

Unclassified

DSTI/DOC(2007)1

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

23-Feb-2007

English text only

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY

Cancels & replaces the same document of 20 February 2007

**CAPTURING THE EVOLVING NATURE OF SCIENCE, THE DEVELOPMENT OF NEW
SCIENTIFIC INDICATORS AND THE MAPPING OF SCIENCE**

STI WORKING PAPER 2007/1

Statistical Analysis of Science, Technology and Industry

Masatsura Igami and Ayaka Saka

JT03222529

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

DSTI/DOC(2007)1
Unclassified

English text only

STI Working Paper Series

The Working Paper series of the OECD Directorate for Science, Technology and Industry is designed to make available to a wider readership selected studies prepared by staff in the Directorate or by outside consultants working on OECD projects. The papers included in the series cover a broad range of issues, of both a technical and policy-analytical nature, in the areas of work of the DSTI. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comments on the papers are invited, and should be sent to the Directorate for Science, Technology and Industry, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

<http://www.oecd.org/sti/working-papers>

CAPTURING THE EVOLVING NATURE OF SCIENCE, THE DEVELOPMENT OF NEW SCIENTIFIC INDICATORS AND THE MAPPING OF SCIENCE*

Masatsura IGAMI, OECD (masatsura.igami@oecd.org)

and

Ayaka SAKA, National Institute of Science and Technology Policy, Japan (saka@nistep.go.jp)

ABSTRACT

There is a long history describing the structure and evolution of science. Recent unprecedented progress in access, use, and analysis of information on scientific publications and patents open innovative ways to study the structure and evolution of science. Especially, a mapping of knowledge has received wide recognition as a new, evolving area of research.

The ultimate goal of this study is to contribute to endeavours to understand and track the changing nature of science. In this study, current trends in scientific activities were mapped and their characteristics were examined. Research areas were explored through a co-citation analysis and a map of science was generated to analyse how research areas were related to each other. Methodology which is commonly used in social network analysis was also applied to examine knowledge networks at the institutional level.

The analysis clearly shows the multi-disciplinary character of some research, such as “Nano materials and devices”, “Genomics”, and “Environment”. A precursor of the emergence of nano-bioscience is also observed. Measurement of countries’ specialisation clearly indicates an increased share of the BRICs (Brazil, Russia, India, and China) in some research. The BRICs are currently in a stage of intense catching up and their importance in knowledge networks is likely to become substantial. Social network analysis at the institutional level reveals that the structure of knowledge networks strongly depends on research and each institution probably has its own function in the network. These results show how science is evolving not only across disciplines but also across countries or regions.

*

This report documents results from collaborative research between OECD and the National Institute of Science and Technology Policy (NISTEP) in Japan. This research would have not been possible without the support and contribution of the NISTEP. The NISTEP provided the expertise in identifying the research areas on which the analysis is based. The paper was prepared under the supervision of Alessandra Colecchia, DSTI, and benefited from valuable comments by Anthony Arundel, DSTI, and Terutaka Kuwahara, NISTEP.

MISE EN ÉVIDENCE DU CARACTÈRE ÉVOLUTIF DE LA SCIENCE, ÉLABORATION DE NOUVEAUX INDICATEURS SCIENTIFIQUES ET TYPOLOGIE DE LA SCIENCE*

Masatsura IGAMI, OCDE (masatsura.igami@oecd.org)

et

Ayaka SAKA, Institut national japonais de la politique scientifique et technologique (saka@nistep.go.jp)

RESUMÉ

Il existe une longue tradition de description de la structure et de l'évolution de la science. Cependant, les progrès récents sans précédent dans l'accès à l'information sur les publications et brevets scientifiques et dans l'utilisation et l'analyse de cette information ouvrent des voies nouvelles pour étudier la structure et l'évolution de la science. En particulier la typologie du savoir est de plus en plus reconnue comme un nouveau domaine de recherche prometteur.

