDISCIPLINARY	11326	0,11	0,03	0,12	-0,0387
11	HISTORY	23604	0,11	0,02	0,06	-0,0652
12	DANCE	1491	0,07	0,03	0,20	-0,2003
13	LITERARY THEORY & CRITICISM	1055	0,00	0,04	0,32	-0,3598
14	LITERATURE, AMERICAN	723	-0,04	0,03	0,33	-0,4694
15	LITERATURE, ROMANCE	4098	0,09	0,02	0,04	-0,5032
16	LITERATURE, BRITISH ISLES	641	-0,09	-0,01	0,16	-12,588
17	ASIAN STUDIES	1659	-0,03	0,00	0,08	-16,701
18	LITERATURE, GERMAN, DUTCH, SCANDINAVIAN	1060	-0,05	-0,01	0,10	-16,942
19	LITERATURE	14051	-0,03	0,00	0,06	-21,130
20	LITERARY REVIEWS	3561	-0,11	-0,02	0,10	-22,098

The list of the 40 most dynamic fields of science was discussed with the ERC and it was concluded to further narrow-down this list to a final list of 20 dynamic fields, which would be analysed in detail using various bibliometric methods. This list contains 13 fields from the sciences, 4 fields from the social sciences and finally 3 fields from arts and humanities (Table 2.5).

During this selection procedure a number of fields of the top 40 dynamic fields (Table 2.4) was rejected for different reasons. The fields "nanoscience and nanotechnology" as well as "mathematical and computational biology" were excluded since in the Science Citation Index Expanded these fields have been defined as aggregations of parts of other fields. Only during recent years they have been defined as independent subject categories. Accordingly, the dynamic analysis for these fields is not comparable to the other fields. The field "computer science and interdisciplinary applications" was also disregarded because of its broad overlaps with other fields due to its interdisciplinary character. The field "health care sciences and services" was combined with the field "nursing". The same holds true for the field "material sciences and biomaterials" which was merged with "materials science, characterisation and testing". The field "statistics and probability" was included in the field "operations research and management science". "Mechanics" and "zoology" turned out to be rather large and not very well defined broad fields, so that they seemed not to be suited for bibliometric analysis.

In the case of social sciences "nursing" was deselected since it already appears within the sciences. "Linguistics" is another case which was not considered further because this field is also part of arts and humanities. However, within arts and humanities it does not appear among the most dynamic fields.

TABLE 2.4. 40 most dynamic fields from the SCIE, SSCI, AHCI (in order to facilitate comparisons, the numbers in the first column correspond to the numbers in the first columns of Tables 2.1–2.3 without adjustments)

No	SCIE Subject Code	Field size (2008)	Factor	Mean 03-08	Std dev 03-08	Sharpe 03-08
1	NANOSCIENCE & NANOTECHNOLOGY	15577	49.41	161.55	188.80	**0.26**
2	MATHEMATICAL & COMPUTATIONAL BIOLOGY	3919	30.60	126.33	124.04	**0.24**
3	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	9744	0.64	10.36	2.33	**0.16**
4	NURSING	4675	1.71	22.38	9.46	**0.15**
5	ORTHOPEDICS	7992	0.59	9.78	2.34	**0.14**
6	ENVIRONMENTAL SCIENCES	25428	0.55	9.20	2.60	**0.11**
7	OPERATIONS RESEARCH & MANAGEMENT SCIENCE	6209	0.70	11.25	4.08	**0.11**
8	OBSTETRICS & GYNECOLOGY	9108	0.40	7.02	1.40	**0.10**
9	MEDICAL ETHICS	663	1.10	16.33	9.17	**0.09**
10	ROBOTICS	995	0.57	9.44	3.44	**0.09**
11	ENGINEERING, BIOMEDICAL	7071	0.60	9.84	3.84	**0.09**
12	ENERGY & FUELS	10452	0.62	10.18	4.14	**0.09**
13	ONCOLOGY	23652	0.34	5.96	0.82	**0.08**
15	BIOTECHNOLOGY & APPLIED MICROBIOLOGY	20245	0.45	7.75	2.26	**0.08**
16	HEALTH CARE SCIENCES & SERVICES	4994	0.58	9.64	3.89	**0.08**
17	MATERIALS SCIENCE, CHARACTERIZATION & TESTING	1914	0.54	9.10	3.69	**0.07**
19	MECHANICS	13772	0.48	8.14	3.12	**0.07**
20	STATISTICS & PROBABILITY	7597	0.54	9.04	4.04	**0.07**
21	MATERIALS SCIENCE, BIOMATERIALS	3257	0.73	11.80	7.09	**0.07**
22	ZOOLOGY	9723	0.46	7.91	2.96	**0.07**
23	ELECTROCHEMISTRY	7595	0.80	12.67	8.11	**0.07**
24	BEHAVIORAL SCIENCES	4805	0.59	9.76	5.07	**0.06**
27	GERIATRICS & GERONTOLOGY	2955	0.56	9.42	4.90	**0.06**
28	EDUCATION, SCIENTIFIC DISCIPLINES	2356	0.59	9.86	5.48	**0.06**
29	REHABILITATION	2613	0.57	9.53	5.19	**0.06**
30	CHEMISTRY, PHYSICAL	37248	0.51	8.71	4.34	**0.06**
31	ENGINEERING, CIVIL	9576	0.68	11.20	7.69	**0.05**
32	OPTICS	18982	0.54	9.17	5.11	**0.05**
	SSCI Subject Code					
1	PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH	7862	0.75	0.12	0.04	**11.16**
2	NURSING	4826	1.56	0.21	0.11	**10.83**
3	TRANSPORTATION	1063	1.20	0.17	0.10	**8.16**
4	GEOGRAPHY	3191	0.81	0.13	0.06	**7.88**
5	LINGUISTICS	3919	1.27	0.19	0.20	**4.57**
6	PSYCHOLOGY, EXPERIMENTAL	4860	0.44	0.08	0.03	**2.72**
7	BUSINESS OR ECONOMICS OR (BUSINESS, FINANCE) OR MANAGEMENT OR PLANNING & DEVELOPMENT) AND DT=(ARTICLE OR (BOOK REVIEW	24892	0.45	0.08	0.07	**1.65**
8	SOCIAL SCIENCES, MATHEMATICAL METHODS	1907	0.43	0.07	0.05	**1.46**
	AHCI Subject Code					
1	ARCHAEOLOGY	2703	0.56	0.10	0.09	**4.86**
2	RELIGION	7876	0.41	0.07	0.07	**4.62**
3	ARCHITECTURE	3421	0.38	0.07	0.06	**4.30**
4	MEDIEVAL & RENAISSANCE STUDIES	2299	0.61	0.12	0.23	**2.19**
5	PHILOSOPHY	9375	0.29	0.05	0.08	**2.14**

2.3 DISCUSSION

In the following section we will have a closer look at the 20 identified growth fields and provide a brief characterisation and discussion of the individual fields. Among the 13 most dynamic fields from the Science Citation Index Expanded (SCIE) we find five fields which can be categorised in to the broad area of medicine. These include "nursing", "orthopaedics", "obstetrics and gynaecology", "medical ethics", and "oncology". They span a large size range with "oncology" being the largest corresponding to 23,652 pub-lications in 2008 and "medical ethics" as the smallest

Table 2.5. Final list of 20 dynamic fields

SCIE	Factor	Mean 03-08	Std dev 03-08	Sharpe 03-08
Nursing	1.71	22.38	9.46	0.15
Orthopedics	0.59	9.78	2.34	0.14
Environmental sciences	0.55	9.20	2.60	0.11
Operations research & management science	0.70	11.25	4.08	0.11
Obstetrics & gynecology	0.40	7.02	1.40	0.10
Medical ethics	1.10	16.33	9.17	0.09
Robotics	0.57	9.44	3.44	0.09
Engineering, biomedical	0.60	9.84	3.84	0.09
Energy & fuels	0.62	10.18	4.14	0.09
Oncology	0.34	5.96	0.82	0.08
Biotechnology & applied microbiology	0.45	7.75	2.26	0.08
Material sciences, characterization & testing	0.54	9.10	3.69	0.07
Behavioral sciences	0.59	9.76	5.07	0.06
SSCI				
Public, environmental & occupational health	0.75	0.12	0.04	11.16
Transportation	1.20	0.17	0.10	8.16
Geography	0.81	0.13	0.06	7.88
Psychology, experimental	0.44	0.08	0.03	2.72
AHCI				
Archeology	0.56	0.10	0.09	4.86
Religion	0.41	0.07	0.07	4.62
Architecture	0.38	0.07	0.06	4.30

dynamic field with 663 publications in 2008. Three of these fields are mainly concerned with the clinical and the therapeutical dimension of medicine[1]. These include "orthopaedics", "obstestrics and gynaecology", and "oncology". "Orthopaedics" covers all issues related to preservation or restorage of the musculosceletal system. "Obstetrics and gynaecology" is concerned with femal reproductive function and reproductive organs. "Oncology" is concerned with all issues related to mechanisms, causes and treatments of cancer.

The remaining two fields of medicine, "nursing" and "medical ethics" are more con-cerned with healthcare issues and the social dimension of medicine. "Nursing" is de-fined as a rather broad category covering all issues of nursing sciences and practice. It is interesting to observe that this field, not directly concerned with clinical or therapeutical developments related to important diseases such as cancer or infectious diseases, turned out to be the most dynamic one among our 20 growth fields. Obviously, the healthcare dimension of medicine has drawn increasing attention within science during the last years.

A second group of dynamic fields can be classified into the engineering domain. These include rather small and more specialised fields like "robotics" and "materials sciences, characterisation and testing" and the larger field of "engineering, biomedical". "Robotics" includes all activities related to the design, training and

[1] More detailed information about the different fields as described in the following is taken from the scope notes of the SCIE, SSCI, and AHCI: http://ip-science.thomsonreuters.com/mjl/scope/scope_sci/; http://ip-science.thomsonreuters.com/mjl/scope/scope_ssci/; http://ip-science.thomsonreuters.com/mjl/scope/scope_ahci/

application of robots. It is closely related to mechanical and electrical engineering, cybernetics and also artificial intelligence. "Materials science, characterisation and testing" covers all sorts of techniques that are used to evaluate and test materials. These include, for example, non-destructive testing but also mechanical testing methods. In particular, also novel approaches such as scanning tunnelling microscopy are included in this category. Finally, "biomedical engineering" has a close relation to medical sciences since it includes various engineering technologies which are used for solving medical problems.

Two of the thirteen growth fields from the Science Citation Index Expanded (SCIE) can be grouped into biological sciences. These include "biotechnology and applied microbiology" and "behavioural sciences". The biotechnology category is a large field covering basically a rather broad range of topics related to the modification of living organisms to make products or improve processes. "Behavioural sciences" on the other hand is a smaller field concerned mainly with the biological dimension of human and animal behaviour.

The three remaining fields do not belong to an additional subgroup but can be considered as stand-alone fields. The first one is "operations research and management science" which has a strong mathematical dimension, including issues related to the definition, analysis and solution of complex problems drawing a lot on mathematical modelling, stochastic modelling or decision theory. "Environmental sciences" and "energy and fuels" are two large fields being concerned with a broad range of different issues. In the case of "environmental sciences" these include, for example, analytical approaches such as detecting contamination in the environment, toxicology, but also environmental health and environmental management. "Energy and fuels" is a similarly broad category including all issues related to non-renewable but also to renewable energy sources. "Nuclear energy and nuclear technology" is not included in this group but covered by own SCIE subject code.

The four most dynamic fields out of the Social Sciences Citation Index (SSCI) interestingly also cover two fields related to medicine. These are "public, environmental and occupational health" and "experimental psychology". The first category is concerned with topics related to social medicine, health behaviour but also health education and safety research. The fact that this group belongs to the most dynamic areas identified during our research together with the already discussed SCIE topics "nursing" and "medical ethics", again points to the notion that the social dimension of medical research has become increasingly important during recent years. "Experimental psychology" is strongly related to neuro sciences since it is concerned with issues such as consciousness, cognition or memory and also visual, auditory, and speech perception. In other words, this category has a strong science relationship. "Transportation" deals mainly with the political dimension of transportation and also with economics and management issues. Finally, "geography" is emphasising human, economic, political and environmental issues of this field.

From the Arts & Humanities Citation Index (AHCI) three quite different fields have been identified as most dynamic ones. "Archaeology" includes issues related to

the study of material remains of past human life and activities, not only covering the social dimension but also methodological issues of detection and analysis. "Religion" is concerned with the major world religions, thereby by nature being quite broad and diverse. "Architecture" as the last of the 20 dynamic fields is not only addressing the design and construction of buildings, but also covers issues such as architectural history, urban and country planning and landscape architecture.

In summary, our 20 most dynamic fields present on the one hand a slight bias towards medical and biological sciences, which on the other hand is not surprising considering the large attention of medical and biological sciences among science funding activities worldwide, and also considering the well established institutional and other infrastructures in this area. On the other hand, we also observe a distinct diversity of these 20 fields ranging from very important grand-challenges-related issues such as "energy supply and environment" to fundamental cultural issues such as "religion" and "archaeology".

2.4 VALIDATION

Considering the overall approach of ERACEP the identification of dynamic fields within the landscape of science does not represent a mandatory step in the search for emerging topics. Rather, the availability of dynamic fields as a starting point facilitates the identification of such topics since it narrows down the search space. Considering this function of the dynamic fields it is not crucial that such fields are validated carefully for example by additional external expert assessments.

Accordingly our approach towards validation of dynamic fields concentrated on an internal assessment through the project team in close coordination with the ERC. During several project meetings with representatives of the ERC each of the fields was discussed intensively. Issues considered during these assessments include relevance of the field in terms of current and future scientific potential, relationship to research currently funded by the ERC, and balance of thematic orientation of the whole portfolio of identified fields in order to avoid biases towards few disciplines. As a result there was consent that the twenty selected dynamic fields represent a solid sample for the further detailed analyses as described in the following sections.

Chapter 3

Structural topic analysis and identification of emerging topics

3.1 INTRODUCTION

The procedure described in Chapter 2 provided a set of ISI Subject Categories in the sciences, social sciences and humanities with remarkable growth in the last decade. In a first step, twenty categories have been selected to undergo further structural analysis in order to identify new and/or emerging topics within these subject matters. In particular, 13 fields have been selected from the sciences, 5 from the social sciences and 3 from the humanities. The underlying data have been retrieved from Thomson Reuters' Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and Arts & Humanities Citation Index (AHCI) for the period 1998-2008.

In this chapter we will briefly summarise the underlying methodology and then present the main results including validation.

Basically the analysis is a combination of different bibliometrical methods: 1. a cluster analysis of dynamically evolving fields provided in Chapter 2 and 2. a fine-grained cluster representation by sets of "core documents". Core documents are documents that have strong links with at least a given number of other documents and are thus, by definition, important nodes in the network possessing high centrality. The 3. component is the diachronic analysis, that is, the analysis of the evolution of links among clusters and topics over time. This has been done on the basis of cross-citations between core documents and clusters in different time periods.

3.2 METHODS

Cluster analysis in bibliometrics is traditionally based either on citation links (bibliographic coupling, cross-citation, co-citation analysis) or textual links (co-word analysis, term representation). Both approaches have advantages and shortcomings. The main advantage of *citation-based methods* is their discriminative power. This is

contrasted by a serious disadvantage: Citation-link matrices are extremely sparse and citation-cased methods tend to "underestimate" links among documents. Furthermore, citation links generate binary measures which are based on value 1 or 0 according as there is a citation link between two documents or not. By contrast, text-based measures are based on term frequencies in documents, which as such provide a natural weight underlying the similarity/distance measures used for the analysis. Link matrices are furthermore less sparse than their counterparts in the citation space. These advantages are abated by two serious problems: The lower discriminative power, which results in "overestimating" links among documents and the dimensionality problem. At least the latter problem can be compensated by Singular Value Decomposition (SVD) or directly by Latent Semantic Indexing (LSI), which uses the first-mentioned method.

3.2.1 *Fine mapping of emerging fields*

The link approach

In scientometrics, the link approach generally refers to all relations that are directly or indirectly created by citations. One distinguishes between direct links that are either directed (as citations to and references from a document, respectively), or cross-citations among document sets, if the direction of information flow does not matter, and indirect links, which are established by the extent of information that is jointly used by two different documents manifested by respective references (called bibliographic coupling, cf. Fano, 1956, Kessler, 1963, Glänzel and Czerwon, 1996) or citations (called co-citations, cf. Small, 1973, Marshakova, 1973). Since citations directly point to documents that are considered relevant by an author for the own research, link-based constructs tend to be strongly discriminative and to indicate rather false negative then false positive relationship. At the same time, citations are selective as they also express affinity to a certain community or even a particular "school". This property is inherited even if indirect links are used. An identified relationship based on this approach is thus reliable but often enough not all relations are recognised. These are two sides of the same coin, one of which could be considered as great advantage, the other one as severe shortcoming.

The lexical approach

The application of textual components goes back to the use of combinations of title words (Garfield, 1969), keywords (e.g., Turner et al., 1988), or subject headings (Todorov, 1992). Initially, this resulted in binary relations, i.e., merely based on presence or absence of terms and their co-occurrence. Characteristics of text-based methods were similar to those of the link approach, the results were, however, often different. Already at that time a combination of these two types of methods was proposed to overcome this discrepancy. As a result of the IT revolution storage capacity as well as CPU and processing speed dramatically increased. Recording, storing and processing abstracts and even full texts of documents has become feasible with the

effect of having to cope with a flood of textual information. Text-mining techniques, e.g., database tomography (Kostoff et al., 1997) made it possible to extract and analyse terms and phrases from large computerised corpora. Not only the mere occurrence of terms and phrases but also their frequency in the document as well as that of the documents containing these text components could be taken into account. However, scientific text is based on natural language and within the same (larger) subject the same or similar vocabulary and paradigms are in use. As a consequence, text-based similarities proved to be less discriminative (as compared with their link-based counterparts), result in less unrecognised relationships but produce more false positives than citation links.

The hybrid approach

Because of the above-mentioned properties of the link and the lexical approach, so-called hybrid solutions, that is, the combination of text- and citation-based similarities are increasingly applied to compensate their individual shortcomings and, at the same time, to synergise their advantages to make the best of the two worlds. The superiority of such hybrid methods over text-based and link-based approaches has recently been shown, among others, by Glenisson et al. (2005) and Boyack and Klavans (2010). The idea of combining citation link-based clustering and text-based approach aimed at pronouncing the advantages of the two components and, at the same time, at reducing the by-effects of their shortcomings. This was first suggested by Braam et al. (1991a,b) and Zitt and Bassecoulard (1994) in the context of improving the efficiency and usability of co-citation and co-word analysis. In particular, one of the objectives was to improve the apparently low recall of co-citation analysis concerning current work. The combination of link-based and textual methods also makes it possible to cluster objects whenever citation links are weak or missing. This feature is, above all, important in the applied sciences, most fields of the social sciences and in the humanities.

In this project, we decided to apply the following *hybrid* approach, which actually forms a combination of two important cluster techniques. The citation-based component uses *bibliographic coupling*, which has several clear advantages compared to *co-citation analysis* (cf. Glänzel and Czerwon, 1996). The most important advantages are that practically all papers have references (even in the social sciences and humanities) and that no response time needed (i.e., for citing the documents under study).

The textual component is based on term frequencies, where terms are extracted from and stemmed for titles and abstracts and stemmed for lexical use. However, the complete indexing is based on *title*, *abstract* and *keywords*; phrases are kept and stop words are removed.

Global text index vs. local concept scores

The first part of the structural topic analysis, the cluster analysis of the relatively small ISI Subject Categories proved to be challenging as all such exercises conducted at *low levels of aggregation*. Usually broader fields like medical sciences or the complete

database are clustered. The problem of huge datasets is often solved by reducing the size using sub-aggregates like journals or highly cited articles (co-citation analysis). The analysis of relatively small disciplines, however, creates different demands. Above all, the textual component is affected; terms and phrases might have different information "value" at different levels of aggregation. For instance, the terms 'algebra' and 'group' have different significance in the global and local environment. The first one provides specific information at the global level but at the local one, within the field of algebra, it is completely redundant. On the other hand, the term 'group' might only be used for global clustering as part of specific phrases like 'Lie group', 'nilpotent group', 'locally compact group', etc. Within the field of algebra, the term 'group' is practically specific enough to identify a topic. At the local level, many terms and phrases become less specific since they express common knowledge base and vocabulary.

The citation-based component is actually, less sensitive to the aggregation-level effect. Only the textual component requires adjustment for our project. This has been solved for the textual component as follows.

- We have used a global input applying *Lucene Text Index* to all documents in the complete Web of Science database (including SCIE, SSCI and AHCI) for the years 1999 and 2000.
- Within all fields, titles and abstracts have been analysed as described above. Keywords are used as they appeared (case insensitive). A term-by-document matrix has been created based on TF-IDF (term frequency–inverse document frequency), where the TF-IDF term weighting scheme is defined as follows.

$$w_{ij} = -f_{ij} \log\left(\frac{n_j}{N}\right), \tag{3.1}$$

  where $f_i j$ is the *term frequency*, i.e., the relative frequency of term t_j in document d_i and the share of documents where the term t_i appears. N denotes all documents and n_j the number of documents containing term t_j.
- A Singular Value Decomposition has been executed on the *term-by-document matrix* to reduce dimensionality and to introduce local concepts. ('algebra' is less specific, 'group' will be more appropriate, to use the above example again.) Similarities/distances have been calculated for the resulting *concept-by-document* matrix. The by-effect of obtaining negative cosine similarity in some cases resulting in angles >90° between documents had to be corrected.

The calculation of the *coupling matrix* for document couples with cosine-similarity measures from the corresponding Boolean document-reference matrix does not require any further remark. The cosine of the angle between two documents in a Boolean vector space is defined as the cosine of two Boolean vectors $\mathbf{d}_i$ and $\mathbf{d}_j$ consisting of '1' and 'o' according as the reference in question is cited or not by the paper that can be obtained from their scalar product as follows (cf. Sen and Gan, 1983)

$$cos(\mathbf{d}_i \mathbf{d}_j) = \frac{\mathbf{d}_i \cdot \mathbf{d}_j}{|\mathbf{d}_i| \cdot |\mathbf{d}_j|}. \tag{3.2}$$

This can readily be reformulated using the number of references shared by two documents (r_{ij}) and their total numbers of references (r_i and r_j, respectively).

$$cos(\mathbf{d}_i\mathbf{d}_j) = \frac{r_{ij}}{\sqrt{r_i \cdot r_j}}. \tag{3.3}$$

The combination of these two components proved, however, far from being trivial. The reason is that bibliographic coupling is based on binary counts, i.e., a reference is shared or not, while the textual component is using weighted counts as described in Eq 3.1. In principle, there are the following four major solution.

1. The *vector concatenation* was out of question because of different underlying similarity measures and other known flaws.
2. The linear combination of similarities/distances is better but sensitive to different "similarity standards" of the two components (e.g., citation component based on Boolean vector space vs. text component based on real vector space).
3. Fisher's inverse χ^2-method can better handle distances from different metrics.
4. The linear combination of angles instead of similarities/distances forms an interesting alternative and offers the opportunity of smoothly adjusting the weight of either component. In addition, it compensates the biases of different similarities/distances from different metrics.

   $$\cos x = \cos(\lambda \cdot x_1 + (1-\lambda) \cdot x_2);\ \lambda \in [0,1],$$

   where x_i ($i = 1, 2$) denote the angle between the same two documents in the citation-based (x_1) and the lexical approach x_2). Several weights are used (λ=0.667, 0.750, 0.833 and 0.875). In the sciences, we have given more weight to the citation-based component ($\lambda \geq 0.83$) than in the social sciences and humanities ($0.83 > \lambda \geq 0.67$).

Cluster analysis and core-document representation

In the previous subsection, we have described the calculation of appropriate similarity/distance measures for the cluster analysis and core-document representation. Here the structural analysis is briefly described. Hierarchical clustering (WARD agglomeration) with different stopping rules for the final number of clusters was applied on the subset of publications identified in the first phase of the project as described in Chapter 2. For obtaining the optimum number of clusters the following methods have been used.

- Local/global maxima of silhouette scores (cf. Rousseeuw, 1987)
- Dendrogram
- Evaluation of classification (based on silhouette values)
- While for global classification best TF-IDF terms and, for aggregates, PageRank can be used we had to search for alternative solutions for local classification. We used the following two methods.

- Keyword representation
- Topic representation by 'core documents'

3.2.2 *Identification of new emerging topics*

The detection of new emerging research topics or sub-disciplines within given Subject Categories is a big challenges. The reason for the theoretical, methodological and practical difficulties are quite complex: As already mentioned in Chapter 1, the most obvious problem is that scientists themselves might not always be aware that their research field is an emerging one. New ideas underlying the research, a rapidly growing number of publications and scientists dealing with this topic is a necessary, however not yet sufficient condition. The question of when such a field is considered "new emerging" is not only related to its "age" but also to the time when its literature has reached a critical mass, which is necessary to exist as and to be recognised as a independent and self-sustaining structure.

The second, rather methodological question refers to finding techniques for detecting these new emerging yet coherent structures. The use of text mining for this purpose is quite obvious. The easiest way of monitoring the emergence of research topics is certainly considering the growing frequency of specific terms within a given research area. However, textual similarity based on shared terms is also related to strong citation links. Thus Jo et al. (2007) have, for instance, analysed the correlation between the term distribution and the link distribution in the citation graph, and have found that citation connectivity is correlated with textual coupling of a term representing a topic. Shibata et al. (2008) have monitored the evolution of clusters by selecting characteristic terms for each cluster, and have shown that similar topics are strongly connected through cross-citation links while papers dealing with different topics are weakly connected. They concluded that the division of a given subject into strongly connected clusters is necessary for the detection of emergence. They also introduced two topological measures to determine the role of each paper in the citation network such that nodes with the same role should be at similar topological positions. These measures were used to decide whether there are emerging clusters.

A different dynamic approach uses sub-classification of subject domains in the sciences, social sciences and humanities, which can be done based on text mining and textual similarities between documents; extracted terms can, in addition, be used in sophisticated ways for labelling and describing the obtained clusters (cf. Lamirel et al., 2008). One possibility to monitor structural changes and the evolution of the number of clusters lies certainly in the application of incremental methods; Lamirel et al. (2010) have recently developed a diachronic multi-source approach for mining research topics.

In the first section the necessity of combining textual and citation-based approach has been stressed. The question arises of how such combinations of text- and link-based components could be applied to diachronic analysis if each of the components causes specific problems in the application to long-term analysis. The inappropriateness of the application of both bibliographic coupling and co-citation analysis over periods, say,

longer than ten years are caused by literature ageing and the genealogy of citations is self-evident. Furthermore, the use of co-citation analysis in the context of new emerging topics is questionable since it takes time before a 'critical mass' of papers on a new research topic is reached, which is needed to produce the highly cited publications needed for co-citation based clustering (see Hicks, 1987). The use of bibliographic coupling for long-term analysis, in turn, is mainly limited by the effect of citation genealogy. Citations form a kind of branching process such that the probability that the reference lists of two related papers would still strongly overlap is rather low if ten, twenty years or more have elapsed between their publication. This experience has lead us to an alternative approach which is described in the following.

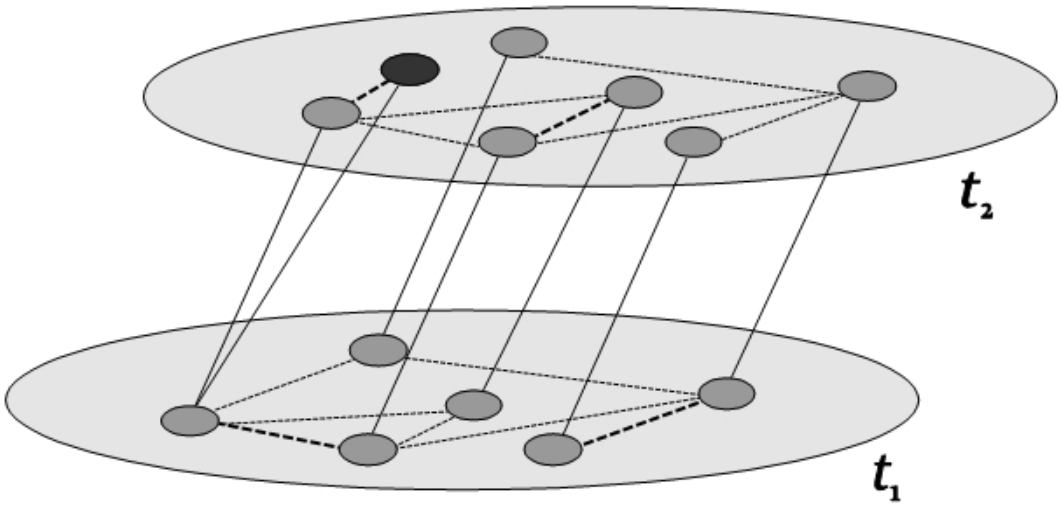

FIGURE 3.1. Sketch of a research field's changing topic structure over time
(dotted lines represent internal structures, solid lines among the time slides t1 and t2)

The basic idea of how to overcome the above-mentioned limitations of text-and citation-based methods is to subdivide the space "vertically" and "horizontally" into disjoint subsets. "Vertical" here means splitting the document space along the time axis, while "horizontal" stands for the classification by topics within an individual time slice. The use of different, disjoint time slices has the advantage that both the textual and the citation-based component will properly work in their own relatively short time window. In shorter and more recent time windows bibliographic coupling has several advantages compared to co-citation analysis (Glänzel and Czerwon, 1996), and is therefore the best candidate for the hybrid classification. The time windows do preferably not overlap in order to obtain independent structures in each time window, where, of course, the same (hybrid) clustering algorithm should be applied. After the clustering has been conducted in each time window separately, all obtained clusters or classes representing research topics in the corresponding time slice form characteristic structures defined on the basis of the underlying (hybrid) similarity measure. These structures are visualised in Figure 3.1; each time window t_1 and t_2

represents an individual document space with its own cognitive structure with, e.g., six clusters in t_1 and seven clusters in t_2, respectively. These link structure among the clusters in each time slice is visualised by dotted lines. The length of the lines stand for the distance, their thickness for the strength. While the same hybrid technique can be applied in both periods (t_1 and t_2), the determination of possible correspondence between t_1- and t_2-clusters – here indicated by solid lines between the two time slices in Figure 3.1 – remains the core problem of the detection of new emerging topics. Both sets represent disjoint document spaces and as argued above, neither bibliographic coupling nor textual analysis should be applied to long-term classification. A second problem arises from the fact that the obtained clusters are, notably at the level of local classification, usually strongly interlinked or possibly heterogeneous. In these cases, monitoring the evolution of complete clusters over different time periods and across disjoint spaces might become rather difficult. One possible solution lies in the representation of clusters or topics by sets of specific documents characterising the topics and in finding an appropriate link representation over time. The solution will be outlined in the following subsection.

Cluster representation for dynamic analysis

The question arises of how the concordance between cluster solutions found in different time periods – as sketched in Figure 3.1 – can be analysed using core-document representation. Before the procedure of detecting emerging topics is described in detail, the notion of core documents is introduced and discussed. 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 has been extended to the hybrid approach by Glänzel and Thijs (2011a). Consequently, core documents can be defined in the hybrid case analogously to the original definition by Glänzel and Czerwon (1996) as follows.
Definition: Core documents are defined as papers that have at least $n > 0$ links of at least a given strength $r > 0$ according to the underlying similarity measure.

Core documents are, by definition, strongly linked with many other documents and thus represent the most interconnected part of the network. Figure 3.2 shows the environment of a (randomly) selected core document according to Glänzel and Thijs (2012); the core document is distinguished from the other publications though its size.

Cross-citation analysis as proved means of classification of publication sets, for instance, scientific journals at a given time (cf. Leydesdorff, 2006, Zhang et al., 2009) can readily be extended to different, not necessarily overlapping time periods such as the results of independent cluster analyses in different time windows. In particular, citation links between core documents in one period to all publications in the clusters of the other period can be used to determine links between the structures of the two periods. This method is expected to reduce noise that might otherwise be caused by cross-citation links of less relevant documents far from the medoids of the clusters. On

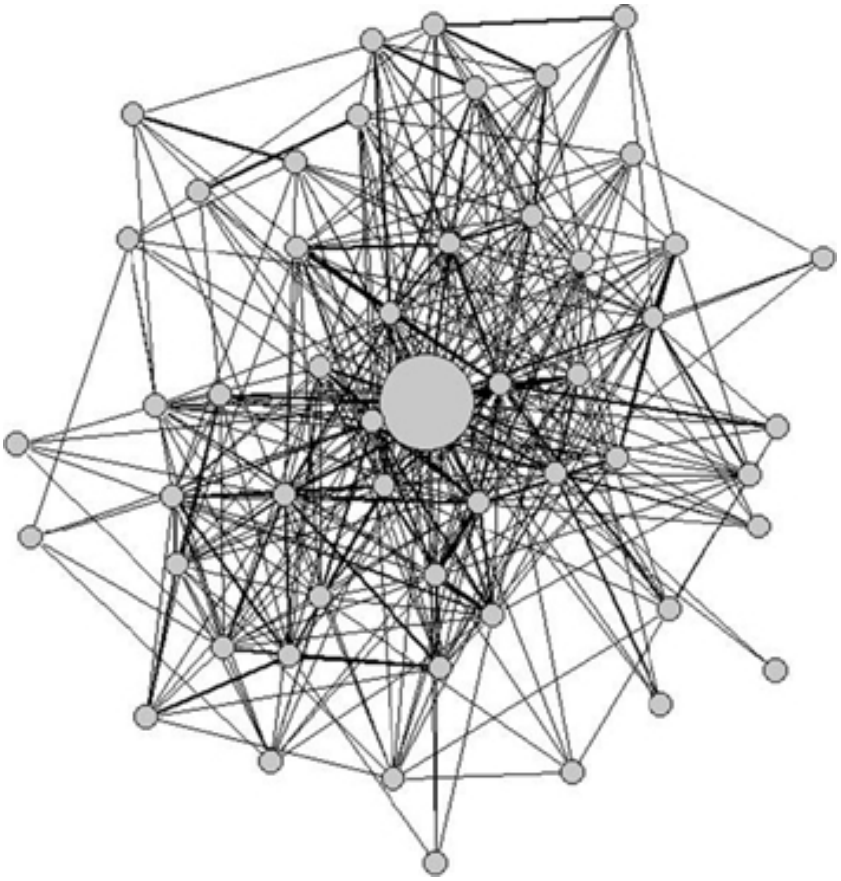

FIGURE 3.2. Visualisation of the link environment of a 'core document'
(Data source: Thomson Reuters Web of Knowledge)

the other hand, cross-citation analysis between core documents alone usually provides an insufficient number of links necessary for creating the concordance. Therefore, whether the cited or the citing side should comprise the complete cluster and the other side should be restricted to the core-document representation only. The outcomes from such direct citation analysis are sketched in Figure 3.1 (solid lines). The following step concerns the detection of new emerging topics, provided that such topics exist in the discipline under study. Before we attempt to identify candidate topics in selected fields, we have a closer look at some main characteristics of emerging topics.

Characteristics of emerging topics

The structure in the time window t_2 in Figure 3.1 shows a new cluster (marked in black), which is strongly linked with the leftmost cluster. We consider this cluster a candidate for a new topic branching off the leftmost cluster which is present in both periods. The detection of emerging topics is not merely based on the number and 'size' of occurring topics or clusters but also related to the cognitive-epistemological

structure of the research field under study. In principle, one can distinguish three paradigmatic cases of cluster evolution. These three cases could indicate new, emerging topics.

(1) Existing cluster with an exceptional growth,
(2) Completely new cluster with its root in other clusters and
(3) Existing cluster with a topic shift.

We just mention in passing that evolution can also occur in the opposite direction in case (1) and (2), say, as declining or vanishing topics.

3.3 SELECTED SUBJECT CATEGORIES AND SELECTED TOPICS

The following dynamic fields as decribes in Chapter 2 have been selected for further structural analysis.

The sciences (SCIE)

1. Nursing
2. Orthopaedics
3. Environmental sciences
4. Operations research & management science
5. Obstetrics & gynaecology
6. Medical ethics
7. Engineering, biomedical
8. Energy & fuels
9. Oncology
10. Biotechnology & applied microbiology
11. Material sciences, characterisation & testing
12. Robotics
13. Behavioural sciences

The social sciences (SSCI)

14. Public, environmental & occupational health
15. Transportation
16. Geography
17. Psychology, experimental

The arts and humanities (AHCI)

18. Archaeology
19. Religion
20. Architecture

From these disciplines, finally ten topics have been chosen out of six of the above disciplines. The cross-citation analysis of core documents and clusters over the two periods have suggested, among others, these topics, which represent one each of the above-mentioned types (1) to (3) of emerging fields. The ten topics, we have labelled according to one of the most specific keywords, are as follows.

- Engineering, biomedical
  - *Brain-Computer Interface*
  - *Kinematics*
- Energy & fuels
  - *Biodiesel*
  - *Fuel Cells*
- Environmental sciences
  - *Nano-pollution*
  - *Radiation*
- Obstetrics & gynaecology
  - *Prenatal diagnosis*
- Public, environmental & occupational health
  - *Aids in Africa*
  - *Environment*
- Geography
  - *State & Region*

3.4 RESULTS

3.4.1 *Introduction*

The emerging topics

The cluster analysis of the six selected fields has been conducted for two 5-year periods each (1999-2003 and 2004-2008) covering the full 10-year period 1999-2008.

The clusters provided by the hybrid analysis for both time windows t_1 and t_2 based on the corresponding silhouette values (cf. Rousseeuw, 1987) are presented in the first figure of each discipline. Both structures are presented in the same diagram using Pajek (Batagelj and Mrvar, 2003). Clusters of the first period are marked in red, those of the second time window in green. The size of the circles is proportional to the number of papers in the clusters. One specific keyword each has been chosen to label the clusters. The selection of this label was arbitrary and only designed to facilitate the identification on the map (Glänzel and Thijs, 2011a,b, 2012).

The description of each of the ten topics is given by *keywords* and *core documents* in separate tables.

1. *Keywords*. This table contains the most specific keywords for all documents in the selected cluster. The keywords contain both the author keywords as well as the keyword-plus (additional keywords added by Thomson Reuters, provider of the database). These data are presented without modification.
2. *Core Documents*. This table contains the title of the core documents. These documents are selected based on their large number of strong links with other documents in the cluster as described in the methodological section. These titles give a good description of the content of the topic.

National contribution and international co-operation in emerging topics

To place the identified fields and topics into an international context, additional analyses were performed within ERACEP aiming at answering the following questions: Which countries are the most active players in the emerging topics and to what extent do they collaborate. In order to answer this question we have analysed both the publication activity and scientific collaboration in the above ten topics *and* their "mother disciplines" according to the six selected ISI Categories. The results are presented in the following manner. Tables first provide the keyword and core-document representation of the corresponding topic (cf. Glänzel and Thijs, 2012). Further tables present the percentage of the most active countries in the world total in the corresponding Subject Category in the period 2004-2008 along with the same data in the emerging topic(s) belonging to the corresponding discipline. Figures show first the topic structure of the corresponding discipline in both periods (cf. Glänzel and Thijs, 2012); the following figures compare the international collaboration network within the Subject Category with that in the emerging topics (cf. Glänzel and Thijs, 2011b).

3.4.2 *Biomedical Engineering*

Thomson Reuters describes the field as follows (TR, 2011a):

> 'Biomedical Engineering covers resources that apply engineering technology to solving medical problems. Resources in this category span a wide range of applications including applied biomechanics, biorheology, medical imaging, medical monitoring equipment, artificial organs, and implanted materials and devices.'

Cluster analyses of the two distinct periods resulted in the identification of eight (first period) and nine (second period) topics, respectively.

Emerging topics: Keywords and Core documents

Two emerging topics are detected. The first topic is labeled *'Brain Computer Interface'*. This topic is already present in the first period but an extensive growth can be observed. Although the phrase 'Brain-Computer Interface' does not occur among the 25 most important keywords, acronyms (BCI) and variants of this phrase (e.g., brain machine interface, asynchronous brain interface) were among the most frequent terms in the titles of the core documents. The second topic is labeled *'Kinematics'*. This completely new topic emerged with links to *'Imaging'*, *'Cartillage'* and *'Bone Cement'*. Most relevant themes are joint and muscle kinematics during motion and the corresponding models. This topic is much smaller than the first topic.

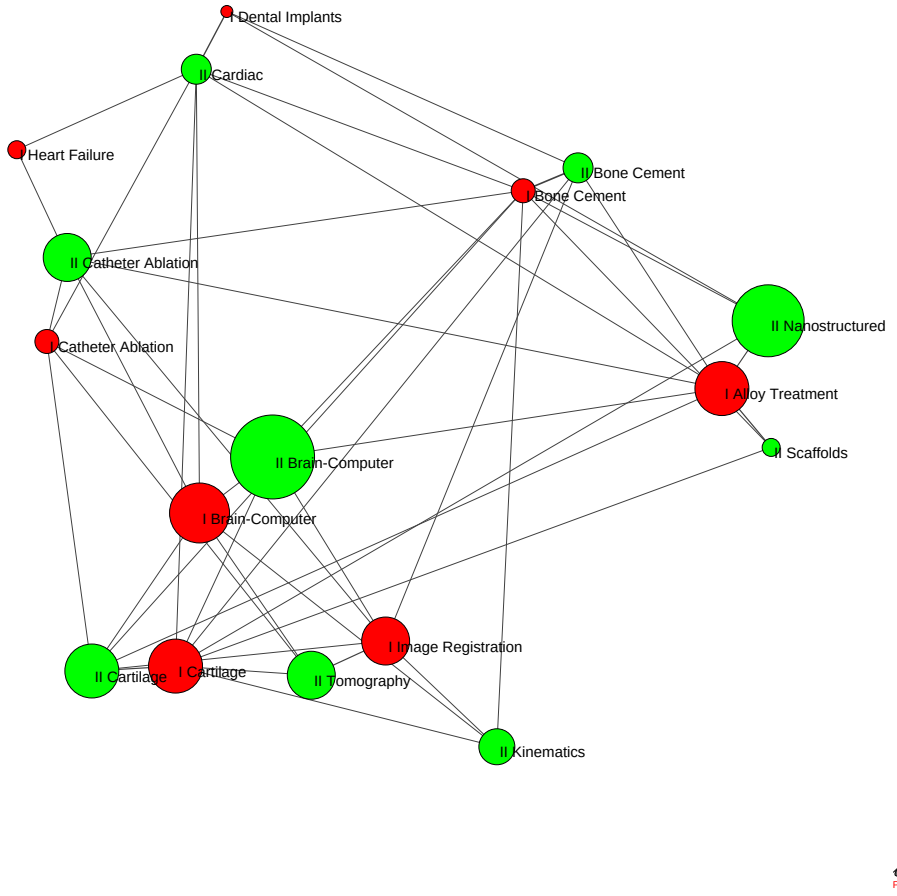

FIGURE 3.3. Biomedical Engineering: relations between clusters
(Data source: Thomson Reuters Web of Knowledge

Brain Computer Interface

Keywords (n = 5638 – 19.4% of the complete discipline)

signals; blood flow; classification; eeg; neural networks; pattern recognition; independent component analysis; hemodynamics; patterns; elasticity; elastography; models; simulation; velocity; identification; time series

Core Documents

ISI UT-code	Document title
000188541100012	"Virtual keyboard" controlled by spontaneous EEG activity
000189183300028	Planar gradiometer for magnetic induction tomography (MIT): theoretical and experimental sensitivity maps for a low-contrast phantom
000220967700004	Adaptive BCI based on variational Bayesian Kalman filtering: An empirical evaluation
000221578000008	Model-based neural decoding of reaching movements: A maximum likelihood approach
000221578000010	Ascertaining the importance of neurons to develop better brain-machine interfaces
000221578000016	Anasynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate
000221578000018	Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms
000221578000019	Support vector channel selection in BCI
000221578000021	Classification of single-trial electroencephalogram during finger movement
000221578000023	BCI2000: A general-purpose, brain-computer interface (BCI) system
000221578000024	The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials
000221578000027	BCI competition 2003 - Data set IIa: Spatial patterns of self-controlled brain rhythm modulations
000221578000029	BCI competition 2003 - Data set IIb: Support vector machines for the P300 speller paradigm
000227747000009	Closed-loop cortical control of direction using support vector machines
000228563700029	A new type of gradiometer for the receiving circuit of magnetic induction tomography
000229850800016	Interpreting spatial and temporal neural activity through. a recurrent neural network brain-machine interface
000231268900006	Spatio-spectral filters for improving the classification of single trial EEG
000231969500013	Sensorimotor rhythm-based brain-computer interface (BCI): Feature selection by regression improves performance
000232112400009	Resonance behaviour of whole-body averaged specific energy absorption rate (SAR) in the female voxel model, NAOMI
000232193200018	Automated methodology for determination of stress distribution in human abdominal aortic aneurysm
000233865100007	A patient-specific computational model of fluid-structure interaction in abdominal aortic aneurysms
000236519000004	Robust classification of EEG signal for brain-computer interface
000236519000005	Steady-state somatosensory evoked potentials: Suitable brain signals for brain-computer interfaces?
000237696200002	Towards adaptive classification for BCI
000237851500029	A fully on-line adaptive BCI
000238294000006	The constitutive properties of the brain parenchyma Part 1. Strain energy approach
000238394700008	The BCI competition III: Validating alternative approaches to actual BCI problems
000238394700028	A practical VEP-based brain-computer interface
000238546800015	User customization of the feature generator of an asynchronous brain interface
000239263400024	A method for detection and classification of events in neural activity
000239673500009	A comparison of optimal MIMO linear and nonlinear models for brain-machine interfaces
000240725300004	Non-invasive determination of zero-pressure geometry of arterial aneurysms
000241242900001	A planar biaxial constitutive relation for the luminal layer of intra-luminal thrombus in abdominal aortic aneurysms
000241536900017	Combined optimization of spatial and temporal filters for improving brain-computer interfacing
000243122900009	A comparison of classification techniques for the P300 Speller
000243270000011	Regional distribution of wall thickness and failure properties of human abdominal aortic aneurysm
000243621200008	Study of discriminant analysis applied to motor imagery bipolar data
000243778300013	Effects of wall calcifications in patient-specific wall stress analyses of abdominal aortic aneurysms
000244241200026	Inverse elastostatic stress analysis in pre-deformed biological structures: Demonstration using abdominal aortic aneurysms
000244498800020	Preprocessing and meta-classification for brain-computer interfaces
000244498800026	Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces
000245565400015	Patient-specific initial wall stress in abdominal aortic aneurysms with a backward incremental method
000245872700006	Combining spatial filters for the classification of, single-trial EEG in a finger movement task
000247947300003	A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals
000251908300031	Generalized features for electrocorticographic BCIs
000251967300001	A comparison of modelling techniques for computing wall stress in abdominal aortic aneurysms
000255744800004	The effects of anisotropy on the stress analyses of patient-specific abdominal aortic aneurysms
000256196100001	QRS complex detection using double density discrete wavelet transform
000257253800012	Control of a humanoid robot by a noninvasive brain-computer interface in humans
000257953000002	Local temporal common spatial patterns for robust single-trial EEG classification

Kinematics

Keywords (n = 2639 – 9.1% of the complete discipline)

muscle; knee; biomechanics; force; forces; joint; motion; stability; kinematics; reliability; movement; rehabilitation; injury; patterns; skeletal muscle; hip; parameters; stroke; performance; spine

Core Documents

ISI UT-code	Document title
000188948300011	Assessment of the functional method of hip joint center location subject to reduced range of hip motion
000188948300019	Estimation of the centre of rotation: a methodological contribution
000221894400014	The effects of muscle stretching and shortening on isometric forces on the descending limb of the force-length relationship
000222157700004	Differences in normal and perturbed walking kinematics between male and female athletes
000224048600009	Sagittal plane biomechanics cannot injure the ACL during sidestep cutting
000225198500013	A new method for estimating joint parameters from motion data
000225321800006	Influence of gender on hip and knee mechanics during a randomly cued cutting maneuver
000225634400005	Scapular kinematics during transfers in manual wheelchair users with and without shoulder impingement
000226884200028	Determination of patient-specific multi-joint kinematic models through two-level optimization
000228803500001	Force enhancement following stretch of activated muscle: critical review and proposal for mechanisms
000229955800005	Scapular kinematics during humeral elevation in adults and children
000231324900014	Association between lower extremity posture at contact and peak knee valgus moment during sidestepping: Implications for ACL injury
000234009000005	A comparison of dynamic coronal plane excursion between matched male and female athletes when performing single leg landings
000234009000006	The influence of gender on knee kinematics, kinetics and muscle activation patterns during side-step cutting
000234117900006	Biomechanical characterization and clinical implications of artificially induced toe-walking: Differences between pure soleus, pure gastrocnemius and combination of soleus and gastrocnemius contractures
000235330500005	Relationship between hip and knee strength and knee valgus during a single leg squat
000236834900013	An optimized protocol for hip joint centre determination using the functional method
000236834900014	Using computed muscle control to generate forward dynamic simulations of human walking from experimental data
000236981400004	Metal-on-metal hip joint tribology
000239379300011	The influence of experience on knee mechanics during side-step cutting in females
000240166900005	Differences in muscle function during walking and running at the same speed
000241489500006	3-D scapular kinematics during arm elevation: Effect of motion velocity
000241489500012	The effect of an impulsive knee valgus moment on in vitro relative ACL strain during a simulated jump landing
000241850100009	Muscles that support the body also modulate forward progression during walking
000242433700008	A survey of formal methods for determining the centre of rotation of ball joints
000243456700001	Force depression following muscle shortening in sub-maximal voluntary contractions of human adductor pollicis
000243868700014	A neuromusculoskeletal tracking method for estimating individual muscle forces in human movement
000244241200013	Prediction of the hip joint centre in adults, children, and patients with cerebral palsy based on magnetic resonance imaging
000245298000007	Modeling initial contact dynamics during ambulation with dynamic simulation
000245872700002	Are patient-specific joint and inertial parameters necessary for accurate inverse dynamics analyses of gait?
000246004500006	Propagation of the hip joint centre location error to the estimate of femur vs pelvis orientation using a constrained or an unconstrained approach
000246898200017	Predictive modelling of human walking over a complete gait cycle
000247558300010	Sex differences in lower extremity biomechanics during single leg landings
000248677300012	Loading characteristics of females exhibiting excessive valgus moments during cutting
000249067900016	Design of patient-specific gait modifications for knee osteoarthritis rehabilitation
000249508900016	Robust estimation of dominant axis of rotation
000251342800026	A framework for the functional identification of joint centers using markerless motion capture, validation for the hip joint
000252889300008	Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors
000253062100018	In vivo measurement of dynamic rectus femoris function at postures representative of early swing phase
000253384500003	A new technique estimating location of mean joint centers of rotation with associated dispersions and assessing to functional coordinate system of moving limb segments
000253435400003	Three-dimensional scapular kinematics during the throwing motion
000254677400018	Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels
000254792800008	Influence of trunk flexion on hip and knee joint kinematics during a controlled drop landing
000255783000018	Research of variations of frequency of non-linear dynamic model of the muscle
000256734000007	Effects of a knee extension constraint brace on selected lower extremity motion patterns during a stop-jump task
000256769200014	Improving joint torque calculations: Optimization-based inverse dynamics to reduce the effect of motion errors
000257278900011	Influences of hip external rotation strength on knee mechanics during single-leg drop landings in females
000258369300013	Fatigue-related changes in stance leg mechanics during sidestep cutting maneuvers
000259129000031	Effects of suppressing arm swing on kinematics, kinetics, and energetics of human walking
000259401600003	Gender differences in lower limb frontal plane kinematics during landing

National contribution

Country	Complete Field	Brain Computer Interface	Share of publications: Kinematics
Belgium	1.38%	1.47%	1.64%
Brazil	1.6%	1.69%	1.26%
Denmark	0.93%	1.67%	0.99%
France	4.01%	4.28%	4.16%
Germany	7.72%	6%	6.29%
Greece	1.17%	1.74%	1.03%
India	1.04%	0.85%	0.04%
Italy	5.56%	5.26%	4.54%
Japan	6.86%	3.92%	4.54%
Netherlands	3.79%	3.48%	4.81%
PR China	6.02%	4.55%	3.05%
Poland	0.96%	1.26%	0.72%
Spain	2.53%	3.68%	1.33%
Sweden	1.84%	2.11%	1.64%
United Kingdom	8.78%	9.69%	11.14%
USA	35.35%	35.33%	40.12%

International co-operation

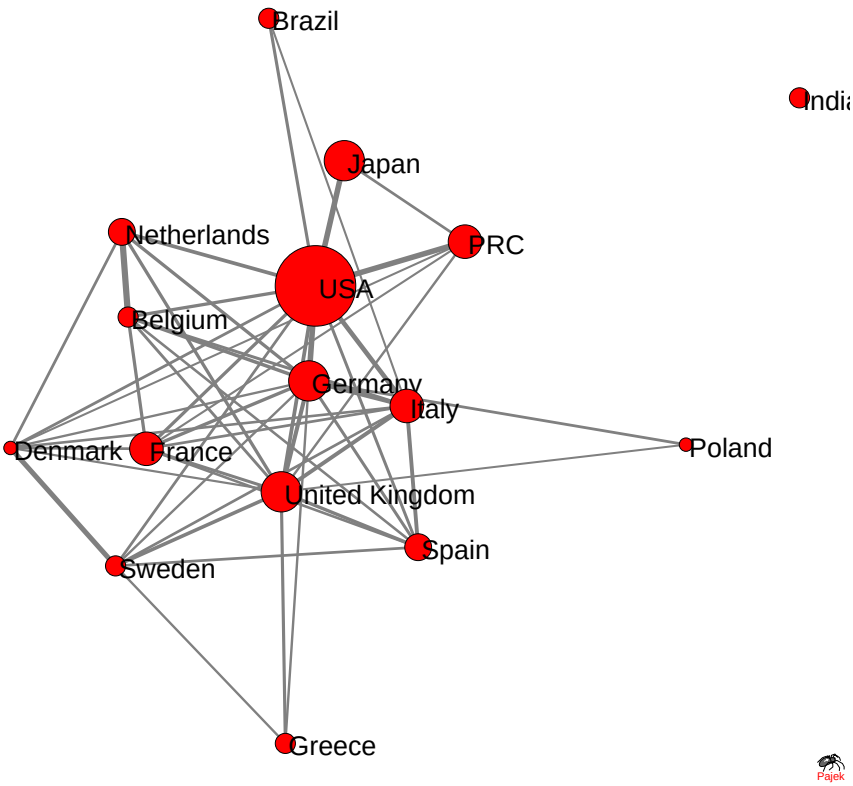

FIGURE 3.4. Biomedical Engineering: Complete field
(Data source: Thomson Reuters Web of Science)

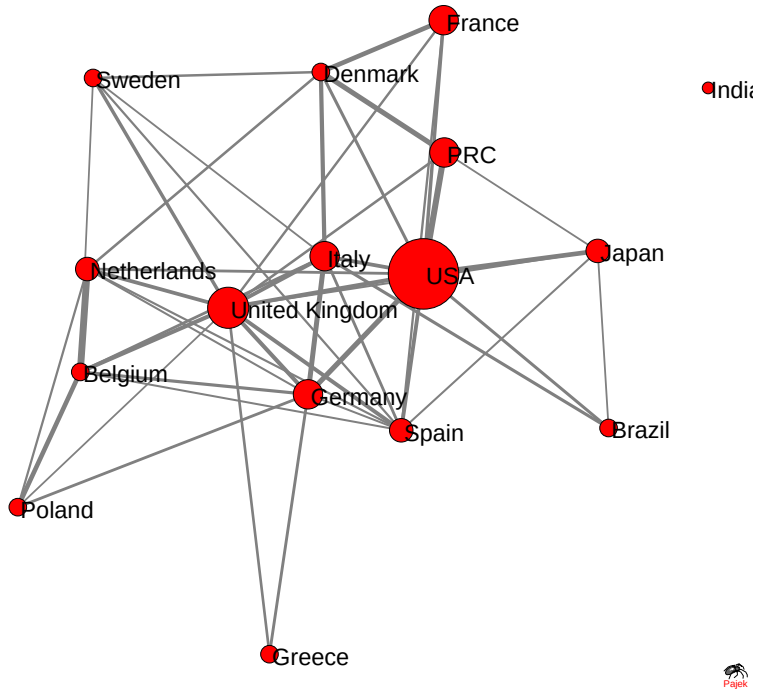

FIGURE 3.5. Biomedical Engineering: Brain Computer Interface
(Data source: Thomson Reuters Web of Science)

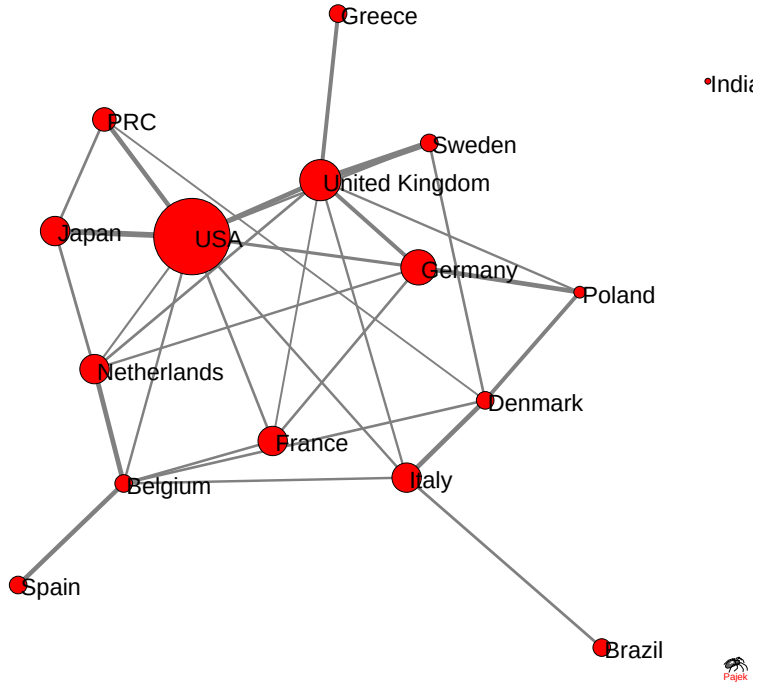

FIGURE 3.6. Biomedical Engineering: Kinematics
(Data source: Thomson Reuters Web of Science)

The analysis of the different national contributions to the field "biomedical engineering" and to the two emerging topics "brain-computer interface" and "kinematics" clearly indicates a leading role of the United States. Their contribution to the whole field is above 35 per cent. A similar position can be observed in the emerging topic "brain-computer interface". The influence of the United States is even more prominent in the emerging topic "kinematics" with a share of about 40 per cent of all publications. China and Japan are interesting cases where the contribution to the two emerging topics is clearly below their contribution to the whole field, which seems to indicate that the attention in these countries to the two emerging topics is less expressed. Main European contributions to the whole field and also to the emerging topics are coming from Germany, France, Italy, and the United Kingdom. Among these, in particular the United Kingdom seems to put more emphasis on our two emerging topics "brain-computer interface" and "kinematics" compared to the role of the United Kingdom in the total field. The cooperation networks for the dynamic field biomedical engineering as a whole and for the two emerging topics present rather dense cooperation patterns within Europe and between European countries and the United States. The Asian countries seem not to be integrated very well into these networks.

3.4.3 *Energy & Fuels*

Thomson Reuters describes the field as follows (TR, 2011a):

> 'Energy & Fuels covers resources on the development, production, use, application, conversion, and management of nonrenewable (combustible) fuels (such as wood, coal, petroleum, and gas) and renewable energy sources (solar, wind, biomass, geothermal, hydroelectric). Note: Resources dealing with nuclear energy and nuclear technology appear in the NUCLEAR SCIENCE & TECHNOLOGY category.'

Cluster analyses of the two distinct periods resulted in the identification of seven (first period) and eight (second period) topics.

Emerging topics: Keywords and Core documents

Two emerging topics are detected. The first topic is labeled *'Biodiesel'*. This topic is related to the topic on *'Diesel'* and other petrol derived fuels from first period. But an extensive growth can be observed and a shift towards non-petrol fuels. The second topic is labeled *'Fuel Cells'*. The seed for this topic lies within the topic on *'Solar Cells and Renewable Energy'* from the first period.