Le but ultime de cette étude est de contribuer aux efforts pour comprendre et retracer le caractère évolutif de la science. Dans cette étude, les tendances actuelles des activités scientifiques ont été mises en évidence et leurs caractéristiques analysées. Les domaines de recherche ont été explorés au moyen d'une analyse de co-citations et une typologie de la science a été dressée pour analyser les liens existant entre les différents domaines scientifiques. Une méthodologie couramment employée dans l'analyse des réseaux sociaux a également été utilisée pour examiner les réseaux de connaissance au niveau institutionnel.

L'analyse a clairement montré le caractère pluridisciplinaire de certaines recherches, comme les « nanomatériaux et nanodispositifs », la « génomique » et l'« environnement ». Un précurseur de l'émergence de la nanobioscience a également été observé. La mesure de la spécialisation des pays a clairement indiqué l'émergence des BRICs (Brésil, Russie, Inde et Chine) dans certaines recherches. Les BRICs sont actuellement engagés dans une phase intense de rattrapage et ils vont certainement prendre une importance significative dans les réseaux de connaissance. L'analyse des réseaux sociaux au niveau des institutions a montré que la structure des réseaux de connaissance est fortement conditionnée par la recherche et que chaque institution occupe sans doute une fonction propre dans le réseau. Ces résultats ont mis en évidence la façon dont la science évolue non seulement entre les disciplines, mais aussi entre les pays et les régions.

*

Ce rapport rend compte des résultats obtenus dans le cadre d'une recherche en collaboration entre l'OCDE et l'Institut national japonais de la politique scientifique et technologique (NISTEP). Cette recherche n'aurait pas été possible sans le soutien et la collaboration du NISTEP. Le NISTEP a apporté son expertise dans l'identification des domaines de recherche sur lesquels repose l'analyse. Le document a été préparé sous la supervision d'Alessandra Colecchia, de la DSTI, et il a bénéficié des précieux commentaires d'Anthony Arundel de la DSTI et de Terutaka Kuwahara du NISTEP.

TABLE OF CONTENTS

1. INTRODUCTION	6
2. METHODOLOGY	8
3. MAPPING OF CURRENT SCIENTIFIC ACTIVITIES	12
4. MEASUREMENT OF MULTI-DISCIPLINARY CHARACTER OF RESEARCH	18
5. MEASUREMENT OF COUNTRIES' SHARE	24
6. CHARACTERISTICS OF INTERNATIONAL RESEARCH IN CITING PAPERS	33
7. CHARACTERISTICS OF INSTITUTIONAL DIVERSITY	38
8. CONCLUSION AND DISCUSSION.....	44
APPENDIX A. ANALYSIS OF CORE PAPERS.....	48
BIBLIOGRAPHY.....	51

CAPTURING THE EVOLVING NATURE OF SCIENCE, THE DEVELOPMENT OF NEW SCIENTIFIC INDICATORS AND THE MAPPING OF SCIENCE

1. Introduction

Bibliometric analysis is a generic tool for measuring international scientific activities. There are many country-level analyses of trends in scientific publications. These analyses constitute effective tools for understanding the macro-scale character of scientific activities (OECD, 2005; NISTEP 2005b). One limitation of these macro-scale analyses is that it is difficult to capture the qualitative evolution of scientific activities through them. A representative case is the development of quantum physics; understanding electrons' behaviour stimulated wide debate in the early 20th century, but scientists now try to control behaviour of electrons on a nano-scale. Similar kinds of remarkable changes have taken place in bioscience since the discovery of the double helix structure of DNA (Watson and Crick, 1953). These dynamics within research are difficult to observe via macro-scale analysis. In this respect, the meso-scale analysis of science activities, *i.e.* the analysis of "research areas", is likely to be a candidate for grasping qualitative changes in scientific activities.

Relationships between types of scientific research are also changing. For example, inter-disciplinary and multi-disciplinary research such as nanoscience and bioinformatics have received wide recognition in recent years. In spite of increased awareness, there have been no methodological guidelines to distinguish the inter-disciplinary or multi-disciplinary character of these types of research. Traditionally, these types of research are simply named as combinations of existing research, such as nano-biology, bio-electronics and so on. The establishment of new systematic methodologies to measure the inter-disciplinary and multi-disciplinary character of science should contribute to monitoring scientific activities.