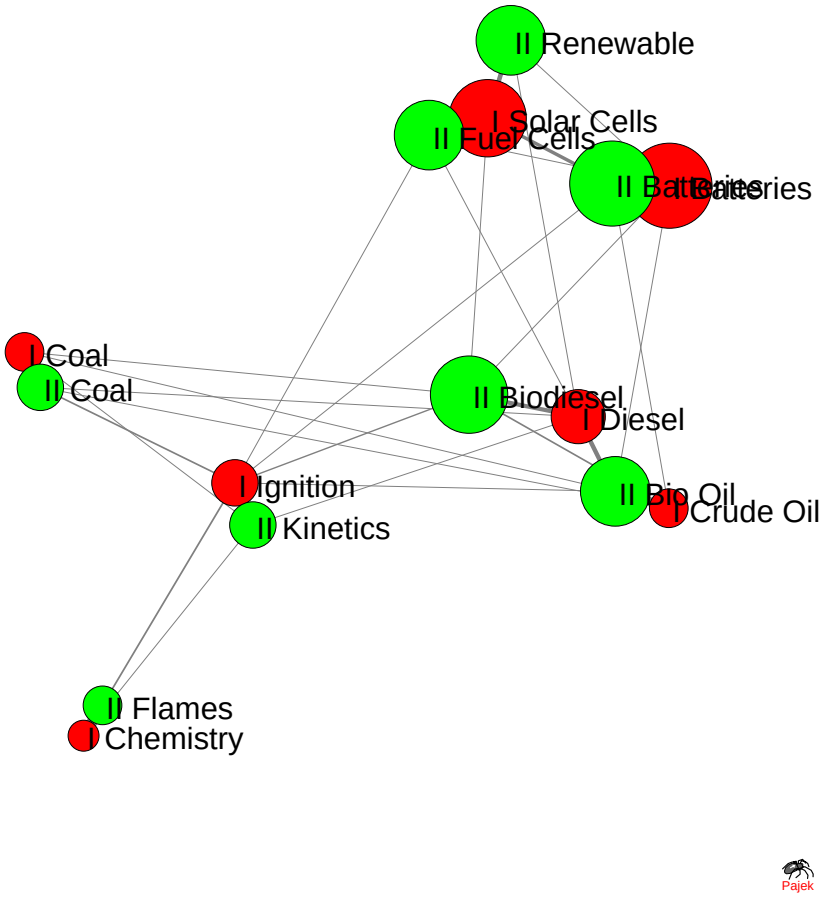

FIGURE 3.7. Energy & Fuels: relations between clusters
(Data source: Thomson Reuters Web of Science

Biodiesel

Keywords (n = 5056 – 17.3% of the complete discipline)

ammonia; biogas; start up; anaerobic digestion; recovery; mechanical properties; carbon dioxide; saccharomyces cerevisiae; biodegradation; synthesis gas; corn stover; mass transfer; enzymatic hydrolysis; selectivity; heat transfer; porous media; silica; crystallization; fischer tropsch synthesis; deactivation

Core Documents

ISI UT-code	Document title
000234152800004	Process optimization for biodiesel production from mahua (Madhuca indica) oil using response surface methodology
000236168700053	Continuous production of biodiesel via transesterification from vegetable oils in supercritical methanol
000236507000004	Temperature effects on biohydrogen production in a granular sludge bed induced by activated carbon carriers
000236683400003	Biohydrogen generation from jackfruit peel using anaerobic contact filter
000237177600012	Biohydrogen-production from beer lees biomass by cow dung compost
000237277000006	Isolation of hydrogen generating microflora from cow dung for seeding anaerobic digester
000237639400004	Fermentative hydrogen production from xylose using anaerobic mixed microflora
000237639400019	Sulfate effect on fermentative hydrogen production using anaerobic mixed microflora
000238277200023	Biodiesel production via non-catalytic SCF method and biodiesel fuel characteristics
000238921700006	Biological hydrogen production in suspended and attached growth anaerobic reactor systems
000239101200004	Biohydrogen generation from palm oil mill effluent using anaerobic contact filter
000239101200006	Continuous bio-hydrogen production from citric acid wastewater via facultative anaerobic bacteria
000241127800005	Improving biohydrogen production in a carrier-induced granular sludge bed by altering physical configuration and agitation pattern of the bioreactor
000241483200014	Encapsulation of polymer photovoltaic prototypes
000242031900004	Evaluation of alternative methods of preparing hydrogen producing seeds from digested wastewater sludge
000243291800007	Fermentative hydrogen production with a draft tube fluidized bed reactor containing silicone-gel-immobilized anaerobic sludge
000244071700005	Microbial hydrogen production with Bacillus coagulans IIT-BT S1 isolated from anaerobic sewage sludge
000244788100004	Biodiesel production facilities from vegetable oils and animal fats
000244997900012	Transparent anodes for polymer photovoltaics: Oxygen permeability of PEDOT
000246113600005	Relationships between viscosity and density measurements of biodiesel fuels
000246671600006	Large-area photovoltaics based on low band gap copolymers of thiophene and benzothiadiazole or benzo-bis(thiadiazole)
000246746800007	Synthesis and photovoltaic properties of soluble fulleropyrrolidine derivatives for organic solar cells
000247500900004	Continuous hydrogen production by anaerobic mixed microflora using a hollow-fiber microfiltration membrane bioreactor
000247500900020	Comparative performance of two upflow anaerobic biohydrogen-producing reactors seeded with different sludges
000247728700004	Production of biodiesel from vegetable oils: A survey
000248022300006	Optimization of process parameters for high-efficiency polymer photovoltaic devices based on P3HT : PCBM system
000251463200028	Feasibility study on fermentative conversion of raw and hydrolyzed starch to hydrogen using anaerobic mixed microflora
000251491200012	Fermentative hydrogen gas production using biosolids pellets as the inoculum source
000251676400007	Mathematical relationships derived from biodiesel fuels
000252259500017	Production of hydrogen in a granular sludge-based anaerobic continuous stirred tank reactor
000252259500018	Bio-hydrogen generation from mixed fruit peel waste using anaerobic contact filter
000252434700014	Continuous biohydrogen production from starch with granulated mixed bacterial microflora
000252434700018	Dark fermentative hydrogen production from xylose in different bioreactors using sewage sludge microflora
000253791300008	Simultaneous biohydrogen production and wastewater treatment in biofilm configured anaerobic periodic discontinuous batch reactor using distillery wastewater
000253791300012	Heavy metal effects on fermentative hydrogen production using natural mixed microflora
000253791300013	Batch dark fermentative hydrogen production from grass silage: The effect of inoculum, pH, temperature and VS ratio
000255116500035	Relationships derived from physical properties of vegetable oil and biodiesel fuels
000255271900009	Lower emissions from biodiesel combustion
000255272000010	Effect of alkali on liquid yields from the pyrolysis of olive oil
000255603600013	HRT-dependent hydrogen production and bacterial community structure of mixed anaerobic microflora in suspended, granular and immobilized sludge systems using glucose as the carbon substrate
000255603600016	Exploring optimal environmental factors for fermentative hydrogen production from starch using mixed anaerobic microflora
000255603600022	Hydrogen economy in Taiwan and biohydrogen
000256008600003	Prediction of higher heating values for saturated fatty acids from their physical properties
000257009900022	Technoeconomic study of supercritical biodiesel production plant
000257150800045	Hydrogen production from sewage sludge using mixed microflora inoculum: Effect of pH and enzymatic pretreatment
000257150800079	Assessment of four biodiesel production processes using HYSYS. Plant
000258518000005	Continuous biohydrogen production in a CSTR using starch as a substrate
000258722000008	Production of biodiesel from tall oil
000259958700014	Effects of different pretreatment methods on fermentation types and dominant bacteria for hydrogen production
000261115300022	Enhanced cellulose-hydrogen production from corn stalk by lesser panda manure

Fuel Cells

Keywords (n = 4135 – 14.2% of the complete discipline)

transport; performance; fuel cell; operation; fuel cells; platinum; layer; pem fuel cell; management; methanol; mathematical model; exchange; conductivity; membrane; membranes; fabrication; cathode; anode; catalysts; gas turbine

Core Documents

ISI UT-code	Document title
000234363900005	Multidimensional modelling of polymer electrolyte fuel cells under a current density boundary condition
000234962600016	Ultra large-scale simulation of polymer electrolyte fuel cells
000236168700042	A three-dimensional, multicomponent, two-phase model for a proton exchange membrane fuel cell with straight channels
000236310800002	Evaluation of the thickness of membrane and gas diffusion layer with simplified two-dimensional reaction and flow analysis of polymer electrolyte fuel cell
000236310800003	Evaluation of the optimal separator shape with reaction and flow analysis of polymer electrolyte fuel cell
000236310800012	Ex situ visualization of liquid water transport in PEM fuel cell gas diffusion layers
000237003000016	A two-phase flow and transport model for PEM fuel cells
000237844700007	Optimization of the operating parameters of a proton exchange membrane fuel cell for maximum power density
000237946400010	PEM fuel cell dynamic model with phase change effect
000238412100015	Numerical study of reactant gas transport phenomena and cell performance of proton exchange membrane fuel cells
000238412100030	Predicting the transient response of a serpentine flow-field PEMFC I. Excess to normal fuel and air
000238412100036	Main and interaction effects of PEM fuel cell design parameters
000238964200004	Three-dimensional transport model of PEM fuel cell with straight flow channels
000240393500010	Numerical modeling of proton exchange membrane fuel cell with considering thermal and relative humidity effects on the cell performance
000240842800062	A two dimensional partial flooding model for PEMFC
000240959500019	Transient analysis for the cathode gas diffusion layer of PEM fuel cells
000240959500042	Modeling non-isothermal two-phase multicomponent flow in the cathode of PEM fuel cells
000241012000089	Humidity of reactant fuel on the cell performance of PEM fuel cell with baffle-blocked flow field designs
000241012000095	Dynamic analysis of gas transport in cathode side of PEM fuel cell with interdigitated flow field
000241067200035	Effects of two-phase transport in the cathode gas diffusion layer on the performance of a PEMFC
000241067200044	Numerical investigation of transport component design effect on a proton exchange membrane fuel cell
000241067200046	Parameter sensitivity examination and discussion of PEM fuel cell simulation model validation Part I. Current status of modeling research and model development
000241781200029	Effects of cathode humidification on the gas-liquid interface location in a PEM fuel cell
000241781200049	Performance/design formulation for a solid polymer based acid electrolyte hydrogen/air fuel cell
000242059700050	A three-dimensional PEM fuel cell model with consistent treatment of water transport in MEA
000243650800014	Effect of plasma nitriding on behavior of austenitic stainless steel 304L bipolar plate in proton exchange membrane fuel cell
000243650800019	Metal bipolar plates for PEM fuel cell - A review
000244359500033	A three-dimensional mixed-domain PEM fuel cell model with fully-coupled transport phenomena
000244772200004	Review of materials and characterization methods for polymer electrolyte fuel cell flow-field plates
000244922600001	Parametric and optimization study of a PEM fuel cell performance using three-dimensional computational fluid dynamics model
000245093800024	Transient computation fluid dynamics modeling of a single proton exchange membrane fuel cell with serpentine channel
000246745700029	A two-phase non-isothermal mixed-domain PEM fuel cell model and its application to two-dimensional simulations
000247359600002	Corrosion study on different types of metallic bipolar plates for polymer electrolyte membrane fuel cells
000247418100010	Innovative gas diffusion layers and their water removal characteristics in PEM fuel cell cathode
000249261700013	Thermal-fluid-dynamic simulation of a proton exchange membrane fuel cell using a hierarchical 3D-1D approach
000249261700018	Simplified model to predict incipient flooding/dehydration in proton exchange membrane fuel cells
000250066900059	Effects of temperature on the location of the gas-liquid interface in a PEM fuel cell
000250066900060	Operation-relevant modeling of an experimental proton exchange membrane fuel cell
000250066900061	Numerical investigation of transient responses of a PEM fuel cell using a two-phase non-isothermal mixed-domain model
000250490600034	Numerical study on PEMFC's geometrical parameters under different humidifying conditions
000250995200042	A two-dimensional steady state model including the effect of liquid water for a PEM fuel cell cathode
000251196800007	Modeling optimizes PEM fuel cell performance using three-dimensional multi-phase computational fluid dynamics model
000251463200021	Growth of Cr-Nitrides on commercial Ni-Cr and Fe-Cr base alloys to protect PEMFC bipolar plates
000251591000028	Protective nitride formation on stainless steel alloys for proton exchange membrane fuel cell bipolar plates
000252634300004	Estimation of flooding in PEMFC gas diffusion layer by differential pressure measurement
000253564000022	Three-dimensional numerical study on cell performance and transport phenomena of PEM fuel cells with conventional flow fields
000255229200010	Two-phase transport in PEM fuel cell cathodes
000256177600002	Transport mechanisms and performance simulation of a PEM fuel cell
000257642500016	Effect of capillary pressure on liquid water removal in the cathode gas diffusion layer of a polymer electrolyte fuel cell
000258993900001	Review of computational heat and mass transfer modeling in polymer-electrolyte-membrane (PEM) fuel cells

National contribution

Country	Complete Field	Biodiesel	Fuel Cells
Belgium	0.68%	0.71%	0.65%
Brazil	1.79%	1.98%	2.18%
Denmark	1.09%	1.06%	0.82%
France	4.34%	4.65%	3.85%
Germany	4.44%	4.62%	5.93%
Greece	1.58%	1.7%	1.33%
India	5.18%	5.96%	3.36%
Italy	2.78%	3.16%	4.59%
Japan	5.23%	7.07%	6.09%
Netherlands	1.65%	1.33%	0.85%
PR China	12.89%	17.25%	15.48%
Poland	1.16%	1.37%	0.68%
Spain	3.84%	3.71%	2.52%
Sweden	1.95%	1.32%	1.11%
United Kingdom	4.88%	4.35%	3.72%
USA	17.61%	14.68%	21.57%

International co-operation

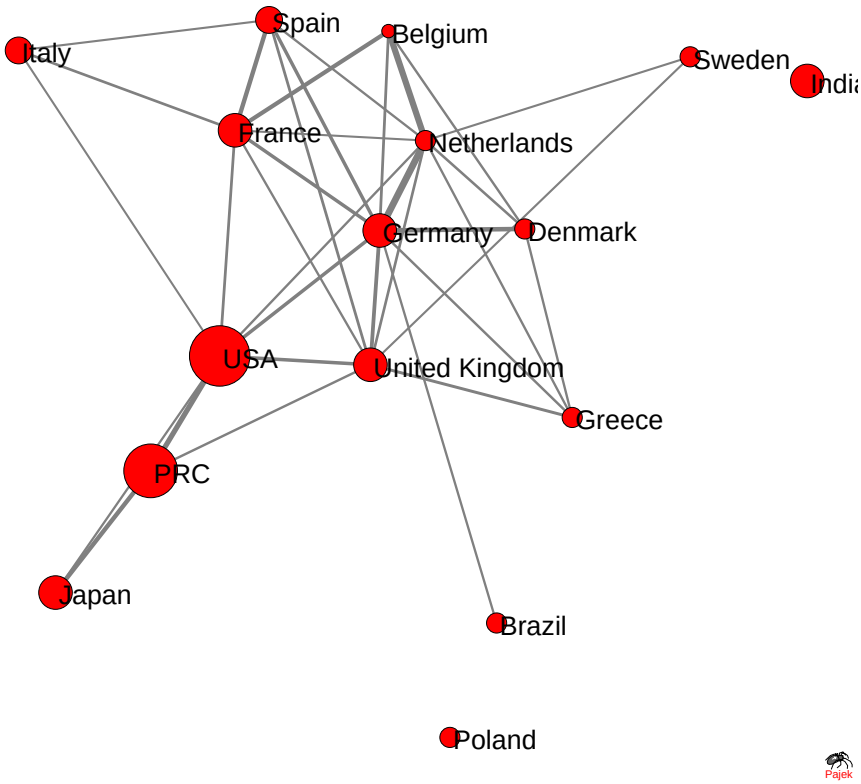

FIGURE 3.8. Energy & Fuels : Complete field
(Data source: Thomson Reuters Web of Science)

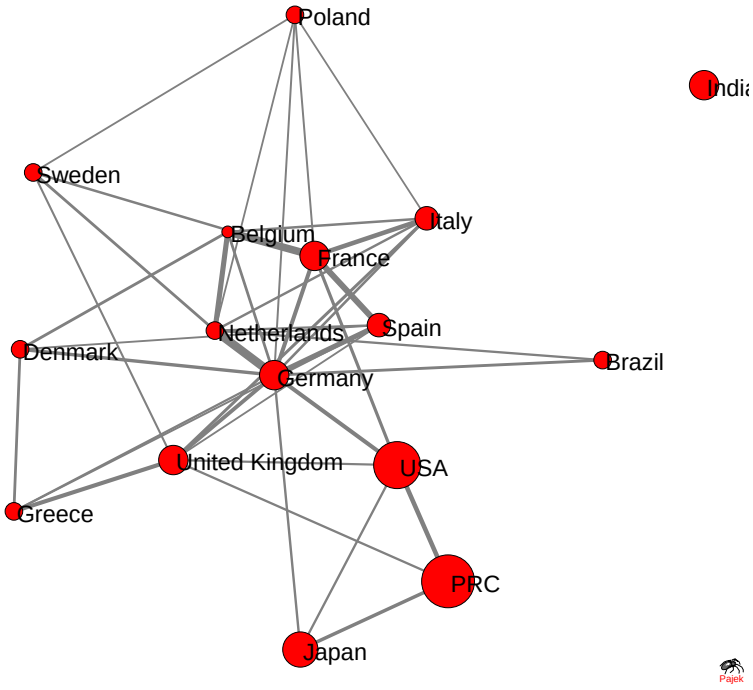

FIGURE 3.9. Energy & fuels : Biodiesel
(Data source: Thomson Reuters Web of Science)

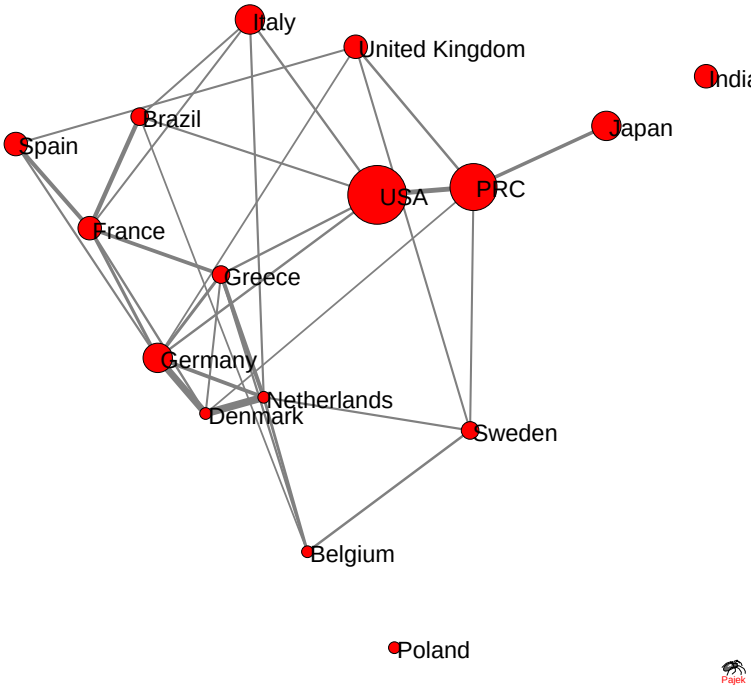

FIGURE 3.10. Energy & fuels : Fuel Cells
(Data source: Thomson Reuters Web of Science)

The international analysis shows that the United States are contributing most publications to the whole dynamic field of "energy and fuels". However, with roughly 18 per cent of all publications their position is by far less dominant compared to "biomedical engineering". Looking at the two emerging topics "biodiesel" and "fuel cells", the United States seem to be much more specialised in the latter one. China is playing a remarkable role in the "energy and fuels" topics contributing almost 13 per cent to the whole field. In the case of "biodiesel" China is the most active country in terms of publication output within our sample contributing a share of more than 17 per cent. Also "fuel cells" is a topic where China is playing an important role and contributes the second highest share of publications after the United States. Within the European countries France, Germany, Spain, and the United Kingdom are playing a similar role in terms of publications targeting the whole field of "energy and fuels". Germany and the United Kingdom are focusing more on "fuel cells", while the other countries seem to have a stronger focus on "biodiesel" which might be explained by the higher significance of agricultural production in these countries.

The cooperation networks in the field "energy and fuels" and the emerging topics "biodiesel" and "fuel cells" are looking quite differently compared to biomedical engineering. In the whole field we find a well-expressed European cluster. The United States and also China and Japan are taking more peripheral positions. Interestingly, there seem to be quite intense cooperations between the United States and China. A similar situation holds true for "biodiesel". Again, we observe a strong European cluster and established cooperation network between the USA and China. Within the emerging topic "fuel cells" the United States and China are moving closer to the European network. Within Europe we find very strong cooperation links in central Europe in the case of "biodiesel" extending to Spain. In the case of "fuel cells" cooperation intensities are strongest between Germany, the Netherlands, and Denmark.

3.4.4 *Environmental Sciences*

Thomson Reuters describes the field as follows (TR, 2011a):

> 'Environmental Sciences covers resources concerning many aspects of the study of the environment, among them environmental contamination and toxicology, environmental health, environmental monitoring, environmental geology, and environmental management. This category also includes soil science and conservation, water resources research and engineering and climate change.'

The complete field of 'Environmental Sciences'experienced an extensive growth over the last decade. Cluster analyses of the two distinct periods resulted in the identification of six (first period) and eight (second period) topics.

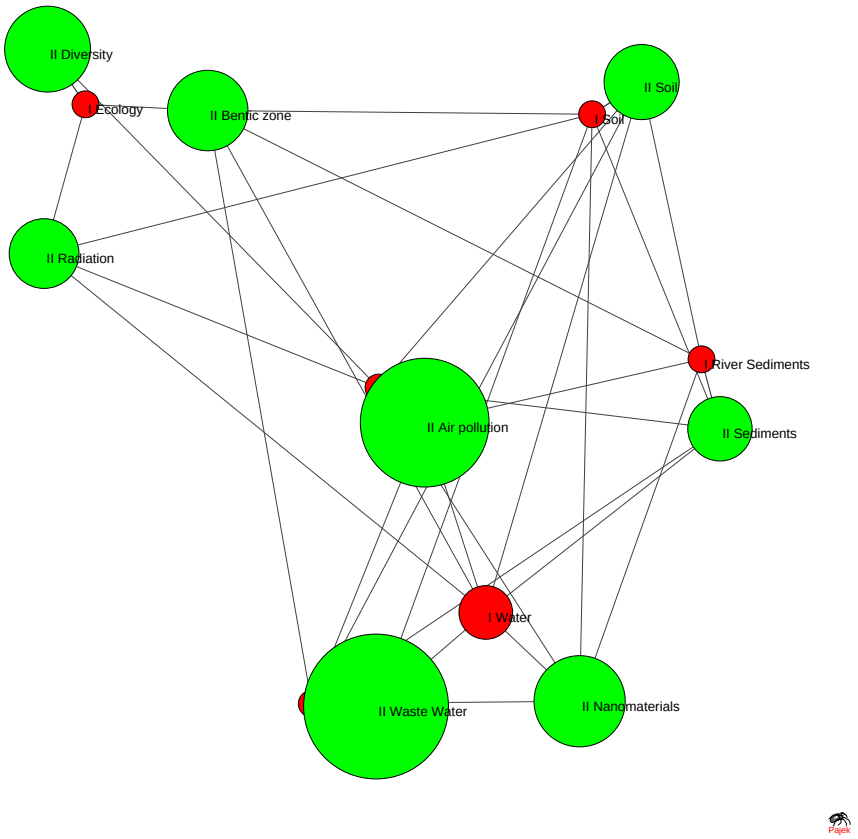

FIGURE 3.11. Environmental Sciences: relations between clusters
(Data source: Thomson Reuters Web of Science)

Emerging topics: Keywords and Core documents

Two emerging topics are detected. The first topic is labeled *'Radiation'*. This is a completely new topic which is not closely related to any of the topics from the first period. The second topic is labeled *'Nanopollution'*. This last topic emerged from the topic on *'Water'*. The topic is much larger but a topic shift towards nanomaterials is also observed.

Radiation

Keywords (n = 2712 – 9.7% of the complete discipline)

energy; consumption; life cycle assessment; industry; sustainability; efficiency; developing countries; greenhouse gas emissions; qsar; industrial ecology; radiation; methodology; uncertainty; spectra; risk; emissions; power; dosimetry; health; policy

Core Documents

ISI UT-code	Document title
000251680900011	CNMAN: A Chinese adult male voxel phantom constructed from color photographs of a visible anatomical data set
000251681000003	Characterisation of mixed neutron-photon workplace fields at nuclear facilities by spectrometry (energy and direction) within the EVIDOS project
000252847100004	Compliance of electronic personal neutron dosemeters with the new International Standard IEC 61526
000252847100041	Monte Carlo simulation of the operational quantities at the realistic mixed neutron-photon radiation fields canel and sigma
000252847100043	IMBA Professional Plus: A flexible approach to internal dosimetry
000252847100046	Neutron dose measurements with the GSI ball at high-energy accelerators
000252847100061	Evaluation of dose equivalent by the electronic personal dosemeter for neutron 'SAPHYDOSE-N' at different workplaces of nuclear facilities
000252847100064	Determination of the full response function of personal neutron dosemeters
000252847100065	Response of neutron detectors to high-energy mixed radiation fields
000252847100068	Bonner sphere neutron spectrometry at nuclear workplaces in the framework of the EVIDOS project
000252847100076	Response of alanine and radio-photo-luminescence dosemeters to mixed high-energy radiation fields
000252847100081	Direction distributions of neutrons and reference values of the personal dose equivalent in workplace fields
000253249900004	Bayesian hypothesis testing - Use in interpretation of measurements
000253584100005	Rebound effects of price differences
000254997400011	Characterization of neutron reference fields at us department of energy calibration fields
000254997400012	Design of a new irsn thermal neutron field facility using monte-carlo simulations
000254997400032	Energy and direction distribution of neutrons in workplace fields: Implication of the results from the EVIDOS project for the set-up of simulated workplace fields
000254997400061	The influence of the type of filling gas on the response of ionisation chambers to a mixed high-energy radiation field
000254997400062	Field calibration studies for ionisation chambers in mixed high-energy radiation fields
000254997400065	Performance of a PADC personal neutron dosemeter at simulated and real workplace fields of the nuclear industry
000254997400074	Neutron measurements in Spanish nuclear power plants with a Bonner Sphere Spectrometer System
000254997400083	Investigation of combined unfolding of neutron spectra using the UMG unfolding codes
000254997400086	Measurement of the spectral fluence rate of reference neutron sources with a liquid scintillation detector
000254997400096	Achievements in workplace neutron dosimetry in the last decade: Lessons learned from the EVIDOS project
000254997400100	The problems associated with the monitoring of complex workplace radiation fields at European high-energy accelerators and thermonuclear fusion facilities
000254997400107	Feasibility study of extremity dosemeter based on polyallyldiglycolcarbonate (CR-39) for neutron exposure
000254997400126	Experimental and theoretical study of the neutron dose produced by carbon ion therapy beams
000255251800008	Shift in the marginal supply of vegetable oil
000256524700006	Study of a gold-foil-based multisphere neutron spectrometer
000256726200009	Internal dose reconstruction under Part B of the energy employees compensation act
000256726200011	Development of rapid methods for assessing doses from internally deposited radionuclides
000256840200075	Internal dosimetry verification and validation database
000256840200114	Estimating uncertainty on internal dose assessments
000256887000010	Increased cause-specific mortality associated with 2003 heat wave in Essen, Germany
000257013600010	The effect of trade between China and the UK on national and global carbon dioxide emissions
000257033100017	Nanotechnology: getting it right the first time
000257973900006	Measuring progress towards carbon reduction in the UK
000260156400017	HIGH-ENERGY RESPONSE OF THE PRESCILA AND WENDI-II NEUTRON REM METERS
000260983600003	Analysis of the CONRAD computational problems expressing only stochastic uncertainties: neutrons and protons

Nanopollution

Keywords (n = 3534 – 12.6% of the complete discipline)

pharmaceuticals; aquatic environment; porous media; constructed wetlands; tandem mass spectrometry; escherichia coli; mass spectrometry; light; performance liquid chromatography; sewage treatment plants; disinfection by products; drinking water; design; personal care products; liquid chromatography; bacteria; mass transfer; aggregation; solute transport; identification