There is a long history describing the structure and evolution of science (Garfield *et al.*, 1964; Small and Sweeney, 1985a; Small *et al.*, 1985b). Recent unprecedented progress in access to, use and analysis of information on scientific publications and patents has opened innovative ways to study the structure and evolution of science. Especially, mapping of knowledge has received wide recognition as a newly evolving area of research. These endeavours aim to develop new tools for easing information access, exploring the structure of knowledge, and assessing the evolution of knowledge (Börner *et al.*, 2003 and references therein; Boyack, 2004 and references therein; Chen, 1999; Chen and Paul, 2001; Shiffrin and Börner, 2004; Mane, K. K. and Börner, 2004).

The ultimate goal of this study is to contribute to the endeavours aimed at understanding and tracking the changing nature of science. Results of the analysis address questions like:

- How are new scientific areas emerging and developing?
- How are scientific areas linked to each other?
- How can the multi-disciplinary character of scientific areas be assessed?
- How can international knowledge flows be measured both at the national and institutional level?

In this study, the citation behaviour of scientists is used as a proxy to measure various information exchanges which translate into knowledge creation and flows. Clusters of papers having similar research subjects are retrieved via co-citation analysis.

A total of 47 218 highly cited papers, *i.e.* the top 1% cited papers, from 1999 to 2004, were clustered through co-citation analysis, so as to obtain a total of 133 research areas, *i.e.* clusters of highly cited papers related to similar research. The Essential Science Indicators (ESI) database, a commercial database provided by Thomson Scientific Inc., was used as the information source for our analysis. Unique names were given to the research areas through a keyword analysis and a consultation with over 100 experts in university, public, and private institutions in Japan.

One remarkable feature of this analysis is that it does not require keywords in identification of research areas. Interaction among researchers, which is observed as a form of citations, is a unique element which is responsible for the formation of research areas. Research areas emerged spontaneously through interactions such as the self-organisation process. This methodological feature makes a holistic understanding on scientific activities beyond borders lying among traditional disciplines possible. Research areas obtained are further evaluated by creation of a map of science. Detailed analysis on statistical characteristics of papers reveals the nature of research. Elaboration of citations will reveal which research areas show rapid development in recent years.

Methodology commonly used in social network analysis is also applied to examine knowledge networks at the institutional level. Study on social networks has been being conducted since the middle of 1930s and many indicators have been developed (Wasserman and Faust, K., 1994; Carrington *et al.*, 2005). Their application to the World Wide Web (Albert *et al.*, 2000), for example, enables one to understand the huge network structure in cyber space. PageRank (Page *et al.*, 1998) is a notable example. Since the same methodology can be applied to co-author networks in science activities, application of the social network analysis to bibliometric analysis could assist the development of new types of science and technology indicators.

This report documents results derived from collaborative research between OECD and the National Institute of Science and Technology Policy (NISTEP) in Japan. Research areas, a fundamental dataset of this research, were prepared by NISTEP. It includes identification of 133 research areas through co-citation analysis, mapping of individual research areas, and content analysis of individual research areas. OECD pursued the development of science, technology, and innovation (STI) indicators in close collaboration with NISTEP. Major achievements are the development of a map showing a holistic view of current scientific activities, identification of 14 research categories based on 133 research areas, assessment of multi-disciplinary character and countries' specialisation in the categories, and application of social network analysis to co-authorship among institutions. This report especially focuses on endeavours to develop new STI indicators among various achievements derived from collaborative research. Full results of the identification of research areas and the content analysis on individual research areas are documented in a report published by NISTEP (NISTEP, 2007).