Core Documents

ISI UT-code	Document title
000251529700002	On colloid retention in saturated porous media in the presence of energy barriers: The failure of alpha, and opportunities to predict eta
000251842100005	The significance of heterogeneity on mass flux from DNAPL source zones: An experimental investigation
000251885900007	Simulating the dissolution of a complex dense nonaqueous phase liquid source zone: 2. Experimental validation of an interfacial area - based mass transfer model
000252037400030	C-60 colloid formation in aqueous systems: Effects of preparation method on size, structure, and surface, charge
000252467100006	Individual and mixture effects of selected pharmaceuticals and personal care products on the marine phytoplankton species Dunaliella tertiolecta
000253185200001	Temporal evolution of DNAPL source and contaminant flux distribution: Impacts of source mass depletion
000253578800001	Nanomaterials as possible contaminants: the fullerene example
000254890400030	Impact of natural organic matter on the physicochemical properties of aqueous C-60 nanoparticles
000255384300013	Mass-removal and mass-flux-reduction behavior for idealized source zones with hydraulically poorly-accessible immiscible liquid
000255444100059	Mechanism of C-60 photoreactivity in water: Fate of triplet state and radical anion and production of reactive oxygen species
000255822100025	Transport and retention of nanoscale C-60 aggregates in water-saturated porous media
000256094200001	Simplified contaminant source depletion models as analogs of multiphase simulators
000256204600027	Colloid transport and retention in unsaturated porous media: A review of interface-, collector-, and pore-scale processes and models
000256852300003	Relationship between mass-flux reduction and source-zone mass removal: Analysis of field data
000257086800001	Exploring e-waste management systems in the United States
000257620000044	Reactive tracer tests to predict dense nonaqueous phase liquid dissolution dynamics in laboratory flow chambers
000258003500001	Predicting dense nonaqueous phase liquid dissolution using a simplified source depletion model parameterized with partitioning tracers
000258325000004	Influence of electrolyte species and concentration on the aggregation and transport of fullerene nanoparticles in quartz sands
000258325000017	Toxicity of aqueous fullerene in adult and larval Fundulus heteroclitus
000258325000018	Effects of particle composition and species on toxicity of metallic nanomaterials in aquatic organisms
000258325000019	Oxidative stress and growth inhibition in the freshwater fish Carassius auratus induced by chronic exposure to sublethal fullerene aggregates
000259387600001	Application of a lumped-process mathematical model to dissolution of non-uniformly distributed immiscible liquid in heterogeneous porous media
000259988400013	Interaction of Fullerene (C-60) Nanoparticles with Humic Acid and Alginate Coated Silica Surfaces: Measurements, Mechanisms, and Environmental Implications
000260561200010	Precise and Accurate Compound Specific Carbon and Nitrogen Isotope Analysis of Atrazine: Critical Role of Combustion Oven Conditions
000260561200015	Evaluation of Toluene Degradation Pathways by Two-Dimensional Stable Isotope Fractionation
000261529100006	Evaluation of simplified mass transfer models to simulate the impacts of source zone architecture on nonaqueous phase liquid dissolution in heterogeneous porous media
000261529100007	Dissolved plume attenuation with DNAPL source remediation, aqueous decay and volatilization - Analytical solution, model calibration and prediction uncertainty
000261529100010	Influence of mass transfer characteristics for DNAPL source depletion and contaminant flux in a highly characterized glaciofluvial aquifer
000261529100012	Changes in contaminant mass discharge from DNAPL source mass depletion: Evaluation at two field sites

National contribution

Country	Complete Field	Radiation	Nanopollution
Belgium	1.66%	2.7%	1.42%
Brazil	1.85%	1.66%	1.61%
Denmark	1.37%	1.81%	1.5%
France	4.44%	5.88%	4.05%
Germany	5.83%	9.32%	6.48%
Greece	1.65%	2.74%	0.99%
India	4.5%	2.07%	5.26%
Italy	4.45%	6.29%	4.05%
Japan	4.15%	4.33%	4.08%
Netherlands	2.52%	4.51%	1.78%
PR China	10.54%	4.81%	10.42%
Poland	2.16%	1.41%	2.58%
Spain	4.81%	5.55%	5.04%
Sweden	2.37%	3.4%	1.78%
United Kingdom	6.44%	10.5%	5.46%
USA	27.62%	27.63%	26.32%

International co-operation

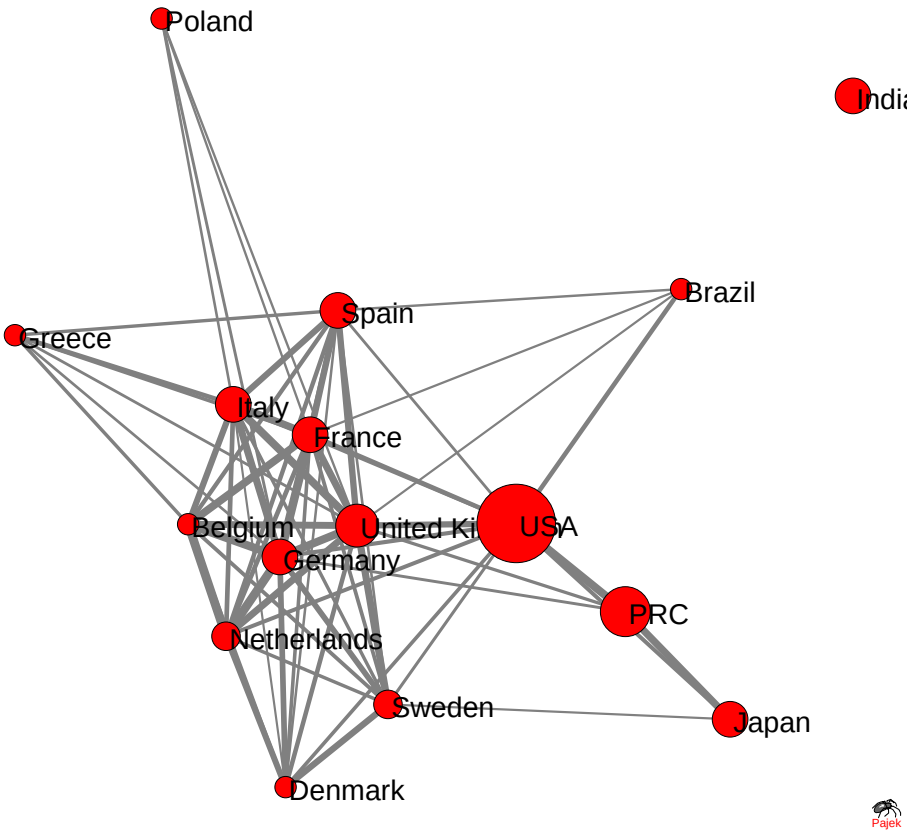

FIGURE 3.12. Environmental Sciences: Complete field
(Data source: Thomson Reuters Web of Science)

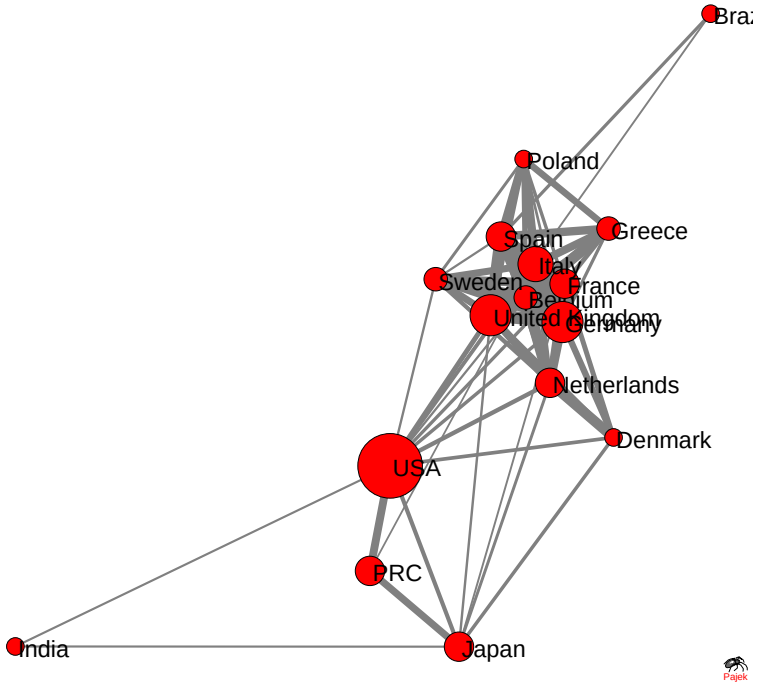

FIGURE 3.13. Environmental Sciences: Nano-pollution
(Data source: Thomson Reuters Web of Science)

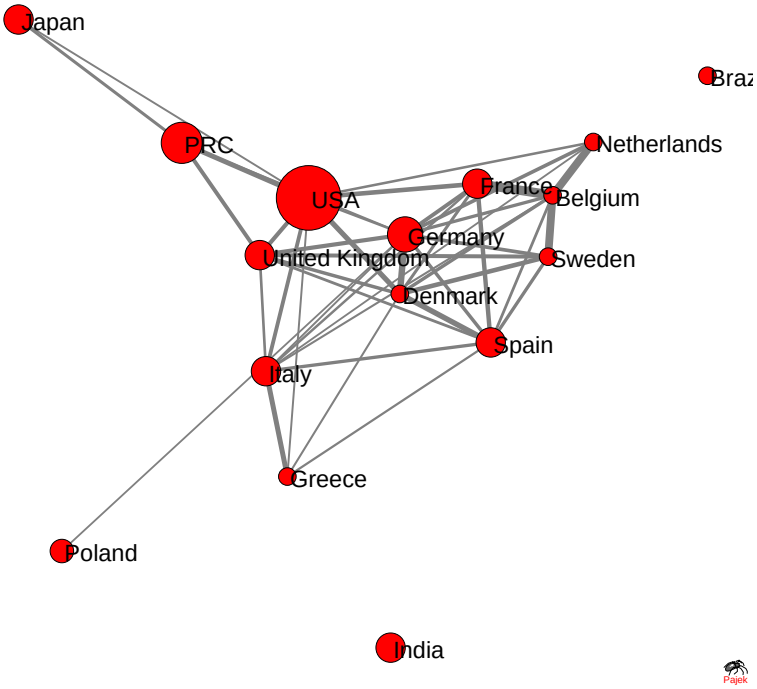

FIGURE 3.14. Environmental Sciences: Radiation
(Data source: Thomson Reuters Web of Science)

The international analysis clearly indicates a leading role of the United States in the whole field and also in the two emerging topics. China is at the second place considering the whole field of "environmental sciences". However, we observe an interesting difference concerning the role of China between our two emerging topics "radiation" and "nano pollution". Scientific activities related to "nano pollution" are at a similar level as those related to the whole field, while "radiation" is playing a far less important role. Within the European countries we find an opposite trend compared to China in the case of the United Kingdom. Here a much stronger focus is put on "radiation", while "nano pollution" seems to be quite less interesting in terms of scientific activities. Germany, on the other hand, is a case where both emerging topics attract more scientific activities compared to the whole field. A similar constellation can be observed for Spain.

Cooperation networks indicate strong links within Europe, notably in "nano pollution". The USA are closely linked to these clusters.

3.4.5 *Obstetrics & gynaecology*

Thomson Reuters describes the field as follows (TR, 2011a):

> 'Obstetrics & Gynecology covers resources on the medical fields concerned with female reproductive function and reproductive organs. Obstetrics covers resources on pregnancy, fetal health, labor, and puerperium. Gynecology covers resources on the health and diseases of female sex organs and their impact on women's overall health. This category also includes resources on fertility, infertility, and contraception.'

Cluster analyses of the two distinct periods resulted in the identification of seven topics in both periods.

Emerging topics: Keywords and Core documents

One emerging topic was detected. It is labeled *'Prenatal Diagnosis'*. This topic is already present in the first period but a topic shift can be observed. Due to technological advancement in ultrasound -3D- the attention shifted towards the fetal nasal bone and to the detection of other syndromes.

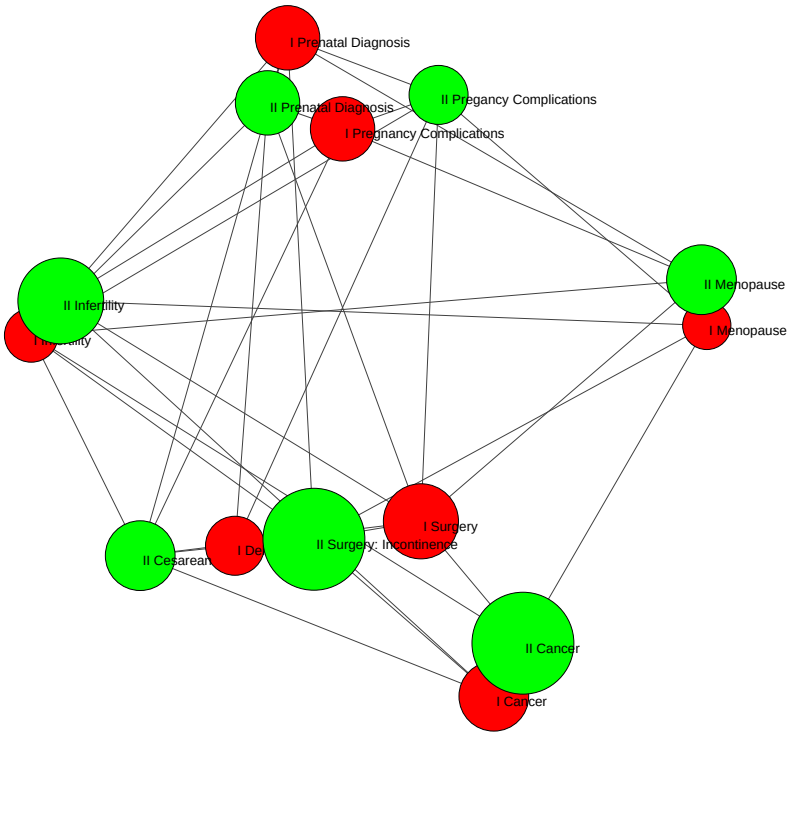

FIGURE 3.15. Obstetrics & Gynaecology: relations between clusters
(Data source: Thomson Reuters Web of Science)

Prenatal Diagnosis

Keywords (n = 4823 – 11.3% of the complete discipline)

prenatal diagnosis; doppler; placenta; flow; ultrasound; hypoxia; anomalies; size; blood flow; fetus; congenital heart disease; blood; abnormalities; ultrasonography; pregnancies; three dimensional ultrasound; phenotype; 3 dimensional ultrasound; sheep; fetal

Core Documents

ISI UT-code	Document title
000188734200003	Middle cerebral artery Doppler velocimetric assessment in two cases of hydrops fetalis without fetal anaemia
000188798500003	Maternal ethnic origin and fetal nasal bones at 11-14 weeks of gestation
000189184600003	Pilot study on the midsecond trimester examination of fetal nasal bone in the Chinese population
000220287300003	Likelihood ratio for trisomy 21 in fetuses with absent nasal bone at the 11-14-week scan
000220287300005	Assessment of the fetal nasal bone at 11-14 weeks of gestation by three-dimensional ultrasound
000220805400009	Fetal anemia: new technologies
000223418300003	How can we diagnose and manage twin-twin transfusion syndrome?
000225437500001	A case of extreme unconjugated fetal hyperbilirubinemia
000225683400004	The association between fetal nasal bone hypoplasia and aneuploidy
000227927800003	Multicenter study of first-trimester screening for trisomy 21 in 75,821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening
000227962700014	Improving the accuracy of fetal foot length to confirm gestational duration
000229361200012	First-trimester ductus venosus, nasal bones, and Down syndrome in a high-risk population
000229361200029	Contemporary treatments for twin-twin transfusion syndrome
000229812000009	Qualitative venous Doppler flow waveform analysis in preterm intrauterine growth-restricted fetuses with ARED flow in the umbilical artery - correlation with short-term outcome
000231276700010	Relationship between nuchal translucency thickness and prevalence of major cardiac defects in fetuses with normal karyotype
000232013700038	Correlation between middle cerebral artery peak systolic velocity and fetal hemoglobin after 2 previous intrauterine transfusions
000232013700056	Antenatal predictors of neonatal outcome in fetal growth restriction with absent end-diastolic flow in the umbilical artery
000232382400009	Insights into the pathophysiology of twin-twin transfusion syndrome
000232382400010	Management of fetofetal transfusion syndrome
000234160600003	Arterial and venous Doppler in the diagnosis and management of early onset fetal growth restriction
000234387300003	Twin-Twin Transfusion Syndrome: Where do we go from here?
000234813400009	Doppler and biophysical assessment in growth restricted fetuses: distribution of test results
000234813400010	Perinatal outcome in monochorionic twin pregnancies complicated by amniotic fluid discordance without severe twin-twin transfusion syndrome
000235556300003	Ultrasonographic evaluation of fetal nasal bone in a low-risk population at 11-13+6 gestational weeks
000235556300011	Mid-facial anthropometry in second-trimester fetuses with trisomy 21: a three-dimensional ultrasound study
000236945900004	The relation between fetal nasal bone length and biparietal diameter in the Korean population
000238491600005	Nasal bone length at 11-14 weeks of pregnancy in the Korean population
000238926700017	Nasal bone in first-trimester screening for trisomy 21
000239896300008	Interest of foetal nasal bone measurement at first trimester Trisomy 21 screening
000239931300012	First-trimester examination of fetal nasal bone in the Chinese population
000240972500004	Brain monitoring in the neonate - The rationale
000242035700006	Nasal bone assessment in prenatal screening for trisomy 21
000242225300008	Cardiac axis in fetuses with abdominal wall defects
000243563700016	Ultrasound assessment prior to laser photocoagulation for twin-twin transfusion syndrome for predicting intrauterine fetal demise after surgery in Japanese patients
000244865100005	Paradoxical scalloped placenta with polyhydramnios in twin-twin transfusion syndrome
000246768900014	Stage-related outcome in twin-twin transfusion syndrome treated by fetoscopic laser coagulation
000246771100003	Predictors of neonatal outcome in early-onset placental dysfunction
000247067200003	Twin-to-twin transfusion syndrome: Is the future getting brighter?
000247265000005	Severe twin-twin transfusion syndrome: outcome after fetoscopic laser ablation of the placental vascular equator
000247321600006	Tocolysis and delayed delivery versus emergency delivery in cases of non-reassuring fetal status during labor
000249962300007	Likelihood ratios for fetal trisomy 21 based on nasal bone length in the second trimester: how best to define hypoplasia?
000250097300023	A prospective, randomized, multicenter trial of amnioreduction vs selective fetoscopic laser photocoagulation for the treatment of severe twin-twin transfusion syndrome
000250540100001	Recent findings on laser treatment of twin-to-twin transfusion syndrome
000251742800018	How useful is the lung-to-head ratio in predicting outcome in the fetus with congenital diaphragmatic hernia? A systematic review and meta-analysis
000255666000009	Laser coagulation of placental anastomoses with a 30 degrees fetoscope in severe mid-trimester twin-twin transfusion syndrome with anterior placenta
000256259100014	Increasing the noninvasive management of rhesus isoimmunization
000256263300005	Prediction of fetal anemia by Doppler of the middle cerebral artery and descending thoracic aorta
000258972100008	Progression of Doppler abnormalities in intrauterine growth restriction
000260585800029	Fetal nasal bone status in Chinese women undergoing first-trimester screening for trisomy 21
000261281400014	Validation of Quintero stage III sub-classification for twin-twin transfusion syndrome based on visibility of donor bladder: characteristic differences in pathophysiology and prognosis

National contribution

Country	Complete Field	Prenatal Diagnosis
Belgium	2.09%	2.26%
Brazil	1.84%	1.66%
Denmark	1.61%	1.54%
France	4.15%	7.51%
Germany	5.06%	7.2%
Greece	1.74%	1.41%
India	1.42%	0.93%
Italy	5.6%	5.5%
Japan	4.12%	6.19%
Netherlands	3.74%	5.38%
PR China	2.44%	2.7%
Poland	1.17%	0.79%
Spain	2.14%	3.2%
Sweden	2.68%	2.64%
United Kingdom	10.82%	13.33%
USA	34.32%	27.63%

International co-operation

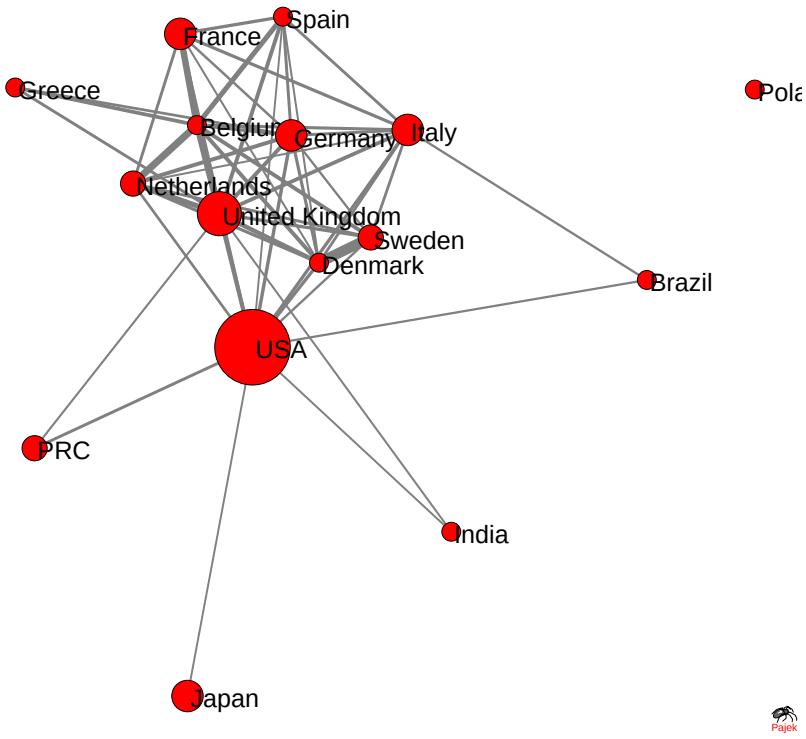

FIGURE 3.16. Obstetrics & gynaecology: Complete field
(Data source: Thomson Reuters Web of Science)

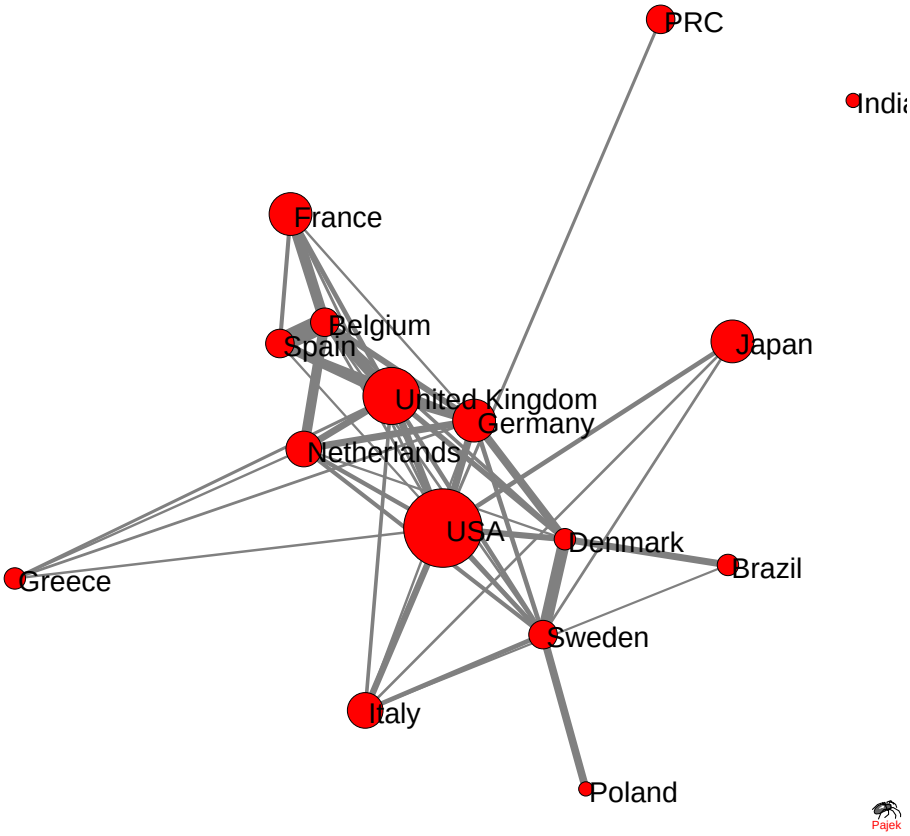

FIGURE 3.17. Obstetrics & gynaecology: Prenatal diagnosis
(Data source: Thomson Reuters Web of Science)

"Obstetrics and gynaecology" belong to "medical research" and as might have been expected we observe a strong position of the United States and the United Kingdom concerning the whole field. On the other hand, China is playing only a minor role in the whole field but also in our emerging topic "prenatal diagnosis". Looking at the European countries a rather strong position of Italy in the whole field of "obstetrics and gynaecology" is remarkable. The strong contributions of France and Germany might have been expected. An additional remarkable finding is the strong role of the Netherlands in the whole field and in particular in the emerging topic of "prenatal diagnosis". Also France and Germany are putting a lot of emphasis onto this emerging topic.

The cooperation network for the dynamic field "obstetrics and gynaecology" as a whole presents a strong European cluster with intense links between many of the European countries. The United States are on the outer edge of this cooperation cluster, however, we also observe quite intensive cooperation links with a number of European countries. Besides the United States and most European countries, all other countries and in particular the Asian countries are not involved in the international cooperation

links. In the emerging topic "prenatal diagnosis" the cooperation network is even denser and the United States are moving into a central position with the European cooperation cluster.

3.4.6 *Public, environmental & occupational health*

Thomson Reuters describes the field as follows (TR, 2011a):

> 'Public, Environmental & Occupational Health covers resources dealing with epidemiology, hygiene, and health; parasitic diseases and parasitology; tropical medicine; industrial medicine; occupational medicine; infection control; and preventive medicine. Also included are resources on environmental health; cancer causes and control; aviation, aerosol, and wilderness medicine.'

Cluster analyses of the two distinct periods resulted in the identification of six (first period) and seven (second period) topics.

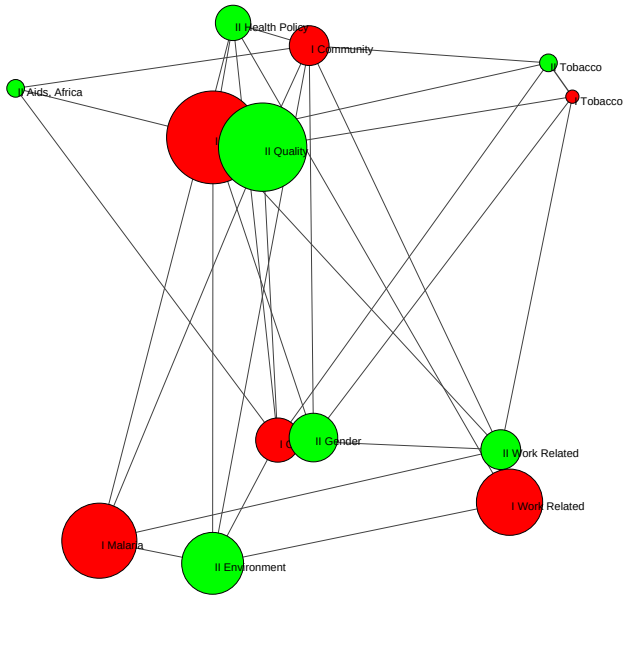

FIGURE 3.18. Public, environmental & occupational health: relations between clusters (Data source: Thomson Reuters Web of Science)

Emerging topics: Keywords and Core documents

Two emerging topics are detected. The first topic *'Environmental Issues'* deals with specific health issues related to the environment or location where people live. Many core documents refer to cities or countries. The second topic, labeled *'Aids in Africa'* emerged from HIV research present in several topics in the first period but is now mainly devoted to the HIV-stigma, to social, socio-political and regional aspects of HIV, with one focus on AIDS in Africa.