This report is organised as follows: Section II explains briefly the methodology; Section III shows the results of mapping 133 research areas and introduces 14 research categories which constitute large domains in modern science; Section IV derives indicators to assess the multi-disciplinary character of these categories; Section V looks at countries' relative advantage in each category; Section VI provides measures of international knowledge flows via analysis of international co-authorship; Section VII shows knowledge networks and their diversity at institutional level; Section VIII concludes and suggests some future work. Finally, Annex A provides fundamental information on the database used.

2. Methodology

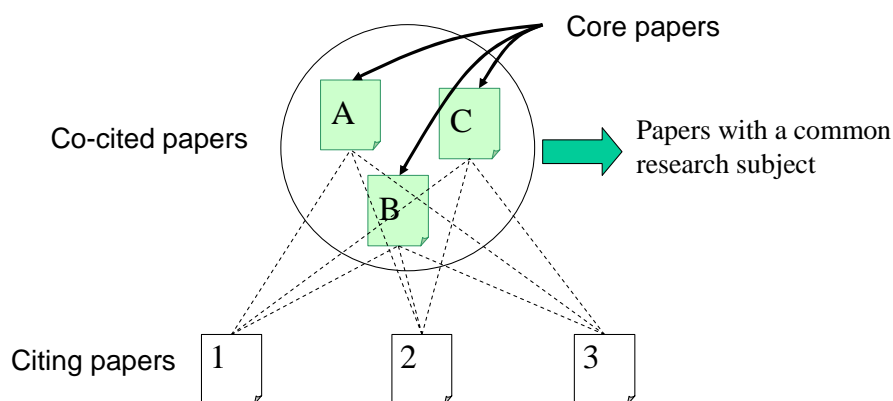
2.1 Basic concept of co-citation analysis

Knowledge creation and flows in cutting-edge research are transmitted through information exchange among researchers. The exchange can take various forms. Citation of scientific papers is one of the major ways for knowledge to flow. Analysis of citations and identification of core papers, the papers which play a central role in research areas, enables examination of characteristics of the research areas and relations among them.

In this research, clusters of papers having similar research subjects were identified via co-citation analysis (Small and Sweeney 1985a). Co-citation is a form of citation in which a set of papers is simultaneously cited by other papers. Figure 1 shows schematic representation of the co-citation, where papers A, B and C are co-cited papers and papers 1, 2, and 3 are citing papers. If papers A and B are co-cited frequently, one can speculate that they have a common research subject. Thus, successive clustering of papers yields a cluster of papers, *i.e.* papers A, B and C in Figure 1, related to a common research subject.

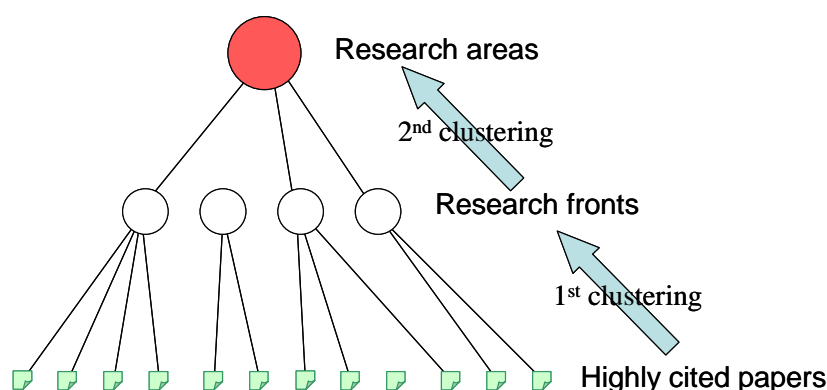
Research areas were identified through two steps to cluster highly cited papers. Clusters of papers obtained through the first and second clustering are referred to as “research fronts” and “research areas”, respectively. Highly cited papers aggregated by co-citation are referred to as “core papers”. The relation between core papers, research front, and research area is shown in Figure 2.

Figure 1. Identification of research areas via co-citation analysis



Note: Schematic figure of co-citation relationships (the dotted lines indicate citations). In the example above, Papers A and B are cited as a pair (co-cited) by papers 1, 2, and 3 at the same time. A cluster of co-cited papers usually has a common research subject. In this research, a cluster of co-cited papers is referred to as “core papers”.

Figure 2. Concept of highly cited papers, research front



2.2 Basic dataset for clustering

A basic dataset for clustering is highly cited papers, *i.e.* the top 1% highly cited papers. They were selected from each field and each year to avoid a bias of fields. Journals are classified into 22 fields¹ in the ESI database. A discipline of a highly cited paper is defined based on the field of the journal. Condition for being selected as highly cited papers vary considerably between fields, for example bioscience papers need about ten times more citations than mathematics papers in order to be highly cited papers. This study covers the period from 1999 to 2004. There were 47 218 highly cited papers in this time period.

A dataset of core papers and research fronts was retrieved from the ESI database. A dataset of citing papers was obtained from Science Citation Index on CD-ROM [1999-2004]. Both datasets are commercial databases provided by Thomson Scientific Inc. The analysis of citing papers was conducted by NISTEP.

2.3 Clustering of papers through co-citations

(Parameters for the clustering)

Clusters of highly cited papers were identified using a single-link clustering method (Small *et al.*, 1985b, see Box 1). There are three parameters which characterise the clustering: *i*) threshold of the clustering, *ii*) maximum cluster size, and *iii*) minimum cluster size. The threshold determines the strength of linkages within the clusters. The maximum and minimum cluster sizes limit the size of clusters. It should be mentioned that the setting of the threshold, maximum cluster size, and minimum cluster size is entirely subjective. The number of research areas strongly depends on the settings. Generally speaking, it is hard to say what parameters are the most adequate for the clustering, because there is no way to measure and define the conceptual expanse of actual research areas. In this study, the parameters were determined so as to obtain a reasonable number of research areas for the analysis; especially for obtaining stable clusters for longitudinal analysis and for obtaining a feasible number of research areas for content analysis (see section 2.5).

1. They are *i*) Agricultural Sciences, *ii*) Mathematics, *iii*) Biology & Biochemistry, *iv*) Microbiology, *v*) Chemistry, *vi*) Molecular Biology & Genetics, *vii*) Clinical Medicine, *viii*) Multidisciplinary, *ix*) Computer Science, *x*) Neuroscience & Behaviour, *xi*) Economics & Business, *xii*) Pharmacology & Toxicology, *xiii*) Engineering, *xiv*) Physics, *xv*) Environment/Ecology, *xvi*) Plant & Animal Science, *xvii*) Geosciences, *xviii*) Psychiatry/Psychology, *xix*) Immunology, *xx*) Social Sciences, general, *xxi*) Materials Science, and *xxii*) Space Science.

(First step of clustering)

Co-citation frequencies for all pairs of the highly cited papers were counted and normalised. There were approximately one million co-citation relations among highly cited papers. Among the various indicators for measuring similarity between a pair of papers (Wasserman and Faust, K., 1994), normalised co-citation frequencies, Pearson product-moment correlation coefficients were calculated using the following equation:

$$N_{ij}^{\text{norm}} = \frac{n_{ij}}{\sqrt{n_i} \sqrt{n_j}}$$

where n_i (n_j) is the citation frequency of a paper i (j), and n_{ij} means the co-citation frequency. The normalised co-citation frequency can be regarded as the strength of the linkage between the papers i and j .

The threshold of the normalised co-citation in the first step was set as 0.3. This means that the linkages whose normalised co-citation was larger than 0.3 were only taken into account during the clustering. The maximum cluster and minimum cluster size were 50 and 2, respectively. If the number of core papers in a cluster exceeded 50, the threshold was increased until the huge cluster was divided into smaller ones including less than 50 core papers. This method is known as variable level clustering (Small and Sweeney 1985).

The results of the first step of clustering were retrieved from the ESI database. The database includes about 5 350 clusters, *i.e.* research fronts. They contain a total of 21 411 core papers which correspond to about half of total highly cited papers. In other words, information on half of the highly cited papers was not included.