Environmental Issues

Keywords (n = 15268 – 20.5% of the complete discipline)

malaria; pcr; outbreak; identification; tests; antibodies; plasmodium falciparum; resistance; regression; infections; meta analysis; polymerase chain reaction; chloroquine; strains; logistic regression; brazil; missing data; models; heterogeneity; bias

Core Documents

ISI UT-code	Document title
000223626700016	Mortality in 13 French cities during the August 2003 heat wave
000225869500009	Ambient carbon monoxide may influence heart rate variability in subjects with coronary artery disease
000226079600009	Temperature and mortality among the elderly in the United States - A comparison of epidemiologic methods
000227430100039	Effects of air pollution on heart rate variability: The VA Normative Aging Study
000229460700026	Association of air pollution with increased incidence of ventricular tachyarrhythmias recorded by implanted cardioverter defibrillators
000231208300007	The estimation of SARS incubation distribution from serial interval data using a convolution likelihood
000231783200004	Mortality displacement of heat-related deaths - A comparison of Delhi, Sao Paulo, and London
000233059200003	Multipoint linkage analysis for a very dense set of markers
000234195400014	Impact of the 2003 heatwave on all-cause mortality in 9 French cities
000236792400010	Temperature and summer mortality: geographical and temporal variations in four Italian cities
000236926500019	Vulnerability to heat-related mortality - A multicity, population-based, case-crossover analysis
000239939200004	Association of ventricular arrhythmias detected by implantable cardioverter defibrillator and ambient air pollutants in the St Louis, Missouri metropolitan area
000242836700016	Focused exposures to airborne traffic particles and heart rate variability in the elderly
000244655200004	Ambient air pollution and cardiovascular emergency department visits in potentially sensitive groups
000247348600009	Japanese adult male voxel phantom constructed on the basis of CT images
000248304500021	Evaluation of two classifications for overweight among Brazilian adolescents
000250036800002	Hindsight: A re-analysis of the severe acute respiratory syndrome outbreak in Beijing
000251245000009	Temperature, temperature extremes, and mortality: a study of acclimatisation and effect modification in 50 US cities
000256865100011	Temperature and mortality in nine US cities

Aids In Africa

Keywords (n = 4145 – 5.6% of the complete discipline)

sex; aids; hiv; south africa; sub saharan africa; infection; human immunodeficiency virus; sexually transmitted diseases; transmission; hepatitis c; risk behaviors; epidemic; users; tanzania; adherence; new york city; individuals; africa; sexual behavior; uganda

Core Documents

ISI UT-code	Document title
000229576400001	Development of a brief scale to measure AIDS-related stigma in South Africa

National contribution

Country	Complete Field	Environmental Issues	Aids in Africa
Belgium	1.15%	1.78%	1.05%
Brazil	3.61%	4.74%	4.89%
Denmark	1.77%	1.8%	0.56%
France	3.32%	5.79%	2.17%
Germany	3.89%	4.46%	1.36%
Greece	0.63%	0.84%	0.19%
India	1.39%	2.49%	1.98%
Italy	2.61%	3.52%	1.86%
Japan	2.66%	3.13%	0.87%
Netherlands	3.6%	4.12%	1.98%
PR China	2.44%	3.25%	2.54%
Poland	0.72%	0.89%	0.06%
Spain	2.51%	3.25%	1.73%
Sweden	3.38%	2.74%	1.18%
United Kingdom	11.66%	11.72%	8.3%
USA	45.96%	38.33%	60.53%

International co-operation

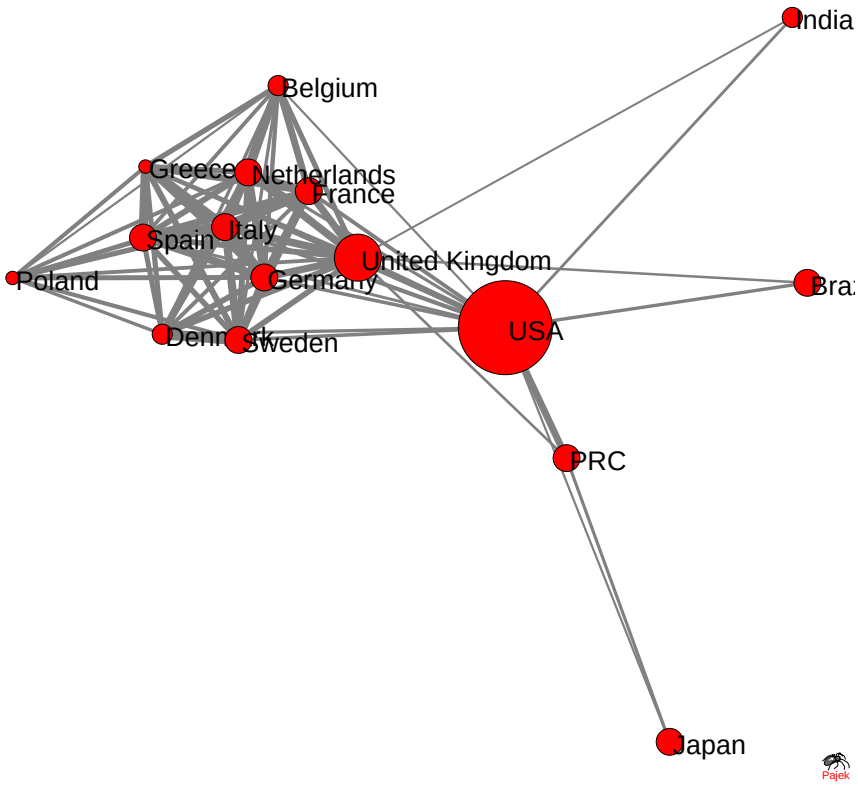

FIGURE 3.19. Public, environmental & occupational health: Complete field (Data source: Thomson Reuters Web of Science)

The complete field is dominated by the United States and the United Kingdom contributing together almost 60 per cent of all publications. Looking at the other European countries, the rather strong position of Sweden with respect to the complete

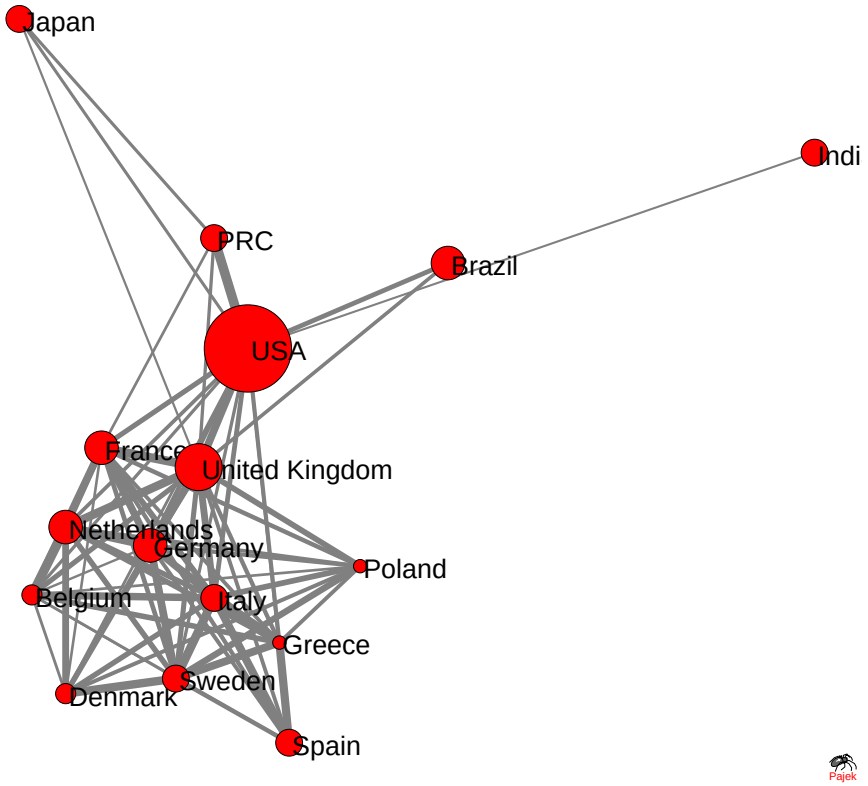

FIGURE 3.20. Public, environmental & occupational health: Environmental Issues
(Data source: Thomson Reuters Web of Science)

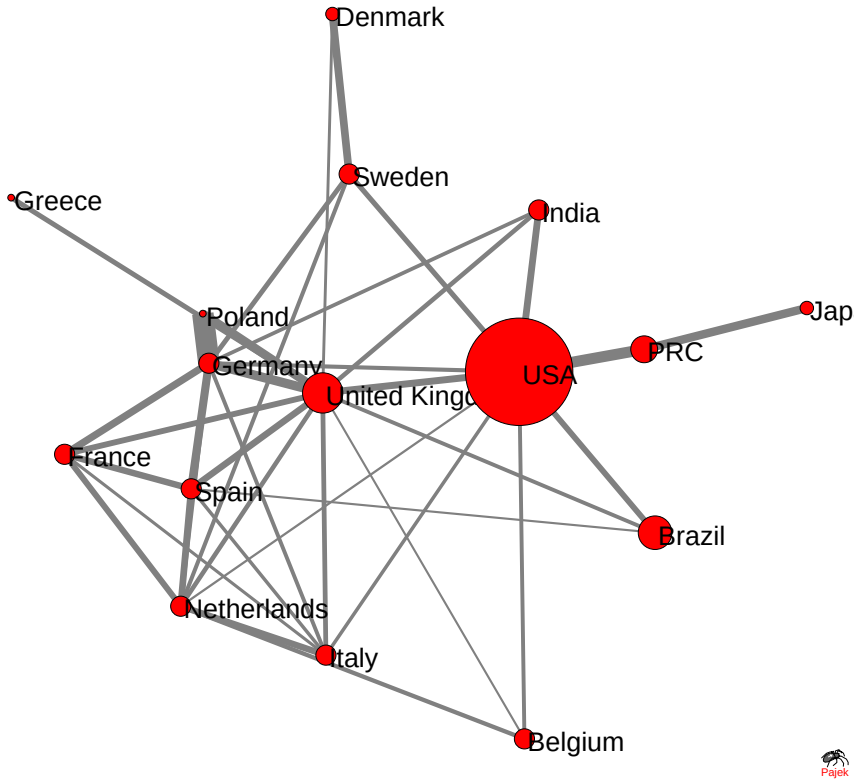

FIGURE 3.21. Public, environmental & occupational health: Aids in Africa
(Data source: Thomson Reuters Web of Science)

field seems remarkable. The emerging topic of "environmental issues" is addressed relative strongly by France, Germany, the Netherlands, and Spain. The emerging topic related to the "HIV stigma is covered largely by research activities in the United States. Interestingly, also Brazil pays strong attention to this field. The international cooperation analysis represents a very dense and intensive network among European countries for the whole topic. But also a number of intense cooperations between the United States and European countries are obvious. In the emerging topic "environmental issues" we observe basically a similar international cooperation pattern. In the case of "AIDS in Africa", cooperation links are less dense and the United States is dominating the situation expressing some cooperation links to the United Kingdom but also to China.

3.4.7 *Geography*

Thomson Reuters describes the field as follows (TR, 2011b):

> 'Geography covers resources concerned with socio-cultural aspects of the Earth's surface emphasizing the human, economic, political, urban, and environmental issues of the discipline. The history of geography and the study of cartography are also covered in this category.'

Cluster analyses of the two distinct periods resulted in the identification of six (first period) and five (second period) topics.

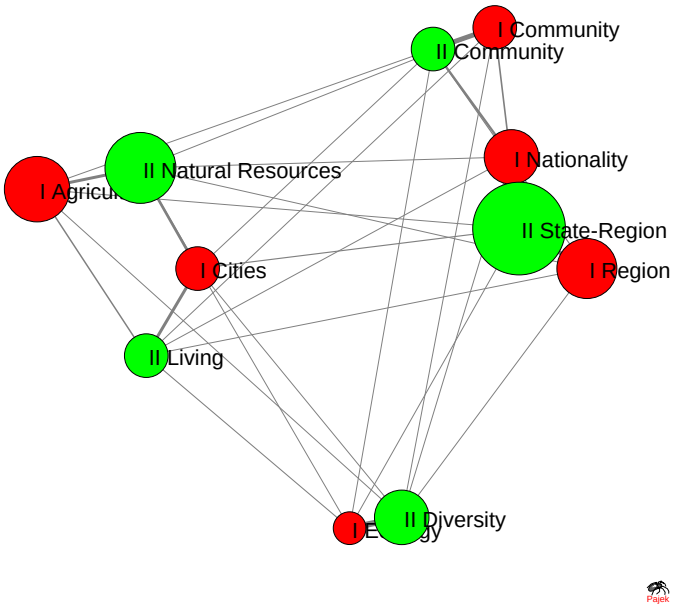

FIGURE 3.22. Geography: relations between clusters
(Data source: Thomson Reuters Web of Science)

Emerging topics: Keywords and Core documents

One emerging topic was detected. It is labeled *'State & Region'*. This topic is a combination of the topics on *'Nationality'* and on *'Region'* from the first period. The new topic is not just a merger of the two but a real shift in the topic can be observed. Now the focus goes to the region as defining identity in a globalized world.

State & Region

Keywords (n = 3551 – 39.2% of the complete discipline)

state; convergence; identity; networks; innovation; regions; neoliberalism; economic development; regional development; labor; geographies; clusters; governance; industry; space; russia; competition; uk; globalization; public space

Core Documents

ISI UT-code	Document title
000187789500002	Spatialities of transnational resistance to globalization: the maps of grievance of the Inter-Continental Caravan
000188417700005	Innovation in complex capital projects: clustering and dispersion in two cases from Argentina and the UK
000189102400003	Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation
000189215400004	Changing scale as changing class relations: variety and contradiction in the politics of scale
000189310800002	Australian trade unions and the politics of scale: Reconstructing the spatiality of industrial relations
000189310800003	Understanding social and spatial divisions in the new economy: New media clusters and the digital divide
000221377900002	Neoliberal nature and the nature of neoliberalism
000221377900006	Privatizing conditions of production: trade agreements as neoliberal environmental governance
000222744500005	Regions and sustainable development: regional planning matters
000222830500006	Regulating economic globalization
000223469500009	Rules of privatization: Contradictions in neoliberal regulation of North Pacific fisheries
000223639600011	Place and region: looking through the prism of scale
000224557600008	Adjustment or renewal in regional clusters? A study of diversification amongst SMEs in the Aberdeen oil complex
000225789800003	Competitive city-regionalism as a politics of space: a critical reinterpretation of the new regionalism
000225980000002	Competitiveness of regions from an evolutionary perspective
000226202600003	Regional spaces, spaces of regionalism: territory, insurgent politics and the English question
000226292500008	Cluster relations in the media industry: Exploring the 'distanced neighbour' paradox in Leipzig
000227301300004	Exporting the German model: The establishment of a new automobile industry cluster in Shanghai
000230516800006	Scaling alternative economic practices? Some lessons from alternative currencies
000231520600003	Beyond rescaling: Reintegrating the 'national' as a dimension of scalar relations
000232054300003	Resources in economic geography: from substantive concepts towards a relational perspective
000232198400006	There's nothing inherent about scale: political ecology, the local trap, and the politics of development in the Brazilian Amazon
000233643500001	Reconfiguring environmental governance: Towards a politics of scales and networks
000239629500007	Offshoring the financial services industry: implications for the evolution of Indian IT clusters
000240110300007	Transnational collaborations, local competitiveness: Mapping the geographies of filmmaking in/through Hong Kong
000241951700003	Situating design in the Canadian household furniture industry
000242497300004	Extraregional linkages and the territorial embeddedness of multinational branch plants: Evidence from the South Tyrol region in northeast Italy
000242803800004	Of scalar hierarchies and welfare redesign: child care in three Canadian cities
000242904000002	International labour migration and tacit knowledge transactions: a multi-level perspective
000243928800011	"We simply have to do that stuff for our survival": Labour, firm innovation and cluster governance in the Canadian automotive parts industry
000244981000010	Introduction to a debate on city-regions: New geographies of governance, democracy and social reproduction
000245959900004	Globalization and the power of rescaled narratives: A case of opposition to mining in Tambogrande, Peru
000246382000012	Exploring the role of professional associations in collective learning in London and New York's advertising and law professional-service-firm clusters
000247069900001	More than 'managing across borders? ' the complex role of face-to-face interaction in globalizing law firms
000251025200003	Neoliberalism, climate governance and the scalar politics of EU emissions trading
000251578800007	Entrepreneurship, proximity and regional innovation systems
000251601200004	Territorial, scalar, networked, connected: In what sense a 'regional world'?
000252258000003	Technological learning and innovation in China in the context of globalization
000253302200010	Neoliberalising nature: the logics of deregulation and reregulation
000254422900004	Alliance-driven governance: applying a global commodity chains approach to the UK biotechnology industry
000255184100002	The predicament of firms in the new and old economies: a critical inquiry into traditional binaries in the study of the space-economy
000255434900003	Rethinking scale as a geographical category: from analystis to practice
000255553100001	The spatialities of contentious politics
000256155000002	Global production networks: realizing the potential
000256155000003	Value chains, networks and clusters: reframing the global automotive industry
000256762600004	'The flea on the tail of the dog': power in global production networks and the restructuring of Canadian automotive clusters
000257267800002	Theorizing sociospatial relations
000257267800010	Making neoliberal states of development: the Ghanaian diaspora and the politics of homelands
000257336500006	Between luminaires and meat grinders: International trade fairs as temporary clusters
000260700700010	The benefits of learning in clusters: analyzing upward mobility for skilled workers in the Cebu furniture cluster

National contribution

Country	Complete Field	State & Region
Belgium	0.85%	0.52%
Brazil	0.46%	0.4%
Denmark	0.88%	0.83%
France	2.07%	1.49%
Germany	3.38%	3.85%
Greece	0.38%	0.32%
India	0.26%	0.11%
Italy	1.41%	1.64%
Japan	1.16%	0.63%
Netherlands	4%	3.47%
PR China	2.28%	0.92%
Poland	0.25%	0.34%
Spain	3.85%	2.7%
Sweden	1.54%	1.81%
United Kingdom	29.37%	35.17%
USA	29.12%	27.33%

International co-operation

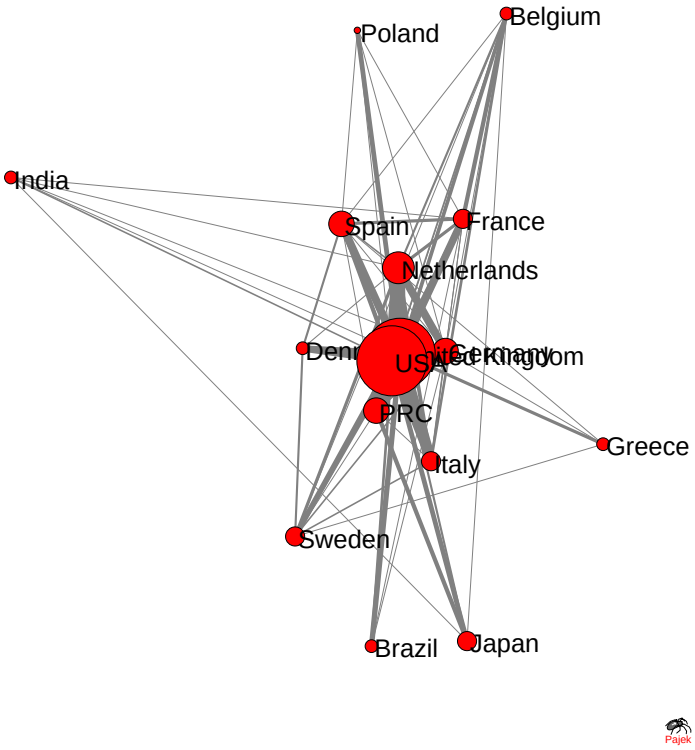

FIGURE 3.23. Geography: Complete field
(Data source: Thomson Reuters Web of Science)

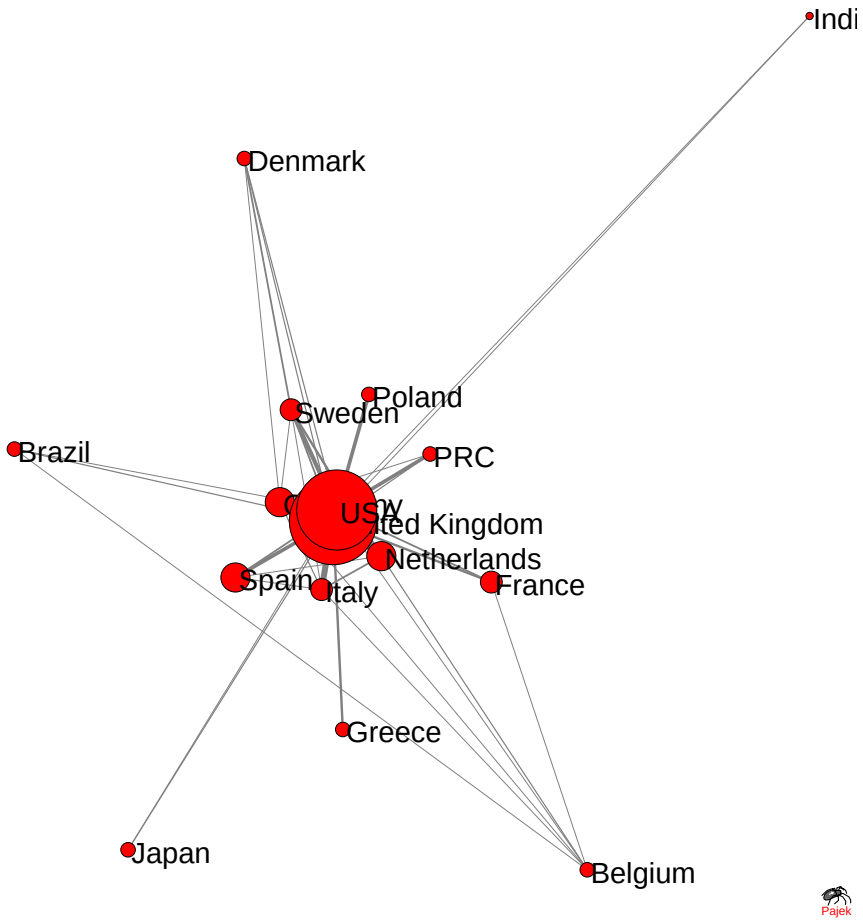

FIGURE 3.24. Geography: State & Region
(Data source: Thomson Reuters Web of Science)

The most interesting observation concerning national contributions to the dynamic field "geography" and also the emerging topic "state and regions" concerns the leading role of the United Kingdom contributing more publications to the whole field and also to the emerging topics than the United States. In particular, in the emerging topic "state and region" the United Kingdom is playing a dominant role with more than 35 per cent of all publications. Accordingly, all other European countries contribute only lower shares to the complete field and also to the emerging topic "state and regions". Considering their rather small size the strong contribution of the Netherlands to both, the complete field and also to the emerging topic seems remarkable.

The cooperation networks reflect the dominating role of two countries, the United Kingdom and the United States. Basically for the whole field we do not find a typical network of collaborations, but rather two strong clusters representing activities of the United Kingdom and the United States. In the case of the emerging topic "state

and regions" a network character is more pronounced, indicating quite intensive collaborations between various European countries and also the United States.

3.4.8 *Further cluster evolution: a stability experiment*

In order to analyse the further evolution of emerging and other topics, the dynamic clustering described in Section 3.2 has been applied also for a more recent period to the discipline *energy & fuels*. In addition to the two periods 1999–2003 and 2004–2008, the three year period 2009–2011 has been added. The period 2009–2011 has been clustered in the same way as has done for the previous years. Again, clusters have been linked with the topics in the period 2004-2008 using core documents, the total topics based on direct citation links. The results are shown in Figure 3.25.

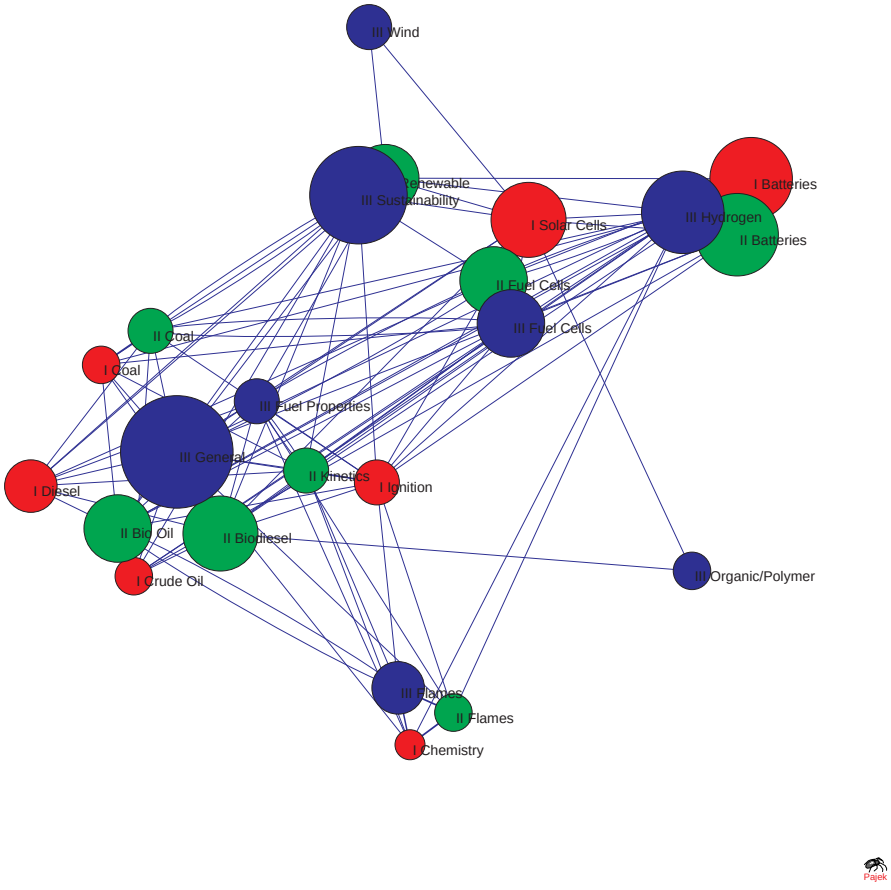

FIGURE 3.25. Cluster evolution of the discipline energy & fuels.
Red: 1999-2003; Green: 2004-2008; Blue: 2009-2011
(Data source: Thomson Reuters Web of Science)

The biofuel cluster has further grown and combines now the bio-oil and bio-diesel topics. It is also interesting that a small new topic emerged in the environment of renewable energy: This topic is devoted to wind energy. The topic on batteries indicates specialisation towards hydrogen. Finally, the small new cluster related to polymers seems to be the result of "immigration" from other disciplines. Summarising, the trends observed in the period 1999–2008 continue but some new developments are already becoming visible.

3.5 VALIDATION OF RESULTS

For a validation of the identified emerging topics as described in the previous chapters we carried out a qualitative assessment of the fields and topics by researchers and experts from respective disciplines. In total 21 scientist and experts from various academic backgrounds from different European countries were consulted. The majority of the experts were from academia or public research institutions and have been selected based on publication record and experience in the field. As the first step, the identified emerging areas were assessed by scientists working in the Fraunhofer Society, representing a great variety of different scientific backgrounds (biology, chemistry, physics, engineering, geography, material sciences). This expert assessment was conducted via face to face meetings. Phone interviews were carried out with international experts when additional external opinion was needed. All the consulted specialists/scientists were asked to validate the identified fields and also to assess the significance of the respective fields within the whole landscape of science but also with respect to their potential future economic relevance. All in all, we considered the following topics:

- Brain-computer interface representing the field biomedical engineering
- Fuel cells and bio diesel representing the field energy and fuels
- Nanomaterials representing the field environmental sciences
- Prenatal diagnosis and 3D-ultrasound representing the field obstetrics and gynaecology
- Environment-related risks representing the field public, environmental and occupational health
- Clustering and evolutionary economic geography representing the field geography

3.5.1 *Biomedical engineering*

The largest emerging topic identified within biomedical engineering was "Brain Computer Interface" (BCI) that is sometimes also called a direct neural interface or a brain–machine interface. It is a direct communication pathway between the brain and an external device such as computers or prostheses. BCIs are often aimed at assisting, augmenting or repairing human cognitive or sensory-motor functions. In other words,

a BCI allows users to act on their environment by using only brain activity, without using peripheral nerves and muscles. It is not a new field, the research on BCIs began already in the 1970s, but during 2004-2008 an extensive growth of the research topic was observed by experts.

The major goal of BCI research is to develop systems that allow disabled users to communicate with other persons, to control artificial limbs, or to control their environment. Based on the expert opinion, one of the groups of patients who could benefit from BCI the most are the individuals suffering from the so-called locked-in syndrome. It is a condition in which patients are fully conscious and aware of what is happening in their environment but are not able to communicate or move and do not have the communication possibilities (Smith, 1992). The condition is caused by a nearly total loss of control over the voluntary muscles. A disease that is known to lead to the locked-in syndrome is amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease. ALS is a progressive, neurodegenerative disease and is characterised by the death of motor neurons which in turn leads to the loss of control over voluntary muscles. Besides ALS also multiple sclerosis, stroke or other cerebrovascular incidents leading to the infarction or degeneration of parts of the brain can cause the locked-in syndrome (Lakerveld et al, 2008).

As an alternative application area for brain-computer interfaces (BCIs) based on the cluster analysis and expert opinion, lies in the field of multimedia communication that could be used as well by persons not suffering from disabilities. For example, in the field of multimedia communication and human-computer interaction, BCIs could possibly be used as an additional modality, together with more traditional modalities, such as the auditive and visual modalities. Multimodal communication with the help of a BCI would help to increase the communication bandwidth between man and machine. Beyond communication, other applications of BCI involving multimedia can also be envisioned. For example one can imagine (multiplayer) games in which BCIs are used for control (Birbaumer, 2006, Millán et al, 2010). Another interesting application area might be the visualisation, or sonification, i.e. the transformation into sound, of brain activity. To develop systems for usage in the field of assistive technology or multimedia communication, many aspects of BCI systems are currently being investigated. Research areas include evaluation of invasive and non-invasive technologies to measure brain activity, evaluation of control signals (i.e. patterns of brain activity that can be used for communication), development of algorithms for translation of brain signals into computer commands, and the development of new BCI applications (Iversen et al, 2008).

Based on expert opinion, a large variety of proof-of-concept systems already exists. Experimental studies have reported some successful implementations of BCIs (Birbaumer, 2006); however, much of the field still remains unexplored as none of the systems described in the scientific literature is suited for daily use by disabled persons or for use in multimedia environments. This could be due to the fact that the majority of technology underlying BCIs is not yet mature enough for usage out of the laboratory. Thereby many challenging and interesting questions in BCI research are

thus still waiting to be explored by the researchers in the upcoming years. In general experts confirmed the high relevance of the topic.

3.5.2 *Energy and fuels*

Biodiesel

Biodiesel as a renewable energy carrier has been receiving increasing attention by research communities and policy-makers, because of rising petroleum price and its environmental advantages (Lin et al, 2011). Biodiesel, which is also known as fatty acid methyl ester (FAME), is produced from transesterification of vegetable oils or animal fats with the addition of methanol. It is quite similar to petroleum-derived diesel in its main characteristics such as cetane number, energy content, viscosity and phase changes (Lim and Teong, 2010). Biodiesel contains no petroleum products, but it is compatible with conventional diesel and can be blended with fossil-based diesel to create a stable biodiesel blend (Lin et al, 2009).

Biodiesel offers a number of technical and environmental benefits over conventional fossil-based fuels. Especially, similarities between the combustion properties of biodiesel and fossil-based diesel have made the former one of the most promising alternatives for a renewable and sustainable transportation fuel. In recent years, the biodiesel industry developed rapidly. According to the International Energy Agency's (IEA) report on biodiesel production that looked at the 21 leading biofuel producing countries, global biodiesel production has increased tenfold from 2000 to 2008 and could be doubled to 21.8 bn liters by 2012. With the global energy crisis approaching, biodiesel fuel will play a more important role in strengthening a nation's energy security. Secondly, as a renewable energy, biodiesel is derived from plant or animal materials which can contribute to the reduction of greenhouse gas (GHG) emissions when replacing fossil oil if they are sustainably managed. Thirdly, the increased demand for oil crops for biodiesel production clearly has a positive effect on net farm income and also reduces government outlays to farmers by raising the market price of oil crops (Tonkin, 2009).

With the global increase in the scale of biodiesel production, biodiesel has become a systemic risk with respect to its economic, ecological, and socio-political impacts. Currently the researchers and policy-makers have discussed various issues linked to first-generation biodiesel and its potential to increase food prices and damage biodiversity; their relatively low GHG abatement capacity yet high marginal carbon abatement costs and their direct and indirect impacts on land use change and related greenhouse gas emissions (Sims et al, 2008).

'First generation' biodiesel is mainly produced from food-grade oils. According to some researches, feedstock acquisition currently accounts for over 75% of biodiesel production expenses, which is a serious threat to the economic viability of the biodiesel industry (Meng et al, 2009). Accordingly, the end cost of the biodiesel mainly depends on the price of feedstock. With vegetable oil price soaring high in recent years, the cost of producing biodiesel will keep raising. Biodiesel can lose its competitive advantage

due to high price. Based on the above, scientists are developing a new generation of biodiesel to help avoid such problems (Lin et al, 2011). One potential solution to this problem is employment of alternative feedstocks of varying type, quality, and cost. These feedstocks may include soapstocks, acid oils, tall oils, used cooking oils, and waste restaurant greases, various animal fats, non-food vegetable oils, and oils obtained from trees and microorganisms such as algae. Additionally, genetic modification is also being used to introduce favourable traits into biodiesel crops, such as higher yields or the ability to grow on non-arable land (Lin et al, 2011).

Currently used diesel engines are not suitable for high viscosity, low volatility and poly-unsaturated character vegetable oils to be applied directly (Srivastava, 2000). Refinement has to be made in order to turn those vegetable oils into quality fuel. Conventional methods of the application of vegetable oil in diesel engines are direct mixing and micro-emulsion. These two physical methods do not require any chemical process and can lower the viscosity of vegetable oil, but they cannot solve the problem of carbon deposits and lube pollution, and the high temperature pyrolysis cracking is hard to be controlled by its reactant at high temperature (Lin et al, 2011). Based on expert opinions, transesterification is the most promising solution to the high viscosity problem, it is widely available technique for industrialised biodiesel production due to its high conversion efficiency and low cost. In recent years, nanotechnology is applied more and more in the catalytic field. Because nanocatalysts have high specific surface and high catalysis activities, they may provide a possible way to solve the above problems. Therefore, they have become the focus of recent research (Shu et al, 2007).

Biodiesel has been receiving increasing attention during the last years because of the relevance it gains from the rising petroleum price and its environmental advantages. It is expected that biofuels will form an important element of global transport energy mix (in the order of 20-30% of total requirement) over the next 40 years and beyond.

Fuel cells

Based on the bibliometric analyses and expert assessment, fuels cells were identified as one of the major clusters in the field of energy research. Indeed, due to the growing concerns on the depletion of petroleum-based energy resources and climate change, fuel cell technologies have received much attention in recent years owing to their high efficiencies and low emissions (Wang et al, 2011). During the years of 2006-2008 more than 4000 articles were published by the research community in well established journals, all tackling the various issues related to fuel cell research and development. Also, since 1986 the EU has funded some 200 projects on hydrogen and fuel cell energy technologies with a total contribution of over € 550 Mio.

Fuel cells, which are classified according to the electrolyte employed, are electrochemical devices that all use the same principal of combining hydrogen fuel and oxygen from the air to produce electricity, heat and water. The reaction occurs in a physical piece of equipment called the proton exchange membrane which contains a cation and anion in separate containment areas. When the cation and anion react

with one another, a chemical reaction occurs and electricity is the final product. Five categories of fuel cells have received major efforts of research (Song, 2002):

- Polymer electrolyte membrane (PEM) fuel cells or PEMFCs (also called PEFCs).
- Solid oxide fuel cells (SOFCs)
- Alkaline fuel cells (AFCs)
- Phosphoric acid fuel cells (PAFCs)
- Molten carbonate fuel cells (MCFCs)

Polymer electrolyte membrance fuel cells (PEMs, also called PEMFCs or PEFCs) are believed to be the best type of fuel cell as vehicular power source to eventually replace the gasoline and diesel internal combustion engines. First used in the 1960s for the NASA Gemini Program, they are currently being developed and demonstrated for systems ranging from 1W to 2kW. PEM fuel cells use a solid polymer membrane (a thin plastic film) as the electrolyte. The fuel for the PEMFC is hydrogen and the charge carrier is the hydrogen ion (proton). At the anode, the hydrogen molecule is split into hydrogen ions (protons) and electrons. The hydrogen ions permeate across the electrolyte to the cathode while the electrons flow through an external circuit and produce electric power. Oxygen, usually in the form of air, is supplied to the cathode and combines with the electrons and the hydrogen ions to produce water. Compared to other types of fuel cells, PEMFCs generate more power for a given volume or weight of fuel cell. This high-power density characteristic makes them compact and lightweight. In addition, the operating temperature is less than 100°C, which allows rapid start-up. These traits and the ability to rapidly change power output are some of the characteristics that make the PEMFC the top candidate for automotive power applications.

The design of a solid oxide fuel cell (SOFC) is different from the typical design of a typical fuel cell. Since SOFC is a solid, the fuel cell does not have to be constructed with the plate arrangement typical of other fuel cell types. SOFC uses a ceramic, solid-phase electrolyte, which reduces corrosion considerations and eliminates the electrolyte management problems associated with the liquid electrolyte fuel cells. To achieve adequate ionic conductivity in such a ceramic, however, the system must operate at high temperatures in the range of 650-1,000°C, typically around 800-1,000°C in the current technology. The SOFC is a solid state device and shares certain properties and fabrication techniques with semiconductor devices. They are considered to be state-of-the-art fuel cell technology for electric power plants and offer the stability and reliability of all-solid-state ceramic construction. Operation up to 1,000°C allows more flexibility in the choice of fuels and can produce better performance in combined-cycle applications.

Alkaline fuel cells were first used by the US Space Program to produce electricity and water for the astronauts. Alkaline fuel cells operate at around 100°C to 250°C, but recent designs have proven that they can operate at lower temperatures from 29°C to 70°C. The AFCs are very high performance fuel cells. This is because of the

fast chemical reaction that takes place in the fuel cell. Although the AFCs are very efficient, they are very susceptible to carbon dioxide. A small amount of carbon dioxide can contaminate the fuel cell and thus ruin the entire fuel cell. The decontamination process is also very costly, which could make the alkaline fuel cell cost more than the value it produces. One advantage is that these types of fuel cells have a very high performance.

Phosphoric acid fuel cells (PAFCs) are typically used for stationary power generation, but some phosphoric-acid fuel cells have been used to power large vehicles such as city buses. This type of fuel cell has certain defects as well which makes them slightly doubtful for mass production. Phosphoric-acid fuel cells are only slightly more efficient than combustion-based power plants. They are also less powerful than other fuel cells, but they have the same weight and volume as the other fuel cells. Because of this, these fuel cells are typically large and heavy. PAFCs are also expensive. They require an expensive platinum catalyst, which raises the cost to produce the fuel cell.

Molten carbonate fuel cells (MCFCs) are high-temperature fuel cells that use an electrolyte composed of a molten carbonate salt mixture suspended in a porous, chemically inert ceramic matrix of beta-alumina solid electrolyte (BASE). Since they operate at extremely high temperatures of 650°C and above, non-precious metals can be used as catalysts at the anode and cathode, reducing costs. Molten carbonate fuel cells (MCFCs) are currently being developed for natural gas, anaerobic digestion gas, and coal-based power plants for electrical utility, industrial, and military applications. The disadvantages of MCFC are that the electrolyte is corrosive and mobile, and a source of CO2 is required at the cathode to form the carbonate ion. All fuel cell types offer a significant advantage over traditional combustion-based thermal energy conversion, in that they provide efficiencies of electrical power supply that can reach up to 60% in electrical energy conversion and overall 80% in co-generation of electrical and thermal energies with >90% reduction in major pollutants (Papageorgopoulos, 2011). Since the fuel is converted directly to electricity, all fuel cells can operate at much higher efficiencies than internal combustion engines, extracting more electricity from the same amount of fuel. The fuel cell itself has no moving parts – making it a quiet and reliable source of power.

Fuel cells can in principle be built in a wide range of power ratings, from a few mW to several MW, and can be used in a wide variety of applications, from miniaturised portable power through transport to power generation in a variety of sizes. They offer advantages of weight compared with batteries, and instantaneous refuelling, similar to combustion engines (http://ec.europa.eu/research/energy/eu/research/fch/background/index_en.htm).

The main technology applications for fuels cells are: transport fuels, propulsion systems and on-board electricity supplies for vehicles, home energy supply systems, alternatives or supplements to batteries in diverse use contexts, industrial combined heat and power (CHP) generation and electricity supply independent from the grid. Expected benefits also include increased security of energy supply, substantial reduction of greenhouse gas and other emissions, achievement of more sustainable

transport patterns, improved performance of home and industrial heat and electricity generation, greater operating times of electric appliances and prospects for constructing sophisticated energy-supply systems for remote locations (Garche et al, 2009).

However, further scientific breakthroughs are required to overcome barriers related to cost and durability to enable fuel cell commercialisation. Breakthroughs in material development, acquisition of fundamental knowledge, and development of analytical models and experimental tools are particularly important for current fuel cell development (Gittleman et al, 2010). For example, avoiding electrode flooding is of critical importance for optimal fuel-cell performance and durability; however this phenomenon is not well understood by the scientific community. The ability to model fuel and reactant transport and electrochemical reactions in electrodes is critical, particularly in the cathode in which the oxygen reduction reaction (ORR) is sluggish and inefficient and water is generated. The fundamental understanding of the electrochemical activity at the triple-phase boundaries would be also a key to breakthroughs of further Pt-loading reduction (Wang et al, 2011).

Next to the basic research questions, manufacturing technologies and industrialisation issues becomes also increasingly important. The topics of analysis tools, quality assurance methodologies, online detection of faulty components, and qualification of components remain as areas of interest and of innovation. Advanced analysis also includes modelling and simulation tools that serve to predict materials durability and degradation processes.

Although great efforts have been made with many breakthroughs achieved, another five to ten years is anticipated being required prior to fuel cell worldwide deployment (DOE-EERE 2008). From the basic research perspective, the knowledge generating activities that enable improved understanding of fuel cell operation principles and engineering of fuel cell technology remain most crucial.

3.5.3 *Environmental Sciences*

Based on the bibliometric cluster analyses, the field of nanomaterials in aquatic environment was identified as one of the new and emerging research areas in Environmental Sciences. Over 3,000 scientific articles were published on this topic only in 2008 and also the expert assessment confirmed the relevance of basic research to look further into this area, as due to the relative novelty of this technology, very little has been known about its risks on biological systems. The concerns about the use of products of nanotechnology have been debated over the last years within the scientific community and among policy-makers.

In spite of the previous research efforts and an increasing amount of published studies on nanomaterials in general, the interviewed experts assessed the knowledge about the harmful effects of nanoparticles still rather limited, especially in aquatic environments.

However, the existing scientific literature provides us already with some examples how nanomaterials end up in environment and what harmful effects they can cause to the living organisms. Industrial products and wastes tend to end up in waterways (e.g. drainage ditches, rivers, lakes, estuaries and coastal waters) despite the implemented safeguards (Daughton, 2004). Consequently as the nanotechnology industries start to come on line with larger scale production, it is inevitable that nanoscale products and by-products will enter the aquatic environment as pollutants.

According to interviewed experts, uptake of nanoparticles into the aquatic biota is a major concern. At the cellular level, prokaryotes like bacteria may well be largely protected against the uptake of many types of nanomaterials, since they do not have mechanisms for the bulk transport of supramolecular and colloidal particles across the cell wall. With eukaryotes (i.e. protists and metazoans) the situation is different, since they have highly developed processes for the cellular internalisation of nanoscale (100 nm or less) and microscale (100 nm - 100,000 nm) particles, namely endocytosis and phagocytosis respectively (Na et al, 2003). In invertebrate animals, the cellular immune system, gut epithelium and hepatopancreas (digestive or midgut gland that is involved in food digestion), where present, is likely to be targeted. In fish, the liver is a probable target following endocytotic transport across the intestinal epithelium into the hepatic portal blood system followed by endocytosis into hepatocytes. Oberdörster (2004) has claimed in one of his studies that colloidal C60 fullerenes are taken up into the brains of largemouth bass and has hypothesised that this transport is via the olfactory nerve.

Studies with cultured mammalian cells have shown that fullerenes can cause oxidative damage and that their cytotoxicity is related to their lipophilicity (Sayes et al, 2004). Uncoated colloidal fullerenes may cause oxidative damage in the brains of largemouth bass. These findings indicate that interpretation of the effects of fullerenes is problematic and underlines the need for more rigorous experimental exposures. In some other studies, nano-size particles have also been demonstrated to enter the digestive gland cells of blue mussels and cockles by endocytosis. Uptake of sucrose polyester into the hepatopancreas of whole mussels from seawater was also demonstrated and was found to increase the up-take (160%) and cellular toxicity (122%) of the PAH anthracene (Moore, 2006). Exploitation of caveolar/endocytotic routes of entry to the cell may allow pollutant nanoparticles to embed themselves within the functional machinery of the cell in ways that are toxicologically quite different from conventional toxic chemicals. Nanoparticles situated in the endoplasmic reticulum, Golgi and endo-lysosomal system could conceivably act as foci of oxidative damage that could not readily be expelled from the cell; while generation of radicals could lead to organelle dysfunction.

Although nanotechnology is a major innovative scientific and economic growth area, it presents a variety of hazards for environmental health. While much attention has been focused on the development and potential benefits of nanomaterials in water treatment processes, results of the bibliometric cluster analysis and expert assessment confirmed the trend that during the last years concerns have been raised regarding their potential environmental toxicity.

Their physical (surface/size) properties may induce binding and transportation of toxic chemical pollutants, as well as possibly being toxic in their own right by generating reactive radicals. Also, release of manufactured nanoparticles into the aquatic environment may pose risks for the health of the aquatic environment which so far have not been explored sufficiently (Moore 2006). Based on the consulted experts' assessment, it is difficult to assess the effect of nanomaterials on health and the environment due to lack of appropriate methods and tools. Therefore experts indicate a requirement for a precautionary approach concerning the individual evaluation of new nanomaterials for risk to the health of the environment. Although the assessment of the field confirmed that current toxicity testing protocols should be generally applicable to identify harmful effects associated with nanoparticles, further research into new methods is still required to address the special properties of nanomaterials.

3.5.4 *Obstetrics and Gynaecology*

Prenatal Diagnosis was identified as an emerging topic within the field obstetrics and gynaecology. There is a variety of non-invasive and invasive techniques available for prenatal diagnosis. Each of them can be applied only during specific time periods during the pregnancy for greatest utility. In some cases, the tests are administered to determine if the fetus will/needs to be aborted. In other cases physicians and patients find it useful to diagnose high-risk pregnancies early so that delivery can be scheduled in a tertiary care hospital where the baby can receive appropriate care.

The prenatal diagnosis enables to detect main birth defects such as neural tube defects, chromosome abnormalities and other conditions, such as spina bifida, cleft palate, Tay Sachs disease, sickle cell anaemia, thalassemia, cystic fibrosis, and also fragile x syndrome and Down syndrome.

Ultrasonography in is a non-invasive diagnosis procedure that is harmless to both the fetus and the mother, it is not a new technology, however, the role of ultrasonography in obstetric practice has continuously evolved since its introduction more than 40 years ago. The introduction to clinical practice of fetal echocardiography of three-dimensional (3D) and real-time 3D (4D) imaging in the last decade has marked a dramatic advance and has been extensively researched by the scientific community. It is also one of the main emerging fields detected by the bibliometric cluster analysis in the Obstetrics and Gynaecology category.

The bibliometric cluster analysis detected that in recent years, the field of prenatal diagnosis has been inundated with three-dimensional (3D) ultrasound imaging tools for diagnosis in obstetrics. The current literature is extensive and filled with articles addressing the application of 3D ultrasound to virtually every aspect of fetal imaging as complementary to two-dimensional (2D) ultrasonography (Goncalves et al, 2005). The advantages of 3D include the enhanced identification of the nature, size and location of certain fetal defects (Xu et al, 2002); visualising structures in reconstructed planes; imaging of the fetal skeleton (Benoit 2003), spatial presentation of blood flow arborisation and vessels (Lee et al. 2003) and the ability to reconstruct a 3D

rendered image of the fetal heart that contains depth and volume which may provide additional information that is not available from two-dimensional (2D) ultrasound images (DeVore et al, 2003).

High hopes are also laid on 3D ultrasound on the precise measurement of the volumes of organs with irregular shapes (Ruano et al, 2011), especially regarding Down syndrome. A common phenotypic feature of individuals with trisomy 21 is a small nose, as originally described by Langdon Down in 1866 (Down, 1866). The definition of short nasal bone varies in different studies primarily because of variations in the reported normal ranges, which were essentially derived from two-dimensional (2D) ultrasound examinations. One possible explanation for the differences in measurements of the nasal bone observed in previous studies is that 2D ultrasound does not provide the ability to confirm that the nose is consistently measured in the exact mid-sagittal plane of the face. Research is currently underway into investigating whether the absence/defect of a nasal bone at the 12-week 3D scan could be a signpost that Down syndrome is present. This, together with the nuchal fold measurement, may give a more accurate indication of the risk of this syndrome and reduce the percentage of women who need an invasive diagnostic test. As the expert assessment concluded, advances in prenatal diagnosis and the use of 3D ultrasound have revolutionised prenatal care providing a powerful instrument for investigating the fetus' profiles. It provides genetic, anatomic, and physiologic information about the fetus(es) that can be used to make informed decisions regarding pregnancy and delivery, thus improving maternal and fetal outcomes. 3D ultrasound therefore has the potential to refine the diagnosis of facial profile anomalies or dysmorphisms, which often are a key sign for syndrome recognition and diagnosis.

3.5.5 *Public, environmental and occupational health*

Environment related risks were one of the research topics identified with the bibliometric analyses. The expert assessment confirmed the area to be an emerging field in public health with a focus on air pollution and temperature effects.

Air pollution

According to the expert assessment, exposure to ambient air pollution has been linked to a number of health outcomes: from modest transient changes in the respiratory tract and impaired pulmonary function; to restricted activity, reduced performance, emergency room visits and hospital admissions; and finally to mortality. The impact on human mortality and/or morbidity of exposure to ambient air pollution is a topic that has received extensive attention in the scientific literature in recent years all over the world (Brunekreef and Holgate, 2002, Samet, 2002).

People are exposed to a mixture of pollutants, emitted by a variety of sources. To a large extent, these sources are related to everyday human activity in a number of sectors, such as transport, energy production, industry and agriculture. Some of the pollutants are created in the atmosphere from precursor pollutants, and some travel

over long distances, cross national borders and create health risks far from the source of the emissions. The main health concerns are associated with particulate matter, that is, solid particles and droplets suspended in the air. The size, composition and origin of particulate matter vary, and it can be emitted directly or can be created from precursor gases in the atmosphere. Another common air pollutant of concern to health is tropospheric ozone, which is created from nitrogen oxides and other gases in the presence of ultraviolet radiation – emitted mostly from processes that involve combustion. Most of the published studies have been trying to describe the trends in relative risk estimates for air pollutants in the last decade. Studies investigating the association between daily time series of mortality and/or morbidity and daily time series of ambient air pollution concentrations have been at the forefront of this research (Puza et al, 2011). The reason for assessing such a trend is that a change in the relationships between air pollution and mortality may be indicative for changes in the composition of the air pollution mixture and, presumably, its toxicity (Shibata et al., 2008). Therefore, regular re-analyses of the associations is seen by the research community as a monitoring activity to keep a finger on the pulse with regard to the toxicity of the ambient air pollution mixture, which is of importance for air pollution policy measures (Fischer et al, 2010).

The studies until now have been mostly contributed to a better understanding of the health effects of exposure to ambient air pollution and added to the weight of evidence that has lead to stricter regulations (Bell et al, 2004). Based on the expert assessment, some of the most influential studies have been the multicity studies conducted in North America and Europe (Samoli et al, 2008). By combining estimates such studies have produced pooled estimates of the effect of pollution, both at national and regional levels. The pollutants of interest in these investigations have primarily been particulate matter air pollution and ozone (Wang et al, 2011).

Temperature

The effects of climate change and global warming on water resources, land resources and biodiversity, as well as their implications in regional communities are already noticeable and of great interest within the scientific community. Recent epidemiological studies show that climate change has affected adversely human health, directly and indirectly. An immediate effect is the physical injury, morbidity and mortality during or after an extreme event as a heat wave, flood or a hurricane, while environmental alterations induced by climate change may help the transmission of infectious diseases and may increase rates of vector borne diseases (malaria, dengue, encephalitis, Lyme), allergies and malnutrition (Costello et al, 2009).

Climate change projections indicate a further rise in global mean surface temperature, global mean sea level (IPCC, 2007) and changes in the frequency and the intensity of extreme weather events and precipitation in the next decades. Although it is not necessary that a hot environment will have adverse effect on human health, since human body has the ability to respond to environmental heat and acclimatise

to it, exposure to extreme heat may overcome the resistance of human body and be harmful for human health. Heat stress could cause rash, cramps, heat exhaustion, and heat stroke and may exacerbate underlying medical conditions, such as heart or lung disease (http://nyc.gov/html/oem/html/hazards/heat.shtml). Heat related mortality is certified as death from heat stroke or hyperthermia, when body's temperature is greater than 40°C. A high number of studies from recent years of heat related morbidity and mortality demonstrates a positive association between meteorological factors and thermal sensation or thermal comfort indices with heat related medical events and deaths (Golden et al, 2008, Schwartz et al, 2004). Europe experienced an extremely warm period in July and August 2003 that resulted in excess mortality of 11,435 people in France from 1st to 15th August and 1316 heat-related deaths in Portugal between 30 July and 12 August. In summer 2003, excess mortality was also reported in other European countries while information and data from that heat wave demonstrated that heat wave effects are probably underestimated. The increased health effect is observed especially among elderly and in particular by those living in older houses without air conditioning as well as due to harvesting effect (WHO, 2003). The number of heat related deaths is likely underestimated since deaths from other causes that can be aggravated by heat, may not be reported as heat related (Goncalves et al, 2006). Mitigation strategies to improve heat stress conditions experienced in the cities are mainly focused on the use of cool materials to decrease the absorption of solar heat and the use of additional urban green. The use of cool sinks like the water and the ground, the employment of solar control techniques and the minimisation of the anthropogenic heat seem to contribute highly to improve local microclimates (Karlessi et al, 2009).

The potential health effects of heat waves are substantial and the management of those effects is a major challenge for scientists and policy-makers (Pantavou et al, 2011). The expert opinion confirms that it is essential to be capable of responding to the adverse health effects of climate change through reliable and updated research.

3.5.6 *Geography*

In recent years, economic geography has been subject to new and promising developments in the research community. Recently, some researchers have sought to delimit and develop an "evolutionary economic geography" (EEG), aiming to create a more systematic theoretical framework for research (MacKinnon et al, 2009). EEG explains the spatial evolution of firms, industries, networks, cities and regions from elementary processes of the entry, growth, decline and exit of firms, and their locational behaviour. Based on the expert opinion, the evolutionary theory is attractive in economic geography, because it may develop into a general theory in economic geography while being applicable empirically to specific processes in space and time. This would make evolutionary theory compatible with a contextual view as advocated in economic geography, without giving up the ideal of developing a theoretical framework that goes beyond the specific and the unique (Frenken and Boschma, 2007).

The bibliometric analyses and expert evaluation identified the role of clusters and regions in a globalised world as one of the main research trends in economic geography. These findings are confirmed by the high number of publications proving that over the past years many scholars have been trying to understand processes of regional growth and change while drawing upon new ideas from for example evolutionary economics.

One of the main areas that has attracted most attention by the research community is clustering of economic activities. The central question for researchers regarding clustering has been how one explains the emergence of clusters in certain places rather than others. There are many different industries studied by various researchers and often the emerge of clusters is explained as a snowball process where clusters emerge through a spinoff process where more successful firms produce more and more successful spinoffs. Since spinoffs tend to locate in the same region as the parent firm, a cluster emerges once a single firm or a few successful firms start to create many successful spinoffs, which, in turn, create successful spinoffs themselves (Boschma and Frenken, 2011).

Based on the expert assessment, the EEG approach to clustering has quite a lot in common with the core model in New Economic Geography (NEG). For example, spatial distributions are derived from location choices and market competition rather than from regional differences in factor prices; competition is driven by scale economies at the firm level; clustering results from a self-reinforcing and irreversible dynamic process and the location of a cluster is path dependent on early decisions in its formative stage (Brakman et al, 2001, Krugman, 1991). However, EEG explains the spatial distribution of economic activity as a historical process, where path dependence in the location of economic activity results from the local spinoff creation. Even if the location of a cluster is historically contingent upon the location of a few well-performing entrepreneurs, the reason why a cluster emerged in a certain region can be explained ex post from the genealogy of entrepreneurs. In contrast, NEG explains spatial distribution of economic activity as resulting from optimizing behaviour by workers and firms leading to an instant (a-historical) equilibrium (Boschma and Frenken, 2011). Thus, NEG describes clustering as a general phenomenon underlying the historical formation of cities and shifts in the urban system that can be expected from changes in transportation costs or trade barriers (Krugman, 2011).

The EEG has had a number of recent empirical advances in the last years. It has provided the research community with new insights on clustering, and on the role of institutions in regional development. However, based on the expert assessment, the EEG is believed to be still under construction. Some of these empirical applications need to be developed further, and more advanced methodologies that can cope with longitudinal data are required to accomplish this task (MacKinnon et al, 2009).

3.5.7 *Conclusion*

All in all, a considerable share of the emerging topics as identified by various bibliometric analyses could be confirmed by the expert validation. However, we also

found examples were the bibliometric methodology alone seems not to be sufficient to identify the most emerging topics within certain disciplines. We observed an interesting and notable correlation between validation of fields and the ISI subject categories. Most discrepancies turned out in the arts and humanities, where experts had most difficulties to confirm the bibliometric results. In a number of cases the core articles identified did not represent the most recent research trends or areas in the respected fields. One of the main reasons for this deviation is that in fields like architecture and religion unlike in natural sciences there is still a substantial amount of articles published in non-English journals that could not be included in the bibliometric analysis. In addition, various regions in Europe have traditionally quite different research focuses in these areas. For example, in theology the research agenda in Germany is quite different compared to the Anglo-Saxon countries and in architecture and archaeology geographical and cultural influences are very important. Accordingly, it turned out that the bibliometric selection was not able to include all the different existing schools.

Regarding the social sciences the bibliometric results were considered to be more accurate in order to present trends compared to arts and humanities. As an example, the field of public health could be mentioned. However, also in this case some of the core articles were not considered as representing sufficiently emerging trends in the disciplines. As another short coming some of the articles seem to represent only very narrow topics that could be considered as important, however, were assessed as far less emerging than other broader themes that the analysis had not been able to detect.

The best results were achieved in the science category. This could be explained by the fact that in engineering and natural sciences geographical and cultural factors do not play a significant role and that the main publishing language is English. However, depending on the field some of the short comings described for the social sciences and humanities also turned out in the sciences category. As an additional problem the heterogeneous level of focus of different clusters was observed. In most cases clusters were rather broad, whereas in some other cases they were defined extremely specifically and could have been belonged easily also to broader identified clusters.

In summary, the bibliometric cluster analysis has proven to be a powerful methodology for identifying emerging fields of science. However, results of the qualitative analysis presented in the previous sections clearly indicate that it needs to be complemented by qualitative expert assessment in order to create robust and validated results.

Acknowledgement

Figures 3.1 and 3.2 are reproduced from Glänzel and Thijs (2012) with permission of the publisher.

Chapter 4

Coverage of emerging topics by ERC-Grant applications

4.1 INTRODUCTION

The second part of the project comprises a mapping of ERC-funded research to the emerging topics. For this purpose, the content of the applications for starting grants is compared to the literature within the detected emerging topics. By such a comparison it is possible to show to what extent the ERC grants refer to these emerging topics or whether they refer to clearly distinct topics. The proposals for the 2009 Starting Grants Call were used.

In addtion the mapping results were also expected to provide added-value for ERC operations and in specific for both ex-ante and ex-post the evaluative or selective procedures. Prior to the review by the panels, the applications could be classified into a structure inherent to the landscape of science, proposals could be labelled as being relevant to emerging or established topics. Such mappings could also provide a quantitative bibliometric mirror for the qualitative expert assessment. Ex-post, the mapping can be used to support a critical reflection on the selection procedure.

When developing a mapping methodology, such as the one presented here, the existence of false-positives and false-negatives cannot be excluded. It is crucial to find the right balance between both types of errors taking into account the requirements of the purpose of the mapping exercise. This always implies qualitative decisions. In the case of ERACEP we assume that the acceptance of false positives is higher than that of false negatives. It is the aim of the procedure to detect applications that are relevant to the emerging topics. As both document sets serve a clearly distinct purpose and they are targeted to a complete different audience, focusing too much on avoiding false positive would deteriorate the mapping and result in a large number of false negatives. The availability of not only emerging topics but also of the other clusters provides a good baseline for comparison as we can expect that the error rate will not differ amongst topics.

4.2 DATA

4.2.1 *ERC Starting Grants*

The applications that have been used for the mapping exercise stem from the 2009 Starting Grant Call. ERC has provided a set of 932 eligible applications of which 165 finally proved to be successful. Part B1 of these proposals was provided in pdf format. This part B1 is expected to contain the following sections.

- A cover page with the Principal Investigator's (PI) name, host institute, full title, acronym and proposal summary.
- Section 1a: The Principal Investigator with the scientific leadership profile containing early research career achievements and important contributions. Also an extensive curriculum vitae and a publication list are included in this section.
- Section 1B: The extended synopsis of the project proposal.

Besides the pdf files containing the proposal as written by the applicant, ERC provided also a data file containing meta-data:

- Proposal ID
- Acronym
- Allocated panel
- Success
- ERC keywords specific to the panels
- Applicant provided keywords

The pdf files have been converted into a plain text format that is more suitable for processing. For some applications it was not possible to convert the file into an appropriate format for analysis. These applications have been omitted from further analysis. After the pre-processing and preparation of the data files a set of 885 applications was retained of which 157 were successful.

Due to the open nature of the application procedure and the respective open structure of proposals, it was extremely difficult to have an automated separation of the different sections in the proposals. As a result the proposals are processed in full and also information about earlier research of the PI is taken into account for the mapping of the proposal to a emerging field.

4.2.2 *Publication Data*

For this mapping exercise, all citable documents indexed by Thomson Reuters in the considered subject categories are taken into account. This is the same data set that has been used for the hybrid clustering in the first part of the project.

- Engineering, biomedical
- Energy & fuels

- Environmental sciences
- Obstetrics & gynaecology
- Public, environmental & occupational health
- Geography

This resulted in a set with 164.220 unique documents of the types *'Article'*, *'Letter'* or *'Review'*. As the assignment to multiple subject categories is possible in the Web Of Science database, a small share (1.7%) of documents is shared among subject categories. All text fields like title, abstract and keywords are relevant for this mapping and are processed. It is important to stress here that the full text of the publication is not available in the Web Of Science database.

4.3 METHODS

For the mapping of ERC applications to the selected emerging topics a full text matching approach was chosen. In Figure 4.1 an overview of the methodology is provided. In the first step the data are collected from the relevant sources. For the topics this is the database built at ECOOM containing the Web Of Science publications and for the applications this are the text files extracted from the PDFs mentioned above. The title, abstract and all keywords appearing in a paper are merged together in order to create a single text field. Also on the side of the applications, only one text field is available. One of the strengths of the method is that the same processing or analysing tools can be used at both sides of the schema. These text fields are then parsed and indexed by the Lucene Text Indexer. More details on this text indexer are given below. From both indexes a set of common terms is extracted. Document by term matrices are created containing the raw frequency of each term in the document. These are of course very sparse matrices. A weighting is then applied to temper the weight of common words and boost the others. These weighted matrices are then combined through the calculation of a cosine between the respective vectors resulting in a paper by application similarity matrix. Taking the average similarity over papers within a certain topic results in a topic to application similarity. The following sections will give a more in-depth description of the steps within the mapping procedure.

4.3.1 *Text Indexing*

For the processing and indexing of the texts we use Lucene which is a high-performance, full-featured text search engine. It is an open source project supported by the Apache Software Foundation and it is free for download at `http://lucene.apache.org/core/`.

In order to enhance the comparability between the two document sets we developed a single set of tools within Lucene that is applicable to both the applications and the scientific papers. Several steps had to be taken in this process, each of them has its own parameters and settings influencing the final result.

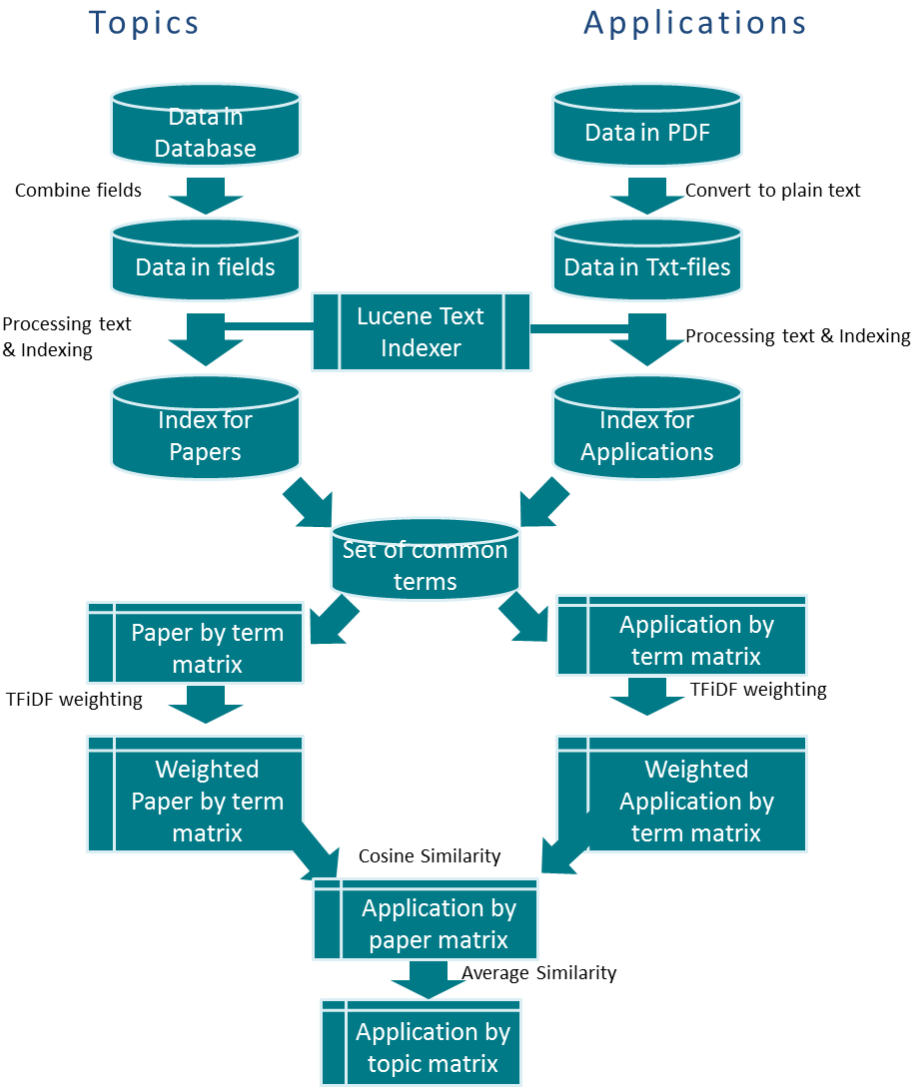

FIGURE 4.1. Schematic overview of the work-flow for the mapping of topics and applications

- Lower-case: All characters are converted to lower case. Although most of the procedures that follow are case-insensitve we made sure to avoid any mismatch due to different casing.
- Tokenization: Punctuation and white spaces are used to separate the distinct words in a text. In this implementation we have chosen to only apply single words and not phrases or keywords.
- Removal of stop-words: Words that occur in many documents without any distinctive power or meaning are removed from the documents in this step. A list of stop-words was compiled prior to the processing. This list contains words like 'and', 'or', 'not', 'is'... but also words that frequently used in the context of abstracts in the publication database: 'copyright', 'by', 'Elsevier'.
- Stemming: In this process, words are reduced to their stem or root form. This stem itself does not have to be a valid word. We use the stemming algorithm that has been developed by Martin Porter in 1980. The Porter-stemmer has become a standard algorithm for stemming and the Lucene software package has a version of the official implementation of the algorithm embedded.

As a result we obtain for each document (paper or application) a list of terms with their frequency and offset in the document. This data is stored in two separate text indexes. Based on both indexes, a list of common words between papers and applications is compiled. Only terms occurring in both sets are considered to be relevant in the remaining analysis. Using this list two document by term matrices were compiled. These matrices contain for each document the raw frequency of each term in the common list. As most terms do not occur in a particular document, most terms within one document have a frequency equal to zero. These matrices are called to be sparse. Raw frequency scores are not appropriate for the purpose of our exercise as they are sensitive to the length of a text. Raw frequencies will be higher for longer texts. A normalized measure is used instead of the raw counts. In this case we use the TFiDF-value.

TF: Term frequency is the count of occurrences of a particular term in the document

iDF: The inverse document frequency. The document frequency refers to share of document in which the terms occurs. In this analysis we use the log of the inverse of this share. The in the following equation N denotes all documents in the particular set and n_j the number of documents containing term t_j (cf. Eq 3.1).

$$w_{ij} = f_{ij} \log\left(\frac{N}{n_j}\right)$$

4.3.2 *Text Matching*

The matching of the applications to the papers is done by calculating the cosine similarity between the respective document vectors. Because both document by term matrices are restricted to only those terms that occur in both data sets, the vectors

representing the documents have the same dimensionality and each dimension refers to the same term. This means that both applications and publications have mathematical representations in one and the same vector space. Within this vector space, similarities among vectors can be calculated using a cosine measure (cf. Eq 3.3).

$$\cos(Appl_a, Pub_b) = \frac{\sum_{j=1}^{n} w_{aj} w_{bj}}{\sqrt{\sum_{j=1}^{n} w_{aj}^2} \sqrt{\sum_{j=1}^{n} w_{bj}^2}} \tag{4.1}$$

The value of this cosine ranges between 0 and 1 where 0 means that application and publication do not share any term.

The mapping of the proposals to topics or detected clusters is based on the average similarity that the applications have to all the linked papers in one cluster. A paper is linked a proposal if they share at least one term. So the average is taken over all non-zero similarities. A threshold of 0.025 was set in order to classify an application relevant to a certain topic. This threshold was decided upon after qualitative validation of the results. Quantitative methods like discriminant analysis or other classification methods were not suitable as there is only one dimension to score the mapping -average similarity- and only two categories 'relevant' or 'not relevant'. Quantitative methods could set the threshold but we would have to decide on the preferred success rate.

Applications can be relevant to multiple topics as they can share high similarity with papers from distinct topics. The purpose of the analysis is not to classify the proposals to unique topics but to identify those applications that are relevant to emerging topics.

4.4 VALIDATION

For the validation of the results of the mapping we used a manual, qualitative procedure. The aim of the validation procedure was three-fold:

- Estimation of the appropriateness of the developed mapping methodology and adjustment if necessary.
- Decision upon the cut-off threshold
- Removal of the false positive applications.

The above mentioned balance between false positives and false negatives is taken into account with a stronger focus on avoiding the latter one. The procedure consists of three steps.

Reading of individual application

The procedure starts with reading the title and summary of the mapped and unmapped applications. Almost all mapped applications were inspected but only a sample was taken from the unmapped applications. The sample was partially based on the presence of journal titles or keywords in the application. This allows for a qualitative judgement

of the relevancy of the application in the subject category. While reading the proposals and comparing their content to the matched fields it became clear that the methodology developed for the mapping was appropriate. The text matching algorithm was able to identify the relevant fields for most of the applications. Two possible sources of error were identified at this level of validation. The different meaning of specific terms across fields could distort the result. As example we refer to the word 'Cell', this has a particular meaning in the life sciences but it is also an important keyword in the field of 'Energy & Fuels' referring to fuel cells or cells as components of batteries. Several applications that were linked to the latter field were removed from the set as they were in the medical or biological sciences.

Applications that were on topic with the emerging clusters were labelled as such. This label was used later to identify false negative mappings.

Selection of papers with highest similarity

For several applications, the similarity with individual scientific papers was validated. At this point we did not make any distinction between relevant or non-relevant subject categories. Those papers that had the highest similarity were inspected. Within each selected pair of application and paper the common words were extracted. This allowed us to extend if necessary the list of stop-words or adjust the preprocessing in the methodology. In some cases we found that the terms that are substantive to the linking are located in the section of the PI's earlier achievements.

Selection of papers within topic

After the calculation of the average similarity between an application and topic, a more detailed analysis of the content of the application and the linked papers was done. The most similar papers within the topic were identified. The application was compared with the content of the papers and was assigned to one of three classes: a) on topic, b) relevant to topic, c) irrelevant. The last category is considered to be a false-positive. As a result of this step several papers dealing with *HIV* or *AIDS* in a pure medical context were removed from the set.

4.5 RESULTS

As mentioned above, after the pre-processing and preparation of the data files a first set of 885 applications and a second set of 164.220 unique papers were retained. All of them shared a least one term with a document in the other set. So no document was omitted due to the restriction of the terms to the set of common terms. After the calculation of the average similarity between proposals and topic and the assignment based on the threshold 298 applications were mapped with at least one topic, 173 applications had multiple assignments. The mapping is not restricted to the identified emerging topics.

The results are presented by subject category. Emerging topics are labelled by an asterisk.The tables list the distinct topics that were identified in the first part of the project within each of the subject categories. For each topic the table presents the number of matched applications in absolute and in relative values and the share of successful applications within the matched ones. It should be pointed out the these values include multiple assignments.

4.5.1 *Biomedical Engineering*

About 112 proposals are mapped with one of the nine topics that were identified in the field 'Biomedical Engineering'. The two emerging clusters are rarely addressed. Only four applications are relevant to 'Brain Computer Interface' and just five to 'Kinematics'. However, a strong focus was found on two other topics namely 'Nanostructuring' and 'Scaffolds'. Ex-post we can conclude that the highest success rate was found in 'Tomography' while the emerging topics had no successful applications. This might of course be influenced by the very low number of proposals that were identified as being relevant to the topic. Overall 18% of all applications in this field were successful.

TABLE 4.1. Mapping results for Biomedical Engineering

Cluster Label	Matched Applications	Relative	Share of Success
Catheter Ablation	3	3%	0%
Brain Computer *	4	4%	0%
Nanostructured	47	42%	17%
Cartilage	19	17%	16%
Tomography	32	29%	28%
Scaffolds	62	55%	15%
Bone Cement	23	21%	9%
Cardiac	13	12%	15%
Kinematics *	5	4%	0%

4.5.2 *Energy & Fuels*

Within the subject category 'Energy & Fuels' we found eight topics of which two were identified as being emergent: 'Bio Fuels/Diesel' and 'Fuel Cells'. The mapping resulted in 64 applications that are relevant to at least one of the topics in this field. Both emerging topics are well addressed by the proposals. Almost half of the topics show some similarities with the content in the literature on 'Fuel Cells' while 'BioFuels' scores a matching rate of 33%. The more conventional topics -coal, batteries and oil- are less covered by the applications. The success rate of the emerging topics is good. One in five proposals on 'Fuel Cells' got selected and one in seven on 'Biofuel'. The proposals on 'Renewable Energy' have the highest success rate 43%. This is well above the average 22% for the subject in total.

Table 4.2. Mapping results for Energy & Fuels

Cluster Label	Matched Applications	Relative	Share of Success
Oil, Bio Oil	4	6%	0%
Batteries, Storage	5	8%	0%
Renewable Energy	14	22%	43%
BioFuel *	21	33%	14%
Kinetics	22	34%	9%
Fuel Cells *	30	47%	18%
Coal	10	16%	0%
Combustion	18	28%	11%

4.5.3 *Environmental Sciences*

We could map 71 applications to the subject category 'Environmental Sciences'. Many of these proposals were linked to the topics 'Bentic Zone' and 'Waste Water', while the topics 'Diversity', 'Sediments' and 'Soil' could attract still 20% or more of the applications in this field. The two emerging topics 'Radiation' and 'Nanopolution' were linked to a small number of proposals. But, remarkably, four out of the five proposals relevant to 'Radiation' were approved and got funding. The field itself had an average success rate of 18%.

Table 4.3. Mapping results for Environmental Sciences

Cluster Label	Matched Applications	Relative	Share of Success
Diversity	17	24%	35%
Radiation *	5	7%	80%
Nanopolution *	4	6%	0%
Waste Water	20	28%	10%
Bentic Zone	23	32%	17%
Sediments	18	25%	11%
Air Pollution	10	14%	10%
Soil	14	20%	21%

4.5.4 *Obstetrics & Gynaecology*

The field of 'Obstetrics & Gynaecology' was linked to 39 applications and most of these proposals showed relevancy for the topic 'Cancer' within the subject field. Of course, here we need to state that many proposals related to other forms of cancer were mapped with this topic which is mostly devoted to 'Ovarian Cancer'. All other topics have very small numbers of applications related to them. As a consequence, no ex-post statements can be formulated.

Table 4.4. Mapping results for Obstetrics & Gynaecology

Cluster Label	Matched Applications	Relative	Share of Success
Prenatal Diagnosis *	1	3%	0%
Cesarean	4	10%	25%
Pregnancy Complications	1	3%	0%
Menopause	3	8%	33%
Surgery: Incontinence	0	0%	
Infertility	5	13%	0%
Cancer	28	72%	18%

4.5.5 *Public, Environmental & Occupational Health*

This field has the lowest number of matched applications, 37 of which only four were successful. Based on the specific nature of the subject category several matched proposals were removed from the set. Most of them were focusing on pure medical aspects of HIV and AIDS and not on any socially relevant issues of the disease. As a result we retained 7 proposals that were considered to be relevant to the topic 'Aids in Africa' of which one got selected. Two topics were most addressed in the linked proposals: 'Health Policy' and 'Gender and Family'. Both topics had only two successful applications. The low number of mapped and successful applications is in contrast to the other subject categories.

Table 4.5. Mapping results for Public, Environmental & Occupational Health

Cluster Label	Matched Applications	Relative	Share of Success
Work related	4	11%	0%
Living Environment *	1	3%	0%
Aids, Africa *	7	16%	17%
Quality of Life	8	22%	13%
Health Policy	17	46%	12%
Tobacco	7	19%	14%
Gender and Family	12	32%	17%

4.5.6 *Geography*

68 proposals are mapped with one of the five topics that were identified in the field 'Geography'. The emerging cluster that is labelled 'State-Region' is addressed very well, nearly two out of three proposals could be mapped with this topic. However, validation showed that due to the more general nature of the language used in the literature and in the proposals, the link between proposals and topics or field seems to be overestimated.

Nevertheless we observed a high success rate for all proposals that were linked to topics in 'Geography'. One out of four applications got selected. This is the highest success rate of all subject categories.

Table 4.6. Mapping results for Geography

Cluster Label	Matched Applications	Relative	Share of Success
Living	13	19%	31%
Natural Resources	23	34%	35%
Community	26	38%	31%
State-Region *	43	63%	28%
Diversity	26	38%	19%

4.6 DISCUSSION

The mapping methodology that has been developed within ERACEP proved to be appropriate for linking applications to topics as defined by a set of scientific papers. These text based links using the full text available in the proposal and the title, keyword and abstract as indexed in bibliographic databases do not require specific parsing of the documents to extract journal titles or references to individual papers. The methodology allows a mapping of the applications as soon as they are submitted with predefined emerging topics and the strength of the link can be expressed by a quantitative indicator. Of course such a mapping methodology is not free of problems. The right balance between the two types of errors has to be found and discussed prior to the mapping. The weight of false positives and false negatives highly depends on the purpose of the exercise. We tried to avoid the latter one and as a consequence had to accept a higher rate for the false positives.

Also methodological problems popped up along the road we took. The applications as they were provided contain all text that has been submitted. The method would improve from separating the different sections in the applications. This could be done at the level of submission by the applicant or while processing the provided text. The first solution would give probably the best results but would hamper the freedom in elaborating proposals that the applicant has been given in the past. The latter solution is more prone to errors.

Another problem is the presence of common words that have a different meaning across fields and disciplines. Such words are *cell*, *migration* or *group*. This could be partially be solved by the use of phrases while indexing the terms. The use of phrases has however also drawbacks. Especially in emerging topics, a common vocabulary has not yet been established and different phrases are used to express the same concept. Also the single terms retain their importance in the particular field. As a consequence, it is difficult to decide on the weights that are given to the phrases and single terms. The application of wrong weights might even distort the mapping instead of improving it.

Some phenomena or subjects are studied from different perspectives. Moreover, some perspective might already have a long tradition while the other is just popping up but has a high relevance. We found this with the topic 'Aids in Africa' which is emerging in 'Public health'. It remains a challenge to separate the applications that deal with the societal impact of the disease from those within a pure medical perspective.

As a result we recommend that the developed methodology is only applied in combination with a qualitative validation.

Nevertheless, the obtained results can be of value from a user perspective both ex-ante as ex post.

Ex ante

The ERC grant applications can be classified according to a dynamically evolving inherent structure of the landscape of science. This landscape and its dynamics can be presented independently from and prior to the classification. The detection and visualization of the dynamics allow the ERC also to organize the evaluation panels to align them along such structures. This could be used to improve the classification of applications to panels as this can be based both on the content of the proposal and the topical structure of the panels. The grants can be labelled as being relevant to thematic clusters (established or emerging) and this mapping could be used for a pre-selection.

Ex post

After the evaluative procedure, the mapping of the successful and non-successful applications in the thematic clusters can facilitate the identification of differences between scientific fields based on coverage of emerging topics and success rate for all topics. The mapping will also provide hints for improving the submission procedures or the evaluative process.

Chapter 5

Conclusions and discussion of results

This final chapter with conclusions and discussions of the results of the ERACEP project will try to close the circle by reflecting on the outcomes within the framework of the initial proposal. This proposal describes the main objectives of ERACEP including possible impact of its results on ERC's operation. It also introduces the concept of the project and the methodological approaches that will be taken or newly developed.

Accordingly, the main aim of ERACEP

> *. . . is the identification of topically emerging research areas and to analyse to what extent the activities supported by the ERC cover and contribute to these research areas.*

In a broader sense the objective of the project was also to contribute to question, whether the bottom-up approach of ERC funding schemes meets the expectation being able to support new lines of research with impact on the development of dynamic creative and innovative areas of research in Europe. It is in this light that the following impact was mentioned:

- The ERC-funded research will be put into a systematic context with the worldwide research activities in emerging areas.
- It will be mapped within the landscape of science with regard to different dimensions of performance (e.g., scope, dynamics, impact, international uniqueness).
- In particular, the analysis will show whether the open calls of the ERC encourage the submission of proposals which really refer to dynamic areas of research and whether the finally selected ones fulfil this criterion. This information will allow to assess to what extent ERC-funded activities contribute to the creation of new scientific knowledge.

- Based on these findings, the appropriateness of the ERC procedures can be reflected in a systematic way so that the strategic objectives of the ERC support can be elaborated further on a rational basis.

Also with respect to perspectives to be taken and new approaches to be developed, the proposal is quite clear. The next section discusses the usability and applicability of the methodology that has been develop during the ERACEP project. In the last section it is shown how the outcomes of the project can contribute to the aims and objectives set in the proposal.

5.1 MAIN RESULTS

The concept of the project was to work from two different perspectives. Both are elaborated in detail in the previous chapters. It was clear from the start of the project that new approaches had to be developed for both tasks at hand. Firstly, for the detection of emerging topics, various methods have already been suggested like Delphi surveys (cf., Cuhls and Grupp, 1998), expert-based brain storming (for instance, Grupp, 1993, where nanotechnology was already discussed as relevant future field and the interfaces between traditional scientific disciplines were recognized as future drivers of research), or expert conferences. Often standard *future* themes were proposed during such sessions and only a few suggestions were forwarded as to really new promising fields at lower level of aggregation.

Secondly, also for the mapping of ERC applications to emerging topics no proven methodology was available at the start of the project. The proposal states that the ERC-grants will be classified to the system of emerging fields but it does not mention any particular technique to accomplish this task.

Therefore it is clear that the project has a true developmental and experimental nature. As a consequence the trajectory of steps to be taken has changed during the course of the project. One of the most important changes was the decision to narrow done the scope of the analysis and to focus more in depth on the evolution of certain clusters instead of a broad mapping distributed over all areas of science.

5.1.1 *Detection of emerging topics*

ERACEP started with an open investigation in all fields of science, social science and arts & humanities. This was chosen to reflect as much as possible the open, flexible bottom-up approach of the ERC funding schemes. In a first step, dynamic fields were identified based on publication activities. As stated in chapter two, the rational behind this identification step is the notion that dynamic growth of a specific topic reflects on the one hand an increasing interest of scientists in this topic, so that more research groups are doing research in the respective area, resulting in an increasing number of scientific publications. On the other hand, dynamics could also indicate that a more or less constant number of scientists are increasing their research activities, resulting again in a growing publication output.

For the actual detection of the emerging topics within the dynamic fields an analysis based on three bibliometric components was developed. First a cluster analysis on the fields was applied, then a fine-grained representation of clusters was built based on core documents and the third component is a diachronic analysis of the evolution of links among clusters and topics over time.

Dynamic Fields

For the identification of the dynamic fields, a Sharpe ratio was calculated for all the subject categories included in the Science Citation Index Expanded (SCIE), the Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI) of Thomson Reuters. This resulted in the selection of twenty categories of dynamic fields (Table 2.5). In particular, 13 fields have been selected from the sciences, 4 from the social sciences and 3 from the humanities. This selection also ensures to cover all areas of science. Separate analyses were conducted in order to comply with the distinct properties of each database.

As the identification of dynamic fields is based on the subject category system from Thomson-Reuters, this analysis cannot escape from the properties and peculiarities inherent to the classification. This system is a journal classification scheme where individual journals are assigned to one or more subject categories. This implies immediately that there is no classification at the level of individual papers. All papers in one journal belong to the same categories. Next, multiple assignments are very common in the system. This resulted in the need for aggregation of small categories with high overlap. The system itself is also dynamic meaning that new categories can be added. Nanoscience and nanotechnology is such a new category.[1] The dynamics of a field could thus be caused just by the changing properties of the bibliographic database[2]. As the scope of a field is extended, more journals become relevant to the field and the field will show a certain growth. The coverage of the database has expanded during the period of time of our study and it is known that this happened especially for the social sciences and arts & humanities. Those artefacts have to be taken into account when judging the quantitative measures used to express the growth of a field.

Another issue related to the choice of subject categories as a starting point is the a priori boundaries that are set in the datasets. Only activities of a topic within a selected category can be analysed. The next steps in the process take this issue into account, the developed methods can be applied independently of the subject categories to any set

[1] This only happens after a certain topic exhibits enough growth to become an own field, which presents a certain critical mass, some established communities and its own communication channels like journals. It is clear that it is not the aim of this project to detect these new fields -indicated earlier as *Standard Future Technologies*, but to detect topics at a much lower level of aggregation.

[2] It is important to note here that the application of other data sources would not be beneficial to the outcome of the project. Only the Scopus database from Elsevier could be a valid alternative but it comes with its own peculiarities. Google Scholar or Microsoft Academic Search are too uncertain as source for this type of research

of selected publications. This means also that any method for the detection of dynamic fields can serve as an input.

Overall, the Sharpe ratio proved to be a powerful indicator of growth and an appropriate criterion for the selection of the most dynamic fields.

Detection of emerging topics

For the structural topic analysis we developed a combination of three bibliometric components. The first component, cluster analysis, is traditionally based either on citation links (bibliographic coupling, cross-citation, co-citation analysis) or textual links (co-word analysis, term representation). Both approaches have advantages and shortcomings. The main advantage of citation-based methods are their discriminative power. This is contrasted by a serious disadvantage: Citation-link matrices are extremely sparse and citation-cased methods tend to "underestimate" links among documents. Furthermore, citation links generate binary measures taking the value 1 or 0 according as there is a citation link between two documents or not. By contrast, text-based measures are based on term frequencies in documents, which as such provide a natural weight underlying the similarity/distance measures used for the analysis. Link matrices are furthermore less sparse than their counterparts in the citation space. These advantages are cancelled out by two serious problems: The lower discriminative power, which results in "overestimating" links among documents and the dimensionality problem. In ERACEP, we decided to apply an hybrid approach, which actually forms a combination of the two traditional cluster techniques. The citation-based component uses bibliographic coupling, which has several serious advantages compared to co-citation analysis (cf. Glänzel and Czerwon, 1996). The most important advantages are that practically all papers have references (even in the social sciences and humanities) and that no response time is needed (for citing the literature). The textual component is based on term frequencies, where terms extracted from and stemmed for titles and abstracts and keywords. Both components are combined to a hybrid measure of similarity between documents using particular weights for each of them.

For the clustering we used Ward's agglomeration with different stopping rules for decision on the final number of clusters. For obtaining the optimum number of clusters the following methods have been used.

- Local/global maxima of silhouette scores (cf. Rousseeuw, 1987)
- Dendrogram
- Evaluation of classification (based on silhouette values)
- Qualitative inspection based on keyword representation and topic representation by 'core documents'

We tried to use almost the same weighting parameters for the combination of citation links and lexical component in fields with similar "hardness" of science. The choice of $\lambda = 0.83$ for the sciences and $\lambda = 0.75$ for the social sciences and humanities provided very good clusters according to the silhouette values. In the case of most

clusters, the keyword representation worked well, however, in some clusters of several disciplines we could not identify sufficiently specific keywords.

For the labeling and representation of the clusters we elaborated the notion of core documents that was coined by Glänzel and Czerwon (1996). According to their definition, core documents are defined as papers that have at least $n > 0$ links of at least a given strength $r > 0$ according to the underlying similarity measure.

These core documents are the second bibliometric component of the methodology and they turned out to be instrumental for the qualitative validation that has been set up. It was much easier to understand the content of a particular topic if it could be represented by a limited set of papers that are really in the centre of the topic.

In the third component of the methodology we applied a diachronic analysis of the links between clusters over time. This analysis starts from independent cluster solutions. In order to be able to detect emerging topics in an early stage, it was important to have available analyses that are separated both in methodology and in time. Therefore we repeated the cluster analysis for a second time period not overlapping the first one. For most fields the periods 1999–2003 and 2004–2008 were chosen.

In particular, citation links between core documents in one period to all publications in the clusters of the other period were used to determine links between the structures of the two periods. This method is expected to reduce noise that might otherwise be caused by cross-citation links of less relevant documents far from the medoids of the clusters. On the other hand, cross-citation analysis between core documents alone usually provides an insufficient number of links necessary for creating the concordance.

The emerging topics can be identified by the above-described method and the danger of false positives is minor and not very likely. However, the requirement of a critical mass of documents might imply the danger of not identifying new emerging fields if the number of relevant publications does not meet this criterion. In such situations most quantitative and computerised techniques will be faced with similar problems that can only be overcome by communication with experts.

This methodology was applied to six selected dynamic fields and resulted in the detection of ten emerging topics

- Engineering, biomedical
  - *Brain-Computer Interface*
  - *Kinematics*
- Energy & fuels
  - *Biodiesel*
  - *Fuel Cells*
- Environmental sciences

  - *Nano-pollution*
  - *Radiation*

- Obstetrics & gynaecology
  - *Prenatal diagnosis*
- Public, environmental & occupational health
  - *Aids in Africa*
  - *Environment*
- Geography
  - *State & Region*

The identification of emerging topics is not only merely based on the number and size of occurring topics but it is also related to the cognitive-epistemological structure of the research field. Based on the different links that were distinguished we found three cases of cluster evolution that can indicate new, emerging topics:

(1) Existing cluster with an exceptional growth,
(2) Completely new cluster with its root in other clusters and
(3) Existing cluster with a topic shift.

Validation

The validation of the obtained results was always considered to be an important phase during the course of the project. For a validation of the identified emerging topics as described in the previous chapters we carried out a qualitative assessment of the fields and topics by researchers and experts from respective disciplines. In total 21 scientist and experts from various academic backgrounds from different European countries were consulted. The majority of the experts were from academia or public research institutions and have been selected based on publication record and experience in the field. All in all, a considerable share of the emerging topics as identified by various bibliometric analyses could be confirmed by the expert validation. However, we also found examples were the bibliometric methodology alone seems not to be sufficient to identify the most emerging topics within certain disciplines. We observed an interesting and notable correlation between validation of fields and the ISI subject categories.

Most discrepancies turned out in the arts and humanities, where experts had most difficulties to confirm the bibliometric results. One of the main reasons for this deviation is that in fields like architecture and religion unlike in natural sciences there is still a substantial amount of articles published in non-English journals that could not be included in the bibliometric analysis.

Regarding the social sciences the bibliometric results were considered to be more accurate in order to present trends compared to arts and humanities.

The best results were achieved in the science category. This could be explained by the fact that in engineering and natural sciences geographical and cultural factors do not play a significant role and that the main publishing language is English.

In summary, the bibliometric cluster analysis has proven to be a useful methodology for identifying emerging fields of science. However, results of the qualitative analysis presented in the previous sections clearly indicate that it needs to be complemented by qualitative expert assessment in order to create robust and validated results.

Additional international analysis and benchmark studies

The availability of set of publications belonging to a certain topic enables a broad range of more bibliometric analyses. Citation indicators could be calculated, as is national contribution and international co-operation. Because the topics are extracted from the dynamic fields we could use these larger publication sets as reference points for the results within the emerging topics. With this analysis questions such as the following can be answered:

- Which countries are the most active players in the emerging topic?
- To what extend do countries active in these topics collaborate with each other?
- Do papers from European scientists differ from American papers?

We observed substantial differences in national contribution and international collaboration amongst the distinct fields and topics.

First there are differences in national activity across the topics. We see that USA is contributing to about 60% of all papers in the topic that we labeled 'Aids in Africa' while its overall share in the field 'Energy & fuels' is just 17.6% (14.7% for biodiesel; 21.6% for fuel cells). For the People's Republic of China the opposite can be detected. Although Africa is of high economical importance to China, they contribute only to 2.5% of all papers on 'Aids in Africa' while their share in papers on biodiesel (17.3%) exceeds the share of US. It is striking that the US takes a clear lead in the topics and fields with a medical nature while their leadership is clearly challenged in the fields like Environmental sciences and also 'Environmental issues' in Public Health.

Collaborative networks among countries could also be built based on the international collaboration. These networks were constructed both for the field in total as for the specific topics. The density of the network in 'Environmental Sciences' is striking, nearly all European countries have very strong links with each other. Also the position of the US located outside the European cluster is remarkable. This can also be observed in the emerging topic 'Environmental issues' in 'Public Health'. This is contrasted by the very central role that the US plays in 'Brain-Computer Interface'. Another interesting observation is the lose network in 'Energy and Fuel' and its two emerging topics. Apparently, nations are well aware of the technological and economical advantages that can come with developments in these fields. Contrary to the medical sciences, nations are more reluctant to share their newly gained knowledge amongst other countries.

These national data and networks of international collaboration enable us to put the activities in the fields and topics into a geo-political context.

5.1.2 *Mapping of ERC grant applications*

The second perspective in this project is build around the mapping of ERC grant applications to the emerging topics that have been detected in the first phase. We used the applications from the 2009 Starting Grant Call.

As already mentioned in Chapter4, the balance between false-positives and false-negatives plays a crucial role in the applicability of the developed methodology. The decision on the importance of both types of error depends only on the objectives of the mapping exercise. As it is the aim to investigate to what extend the ERC supported activities cover and contribute to the detected emerging topics, the avoidance of false negatives is more crucial to the procedure. As a consequence we had to build in a higher acceptance of false-positives. Of course all possible efforts were taken to reduce both types of error. The availability of not only emerging topics but also the other clusters provide a good baseline for comparison as we can expect that the error rate will not differ amongst topics.

While developing this mapping methodology, it became clear that the distinct targeting of the documents had to be taken into account. Scientific publications and grant applications serve a complete different purpose and are each written for another audience. Scientific papers are meant to be read by peers who are aware of the current state-of-the art and thus assume a lot of tacit knowledge. Also the vocabulary in the paper can be highly specialised. Applicants for grants do not know how familiar reviewers will be with their field. In fact, applicants often can assume that the reviewers are not experts in the field of the application. Therefore, writers try to avoid specialised terms and keywords.

One of the problems that we encountered while processing the applications is the open format used to submit the form. It was not possible to automatically parse the documents to extract the relevant sections from the PI's prior achievements. As a consequence, the complete application form B1 was analysed and used for matching.

We tried to develop a mapping in two stages that would mirror the two steps in the first perspective of the project. For the detection of the emerging topics we started with the identification of dynamic fields and later we used clustering to find the final emerging topics. For the mapping we tried first to assign applications to ISI subject categories and later to map the proposal to individual topics.

The assignment to subject categories was conceived analogous to the ISI classification system, where journals are assigned to classes. In this stage, we identified journal titles as they occur in the application. Through these journals, corresponding subject categories were mapped to the proposal. After validation, this mapping turned out to be not successful. Often journals were abbreviated in the references or journal titles only consist of common words that occur in many texts (eg. Cell, Nature, Science). The validation exercise proved however to be very useful as many applications were

indeed manually assigned to fields and topics. For example all applications to the 2009 Starting Grant Call have been classified manuallly into ISI fields.

The second stage provided us with a successful mapping of the applications. Here, a text mining tool was developed and its strength was that the same set of procedures for processing and analysing could be applied to both distinct datasets. As a result two comparable sets of data describing both the scientific papers and the grant applications were attained. Based on these dataset similarities among the papers and applications could be calculated. For the mapping of a grant application to a topic the average similarity of the proposal to all the papers in the topic is calculated. If this average similarity exceeds a certain threshold, the application is said to be relevant to the topic. Based on the validation of the data we set the threshold to 0.025. As mentioned earlier, this mapping exercise requires manual input at some points and this decision of the threshold is the most important one. This decision cannot be automated as it depends completely on the required balance between the two types of error. Any computerized method would require some parameter to set this balance. This means that the problem is just shifted from one parameter to another one.

However, we feel confident about the choice of manual input as this decision was supported by extensive validation.

For the validation of the results of the mapping we used a manual, qualitative procedure. The aim of the validation procedure was three-fold:

- Estimation the appropriateness of the developed mapping methodology and adjustment if necessary.
- Decision upon the cut-off threshold
- Removal of the false positive applications

The above mentioned balance between false positives and false negatives was taken into account with a stronger focus on avoiding the latter one. Here, the validation of the first stage of the matching proved to be very helpful, many of the proposal were already read and assigned manually to one or more fields or topics. This allowed for a qualitative judgement of the relevancy of the application in the subject category. While reading the proposals and comparing their content to the matched fields it became clear that the methodology developed for the mapping was appropriate. The first possible sources of error were identified , especially common words with distinct meaning across fields seemed to distort the mapping.

The document-application pairs with the highest similarities have been inspected and common words were extracted. This allowed us to adjust the processing and analysis of the text fragments in the papers and proposals. Another possible error was identified in this validation step: sometimes papers are linked based on terms in the PI's prior achievements. This can easily solved in the future if applications could be submitted to the ERC with distinct sections.

In the last validation step we investigated the mapping of applications to particular topics. After this step the decision on the threshold was finalized and some false-

positives as several papers dealing with HIV or AIDS in a pure medical context were removed.

The results of the mapping showed substantial differences across different topics. In Biomedical research we found only a very small number of applications relevant to the emerging topics and none of them got funded. While 'Fuel cells' and 'Bio diesel' in 'Energy & Fuels' could be linked to a substantial number of proposals of which several were successful. In Environmental Sciences we found a small number of proposals being mapped with the topics but in 'Radiation' four out of five applications got granted while for 'Nano pollution' it was none.

To conclude, this methodology has several advantages as the text based link can be quantified and calculated as soon as the application is submitted. It uses the full body of the text and it does not involve parsing of the text to extract references that have to be matched with individual papers or journals. Of course, any improvement to the body of the text will has a beneficial effect on the mapping.

5.2 IMPACT ON ERC'S OPERATION

Once again the distinction in two perspectives taken in this project finds its mirror image in the dichotomy of the possible impacts of ERACEP on ERC's operation. Both the detection of emerging topics and the mapping of grant applications are conceptually independent from each other, they can serve different purposes within the work flow at ERC. While the emerging topics are more relevant to the internal organization of ERC, the mapping of proposals relates to the main objectives of ERC to support and encourage creative scientists to be adventurous and take risk.

Both methods can also be implemented in a stand-alone environment and have their own usage within ERC. We want to stress here that the outcomes of ERACEP can only support the different work packages within ERC and never serve as a replacement. Within this final report we have mentioned in several paragraphs the importance of human validation and even of human input.

5.2.1 *Immediate consequences of proposal assignment to emerging topics*

One of the most important and immediate impacts is gained through the identification of proposals within emerging research topics. Based on the results of the detection of emerging topics one might expect two main applications of this identification. First, the result can be used to initiate an internal discussion of the structure and scope of the different panels. The description of the existing panels can then refer explicitly to these new topics. Also, the list of featured keywords can be updated. And finally new panels could be created or existing ones revised. Such a restructuring can thus be driven by the inherent dynamics of the science landscape and not by external classifications. *ERC could therefore benefit from this through a better adjustment of the procedures to the evolution of science.*

A further immediate but not obvious aspect of application is that these topics could assist ERC in finding appropriate panel members with the right expertise needed for emerging research directions. Additional bibliometric analysis can be performed on any set of publications, so that it is possible to detect the leading and most appropriate scientists in these emerging topics. In this report we did not elaborate this possibility as this was beyond the scope of the project.

Beeing able to detect emerging clusters implies as well the possibility that topics might evolve in the opposite direction. These vanishing topics can show a strong decrease in activity or they could be completely absorbed by another established one. The detection of clusters that show such dynamics enriches the debate on the structure and scope of the panels. Even the labelling of clusters as being an established research topic can shed another light on science. Often there exists some kind of public consensus about what future technologies or science can be. The present methodology allows us to detect the already ongoing trajectories in these topics and show that they might be not so emergent as generally is thought. As an example we refer to 'nano-science and -techonology' that can already be considered an established field. Within this field, we found, however, that research on pollution of nano-materials is emerging. Analogously, the field of renewable energy was identified as being an established field whereas the research on fuel cells proved emergent within the field.

5.2.2 *Mapping of Applications*

In this subsection we divide the possible usage of the developed methodology into *ex ante* and *ex post* for the evaluation of grant applications.

Ex Ante

- In the present situation, the PI has to decide upon the most suitable panel. The mapping of applications to certain topics could be helpful for the assignment to evaluation panels after submission.
- External experts can more easily be found and assist in the review process. The application can be mapped with a set of papers and bibliometric analysis can then identify the most appropriate researchers.
- Grant applications can be mapped on thematic clusters. These clusters could be labelled as 'emerging', 'established' and even 'vanishing'. Based on this classification, applications can be selected that require special attention.

In this context we have also to mention that similarly to other bibliometric methods, the assignment of proposals to emerging research topics too, is not suited as evaluation tool in the reviewing process of grant applications as the research orientation as such does not reveal any quality aspects. However, the methodology developed in our study can readily be used in conjunction with other bibliometric techniques in the framework of an ex-post evaluation. The expected impact will be outlined in the following.

Ex Post

The results of the mapping of the grant applications combined with the outcomes of the review process can be analysed at an aggregated level and be used during a reflective exercise.

- It might facilitate the identification of differences in procedures amongst panels or across different fields based on coverage of emerging topics and success rates for all topics.
- It can point to topics where the reasons for selection/non-selection may need additional consideration.
- The combined analyses of mapping, the results of the reviewing process and the outcomes achieved by supported grants *after* successful application might give important feedback and help the ERC to adjust or optimise their policies.

5.3 OUTCOMES AND OBJECTIVES

Within the ERACEP support action we developed a robust methodology for the detection of emerging research topics based on advanced bibliometric tools and expert validation. The key elements of the ERACEP approach comprise as a first step an analysis of dynamic fields within the landscape of science. As a second step a cluster analysis within science fields is carried out using a hybrid approach combining bibliographic coupling and text mining. This is complemented by a third step, a diachronic investigation of the links between clusters over time. Finally a qualitative validation by content analysis and expert assessment is performed. This approach was elaborated in an exemplary but in-depth way focussing on a limited number of scientific fields which however represent a broad diversity of science, social sciences, and arts and humanities.

A main advantage of the ERACEP approach for the detection of emerging topics is its open design. It allows open searches in any field of science not restricted to any pre-classification. This feature reflects well the open bottom-up approach of ERC's operation. The ERACEP approach can be applied timely as soon as publications are available. It does not depend on any citation-time-gaps which are typical for citation based bibliometric methods. Finally the approach is generic by nature; it can be applied to any field of scientific activities provided that these are represented in documents which are stored in databases. One example of an additional application field would be the FET OPEN funding activities of the EC which had been developed and tested as an open bottom-up funding scheme in the ICT domain but will be expanded to other science domains.

However, we also note that the applicability of the ERACEP tool box requires a certain expertise in doing bibliometric analyses and the ability to handle and process large amounts of data, which in turn relies on suitable databases and the availability of sufficient computing power and infrastructures. Accordingly a simple transfer of the ERACEP methods is not feasible. On the other hand it can be applied easily in a

cooperative way between external specialized expertise and internal competences of a funding organization.

Despite theses limitations, the overall outcomes of ERACEP indicate that the first main objective, the identification and characterization of emerging topics within the landscape of science could be achieved.

ERACEP research addressing the second main objective, the analysis to what extent ERC funding is able to cover and support emerging topics turned out to be rather demanding. As a first challenge we identified data availability. The original intention to base this part of ERACEP on published information about ERC funding activities turned out not to be feasible since such information was not detailed enough. Rather we had to use detailed content information on the level of individual grant applications. This also implied that the scope of the ERACEP analysis had to be narrowed down to a limited number of grant applications. We used the 932 applications to the 2009 Starting Grant Call.

A full text matching approach was developed to map these applications to the dynamic fields and emerging topics of science. The complete text of each application was transferred into a text body which could be compared with the corresponding text bodies derived from all publications characterizing the selected fields of science including emerging topics. For the selected sample of grant applications we demonstrate that ERC funding procedures indeed are suitable to address emerging topics. However we also observe substantial differences between dynamic fields and emerging topics.

This raises a number of interesting questions for further in-depth analyses: Why do scientific communities from different scientific fields consider emerging topics differently in their research activities? What are the reasons for the differences in success rates of grant applications in the different science domains and in particular considering emerging topics? More generally, what are the main motivations of scientists to select their preferred topics of research? How important are factors like curiosity, relevance, impact, or likelihood of improving publication records and obtaining funding support in such considerations?

Taken together ERACEP outcomes demonstrate that also its second main objective, the analysis to what extent ERC funding is able to cover and support emerging topics, could be achieved.

ERACEP outcomes are clearly relevant to the issue of putting ERC-funded research into a systemic context of worldwide research activities in emerging topics, which was identified as a first possible impact dimension of ERACEP. ERC-funded research is mapped to the landscape of science which could also be characterized by its international collaboration networks and specializations. ERACEP outcomes also indicate that the open call approach of the ERC is able to attract applications which are addressing emerging topics. Finally, ERACEP outcomes are useful for supporting and critically reflecting ERC's funding procedures from an *ex ante* and an *ex post* perspective. Ex ante preselection of grant applications, defining evaluation panels, and attributing applications to panels are facilitated by mechanisms which reflect the

inherent structure and dynamics of science. From the ex post perspective, the ERACEP matching approach can be used for supporting a critical reflection of ERC-procedures.

Annex

LIST OF CONSULTED EXPERTS

Allebeck, Peter – Department of Public Health Sciences, Karolinska Institute (SE)

Bradke, Harald – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Bührlen, Bernhard – Psychiatric University Clinics Basel (CH)

Delnoij, Diana – Faculty of Social and Behavioural Sciences, Tilburg University (NL)

Deprest, Jan – Faculty of Medicine, University of Leuven (BE)

Falzetti, Marco – The European Technology Platform on Advanced Engineering Materials and Technologies (IT)

Gowen, Margaret – Institute of Archaeologists of Ireland (IRL)

Hägele, Martin – Fraunhofer Institute for Manufacturing Engineering and Automation IPA (DE)

Heinze, Cornelia – Department of Nursing, Science, Charite´ University Berlin (DE)

Hüsing, Bärbel – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Keren, Gideon – Faculty of Social and Behavioural Sciences, Tilburg University (NL)

Langanke, Martin – University of Greifswald, Faculty of Theology (DE)

Marscheider-Weidemann, Frank – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Musso, Stefano – Faculty of Architecture and the Faculty of Engineering, University of Genoa (IT)

Neuper, Christa – Department of Psychology, University of Graz (AT)

Schade, Wolfgang – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Schröter, Marcus – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Stahlecker, Thomas – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Strech, Daniel – Hanover Medical School, Institute for History, Ethics and Philosophy of Medi-cine (DE)

Walz, Rainer – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Wietschel, Martin – Fraunhofer Institute for Systems and Innovation Research ISI (DE)

Bibliography

Batagelj, V., Mrvar, A. (2003). *Pajek – analysis and visualization of large networks*. In M. Jünger & P. Mutzel (Eds.), Graph drawing software. Springer, 77–103.

Bell, M.L., Samet, J.M., Dominici, F. (2004), Time-series of particulate matter. *Annu Rev Public Health*, 25 (2004), 247–280.

Benoit, B. (2003), The value of three-dimensional ultrasonography in the screening of the fetal skeleton. *Childs Nerv Syst*, 19, 403–409.

Birbaumer, N. (2006), Brain-computer-interface research, Coming of age. *Clinical Neurophysiology*, 117, 479–483.

Boschma R., Frenken, K. (2011), The emerging empirics of evolutionary economic geography. *J Econ Geogr*, 11(2), 295–307.

Boyack, K.W., Klavans, R. (2010), Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *JASIST*, 61 (12), 2389–2404.

Brakman, S., Garretsen, H., van Marrewijk, C. (2001), *An Introduction to Geographical Economics, Trade, Location and Growth*. Cambridge, Cambridge University Press.

Braam, R.R., Moed, H.F., van Raan,A.F.J. (1991a). Mapping of science by combined cocitation and word analysis, Part 1: Structural aspects. *Journal of the American Society for Information Science*, 42 (4), 233–251.

Braam, R.R., Moed, H.F., van Raan, A.F.J. (1991b). Mapping of science by combined cocitation and word analysis, Part II: Dynamical aspects. *Journal of the American Society for Information Science*, 42 (4), 252–266.

Brunekreef, B., Holgate, S.T. (2002), Air pollution and health. *Lancet*, 360(2002), 1233–242.

Costello, A., Abbas, M., Allen, A., Ball, S. et al. (2009), Managing the health effects of climate change. *Lancet*, 373, 1693–1733.

Cuhls, K., Grupp, H. (1998), *Studie zur globalen Entwicklung von Wissenschaft und Technik*. BMBF, Bonn.

Daughton, C.G. (2004), Non–regulated water contaminants, emerging research. *Environ Impact Asses Rev*, 24, 711–32.

DeVore, G.R., Falkensammer, P., Sklansky, M.S., Platt, L.D. (2003), Spatio–temporal image correlation (STIC), new technology for evaluation of the fetal heart. *Ultrasound Obstet Gynecol*, 22, 380–387.

DOE-EERE (2008), Fuel cell technology challenges. http,//www1.eere.energy.gov/hydrogenandfuelcells/fuelcells/fc_challenges.html

Down, L.J. (1866), Observations on an ethnic classification of idiots. *Clinical Lectures and Reports, London Hospital*, 3, 259–262.

ERC (2009), *Work Programme*. ftp://ftp.cordis.europa.eu/bup/fp7/docs/wp/ideas/l_wp_200901_en.pdf

Fano, R.M. (1956), *Information theory and the retrieval of recorded information*. In J. H. Shera, A. Kent, J.W. Perry (Eds.), Documentation in action. New York: Reinhold Publ. Co., 238–244.

Fischer, B.R. (2001), *Performance Analyse in der Praxis. Performanzmaße, Attributionsanalyse, VFA-Performance Presentation Standards*. München/Wien, R. Oldenbourg Verlag.

Fischer et al. (2010), Trends in relative risk estimates for the association between air pollution and mortality in The Netherlands, 1992–2006. *Environmental Research*, 111 (1), 94–100.

Frenken, K., Boschma, R. (2007), A theoretical framework for evolutionary economic geography, industrial dynamics and urban growth as a branching process. *Journal of Economic Geography*, 7, 635–649.

Garche, J., Bonhoff, K., Ehret O., Willmetz, W. (2009), The German National Innovation Programme Hydrogen and Fuel Cell Technology. *Journal Fuel Cells*, 3 (2), 192–196.

Garfield, E. (1969), Permuterm Subject Index – the primordial dictionary of science. *Current Contents*, 12 (22), 4.

Geller, G., Strauss, M., Bernhard, B.A., Holtzman, N.A. (1997), Decoding informed consent, Insights from women regarding cancer susceptibility testing. *Hastings Center Report*, 27(2), 28–33.

Gittleman, C., Masten, D., Scott, J. et al. (2010), *Automotive fuel cell R&D needs. DOE fuel cell pre-solicitation workshop*. Department of Energy, Lakewood, Colorado.

Glänzel, W., Czerwon, H.J. (1996), A New Methodological Approach to Bibliographic Coupling and Its Application to the National, Regional and Institutional Level. *Scientometrics*, 37 (2), 195–221.

Glänzel, W., Thijs, B. (2011a), Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88 (1), 297–309.

Glänzel, W., Thijs, B. (2011b), Research in Emerging Fields: Who takes the lead? *ISSI Newsletter*, 7 (4), 85–95.

Glänzel, W., Thijs, B. (2012), Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics*, 91 (2), 399–416.

Glenisson, P., Glänzel, W., Janssens, F., de Moor, B. (2005), Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management*, 41 (6), 1548–1572.

Golden, J.S., Hartz, D., Brazel, A. et al. (2008), A biometeorology study of climate and heat-related morbidity in Phoenix from 2001 to 2006. International *Journal of Biometeorology*, 52 (6), 471–480.

Goncalves, L.F., Lee, W., Espinoza, J., Romero, R. (2005), Three- and 4-dimensional ultrasound in obstetric practice, does it help? *J Ultrasound Med*, 24, 1599–1624.

Haines, A., Kovats, R.S., Campbell–Lendrum, D., Corvalan C. (2006), Climate change and human health, impacts, vulnerability, and mitigation. *Lancet*, 367 (2006), 2101–2109.

Grupp, H. (1993), *Technologie am Beginn des 21. Jahrhunderts*, Physica, Heidelberg.

Hicks, D. (1987), Limitations of co-citation analysis as a tool for science policy. *Social Studies of Science*, 17, 295–316.

High-Level Expert Group (2005), *Frontier Research: The European Challenge*. Luxembourg, Office for Official Publications of the European Communities. ISBN 92-894-9209-0.

IPCC, Intergovernmental Panel on Climate Change (2007), *The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge, University Press.

Iversen, H.I. et al. (2008), A brain-computer interface tool to assess cognitive functions in completely paralyzed patients with a myotrophic lateral sclerosis. *Clinical Neurophysiology*, 19(10), 2214–2223.

Jo, Y., Lagoze, C., Giles, CL (2007), *Detecting research topics via the correlation between graphs and texts*. KDD-2007 Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 370–379.

Karlessi, T., Santamouris, M., Apostolakis, K. et al. (2009), Development and testing of thermochromic coatings for buildings and urban structures. *Solar Energy*, 83 (4) (2009), 538–551.

Kessler, M.M. (1963), Bibliographic coupling between scientific papers. *American Documentation*, 14, 10–25.

Kostoff, R.N., Eberhart, H.J., Toothman, D.R. (1997), Database tomography for information retrieval. *Journal of Information Science*, 23 (4), 301–311.

Krugman, P. (1991), Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483–499.

Krugman, P. (2011), The New Economic Geography, now middle-aged. *Regional Studies*, 45(1), 1–7.

Lakerveld, J., Kotchoubey, B., Kubler, A. (2008), Cognitive function in patients with late stage amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*, 79, 25–29.

Lamirel, J.C., Ta A.P., Attik, M. (2008), Novel labeling strategies for hierarchical representation of multidimensional data analysis results. In: A. Gammerman (Ed.), Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, 11-13 February 2008, Innsbruck, Austria. ACTA Press, Track 595-138, p. 169–174.

Lamirel, J.C., Safi, Gh., Pryankar, N., Cuxac, P. (2010), *Mining research topics evolving over time using a diachronic multi-source approach*. The Fourth International Workshop on Mining Multiple Information Sources - ICDM 2010.

Lee, W., Kalache, K.D., Chaiworapongsa, T. et al. (2003), Three–dimensional power Doppler ultrasonography during pregnancy. *J Ultrasound Med*, 22, 91–97.

Leydesdorff, L. (2006), Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, 57 (5), 601–/22613.

Lim, S., Teong, L.K. (2010), Recent trends, opportunities and challenges of biodiesel in Malaysia, an overview. *Renew Sustain Energy Rev*, 14, 938–54.

Lin, L., Dong, Y., Chaitep, S., Vittayapadung, S. (2009), Biodiesel production from crude rice bran oil and properties as fuel. *Appl Energy*, 86, 681–688.

Lin, L., Cunshan, Z., Vittayapadung, S. et al. (2011): Opportunities and challenges for biodiesel fuel. *Applied Energy*, 88, 1020–1031.

MacKinnon, D., Cumbers, A., Pike, A., et al. (2009), Evolution in Economic Geography, Institutions, Political Economy, and Adaptation. *Economic Geography*, 85, 129–150.

Marshakova, I.V. (1973), System of connections between documents based on references (as the Science Citation Index).

Meng, X., Yang, J., Xu, X. et al. (2009), Biodiesel production from oleaginous microorganism. *Renew Energy*. 34, 1–5.

Millán, J. et al. (2010), Combining Brain-Computer Interfaces and Assistive Technologies, State-of-the-Art and Challenges. *Frontiers in Neuroscience*, 4.

Moore, M.N. (2006), Do nanoparticles present ecotoxicological risks for the health of the aquatic environment? *Environment International*, 32, 967–76.

Na, K., Lee, T.B., Park, K.H., Shin, E.K., Lee, Y.B., Choi, H.K. (2003), Self–assembled nanoparticles of hydrophobically modified poly-saccharide bearing vitamin H as a targeted anti-cancer drug delivery system. *Eur J Pharm Sci*, 18, 165–73.

Oberdörster, E. (2004), Manufactured nanomaterials (fullerenes, C60) induce oxidative stress in the brain of juvenile largemouth bass. *Environ Health Perspect*, 112, 1058–62.

Pantavou, K., Theoharatos, G., Mavrakis, A. et al. (2011), Evaluating thermal comfort conditions and health responses during an extremely hot summer in Athens. *Building and Environment*, 46 (2), 339–344.

Papageorgopoulos (2010), *DOE fuel cell technology program overview and introduction to the 2010 fuel cell pre-solicitation workshop in DOE fuel cell pre-solicitation workshop*. Department of Energy, Lakewood, Colorado

Puza, B., Roberts, S., Yang, M. (2011), Constrained confidence intervals in time series studies of mortality and air pollution. *Environment International*, 37 (1), 204–209

Reiss, T., Schmoch, U., Schubert, T. et al. (2007), *Aussichtsreiche Zukunftsfelder der Biotechnologie – Neue Ansätze der Technologievorausschau*. Stuttgart: Fraunho-fer IRB Verlag (ISI-Schriftenreihe Innovationspotenziale)

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Ruano, R. , Martinovic, J., Dommergues, M. et al. (2005), Accuracy of fetal lung volume assessed by three-dimensional sonography. *Ultrasound Obstet Gynecol*, 26, 725–730.

Samet, J.M. (2002), Air pollution and epidemiology, "déjà vu all over again?" *Epidemiology*, 13 (2) (2002), 118–119.

Samoli, E., Peng, R., Ramsay, T. et al. (2008), Acute effects of ambient particulate matter on mortality in Europe and North America, results from the APHENA study. *Environ Health Perspect*, 116 (2008), 1480–1486.

Sayes, C.M., Fortner, J.D., Guo, W. et al. (2004), The differential cytotoxicity of water soluble fullerenes. *Nanoletters*, 4, 1881–7.

Schmoch, U., Wang, J., Stoica, R. (2005), *Research and Development in the Turkish and North African Landscape of Science*. Report to the German Federal Ministry of Education and Research.

Schwartz, J., Samet, J., Patz, J. (2004), A hospital admissions for heart disease, the effects of temperature and humidity. *Epidemiology*, 15 (6) (2004), 755–761.

Sen, S.K., Gan, S.K. (1983), A mathematical extension of the idea of bibliographic coupling and its applications. Annals of Library Science and Documentation, 30, 78–82.

Sharpe, W.F. (1966), Mutual Fund Performance. *Journal of Business*, 39 (1), 119-138.

Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2008), Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* 28 (11), 758–775.

Shin, H.H., Stieb, D.M., Jessiman, B. et al. (2008), A temporal, multicity model to estimate the effects of short-term exposure to ambient air pollution on health. *Environ Health Perspect*, 116, 1147–1153.

Shu, Q., Yang, B., Yuan, H. et al. (2007), Synthesis of biodiesel from soybean oil and methanol catalyzed by zeolite beta modified with La3+. *Catal Commun*, 8(12), 2159–65.

Sims, R., Taylor, M., Saddler, J., Mobee, W. (2008), *From 1st- to 2nd-generation biofuel technologies*. IEA 2008

Small, H. (1973), Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265–269.

Smith, R.A. (1992), *Handbook of amyotrophic lateral sclerosis*. New York, Marcel Dekker Inc.

Song, C. (2002), Fuel processing for low-temperature and high-temperature fuel cells, Challenges, and opportunities for sustainable development in the 21st century. C*atalysis Today*, 77, 17–49.

Srivastava, A., Prasad, R. (2000), Triglycerides-based diesel fuels. *Renew Sustain Energy Rev*, 4, 111–33.

Takeda, Y., Kajikawa, Y. (2008), Optics: a bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*,78 (3), 543–558.

Todorov, R. (1992), Displaying content of scientific journals: A co-heading analysis. *Scientometrics*, 23 (2), 319–334.

Tonkin, I. (2009), *Global biodiesel production reaches record high*. http,//www.roadtransport.com/

Thomson Reuters, (2011a), *Science Citation Index Expanded*, *Scope Notes*, accessible via http://science.thomsonreuters.com/mjl/scope/scope_scie/.

Thomson Reuters, (2011b), *Social Science Citation Index*, *Scope Notes*, accessible via http://science.thomsonreuters.com/mjl/scope/scope_ssci/.

Turner, W., Chartron, G., Laville, F., Michelet, B. (1988), *Packaging information for peer review: new co-word analysis techniques*. In: van Raan, A., (Ed.), Handbook of quantitative studies of science and technology. North Holland, 291–323.

Wang, Y. et al. (2011), A review of polymer electrolyte membrane fuel cells, Technology, applications, and needs on fundamental research. *Journal Applied Energy*, (88), 981–1007.

WHO, World Health Organization (2003), *Heat-waves, impacts and responses*. Briefing note for the fifty–third session of the WHO Regional Committee for Europe, Vienna, Austria.

Xu, H.X., Zhang, Q.P., Lu, M.D., Xiao, X.T. (2002), Comparison of two-dimensional and three–dimensional sonography in evaluating fetal malformations. *J Clin Ultrasound*, 30, 515–525.

Zhang, L., Glänzel, W., Liang, L. (2009), Tracing the role of individual journals in a cross-citation network based on different indicators. *Scientometrics*, 81 (3), 821–838.

Zitt, M., Bassecoulard, E. (1994), Development of a method for detection and trend analysis of research fronts built by lexical or cocitation analysis. *Scientometrics*, 30 (1), 333–351.