Usually the number of research fronts speculated from the core papers tends to be smaller than research areas. For example, there are about 60 research fronts associated with carbon nanotubes. Each research front represents specific research on carbon nanotubes such as synthesis of carbon nanotubes, application of carbon nanotubes to field effect transistors, hydrogen storage in carbon nanotubes, etc. This means that a research front corresponds to a research topic rather than a research area. Thus research fronts were further aggregated in order to obtain their clusters, *i.e.* research areas.

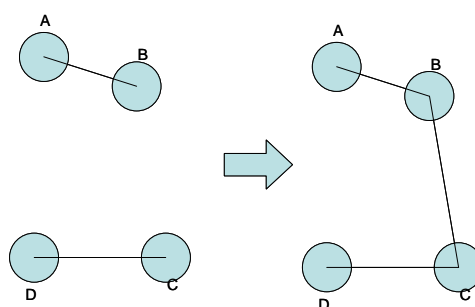
(Second step of clustering)

The second step of clustering is almost identical to the first step. One major difference is that an individual research front is treated as a single super-document. This means that if a research front contains a number of core papers, all citing papers of these core papers were treated as citing papers of this single super-document. The threshold of the normalised co-citation was set as 0.1. The maximum cluster and minimum cluster size are 100 and 6, respectively. The second clustering yielded 133 research areas. These are the basic dataset of this analysis.

Clustering is a useful tool for understanding current scientific activities, but at the same time its limitations should be recognised. This analysis is mainly targeting research areas whose core papers exceed a certain volume, thus small areas are not detectable in the current parameter setting. Tuning the parameters will be effective to detect very small research areas that are about to emerge.

Box 1. Single link clustering method

The following is the simplest example of the single-link clustering. Suppose that there are two pairs of core papers A-B and C-D, where A-B (or C-D) is connected by co-citation which is larger than the clustering threshold. In the single-link clustering regime, if B-C is connected by co-citation, we treat A, B, C, and D as a single cluster even if there is no linkage between A-D. By continuing this procedure, a cluster of papers which are connected by co-citations is formed. The linkage between A-C, A-D, and B-D are also necessary to form a cluster in a complete link clustering. Complete link clustering is suitable to find clusters tightly linked to each other.



2.4 Identification of “hot” research areas

“Hot” research areas, *i.e.* research areas with a high level of activity, were also identified through the following steps (NISTEP 2005a). First, number of citations and average annual growth rate in citations in past six years were calculated in all research fronts. Their field averages were also computed for the ESI’s 22 fields. Next, research fronts in which both the number of citations and the average annual growth rate in citations were larger than the field average were identified. This analysis assigned approximately 1 000 research fronts as hot research fronts. Finally, the number of hot research fronts in each research area was counted and research areas including four or more hot research fronts were selected as hot research areas. Forty-two out of 133 research areas were identified in this research.

2.5 Content analysis of selected research areas

Clustering of highly cited papers yielded 133 research areas. These areas are the basic dataset of this research, but are only lists of papers just after clustering. Thus determination of unique names of research areas is a necessity. In this study, elaborated expert analysis was conducted to understand the contents of areas. To our knowledge, the names of clusters in the usual mapping exercise were determined via keyword analysis or consultation with a few experts.

Preliminary analysis was conducted by experts in NISTEP. They were asked to assign proper names and a brief overview of research areas referring to a map of research area and a list of core papers obtained via the co-citation analysis. These results were sent to over 100 experts in universities, public, and private institutions in Japan in order to check their reliability. Experts were asked to modify the results of the preliminary examination by referring to a map of research areas and a list of core papers.

The content analysis required the most efforts. It took approximately six months to finalise the results. This process was a good opportunity to listen to the experts’ opinions on the validity of the mapping and get valuable insights that are hardly perceived by purely statistical views. The full results of the content analysis are documented in a report published by NISTEP (NISTEP, 2007).

2.6 Mapping of research areas

A mapping of research areas aims to generate a bird's eye view of science activities and trace their changing nature (NISTEP 2005a). How are scientific areas linked to each other? There are immense endeavours to answer this question (Garfield *et al.*, 1964; Small and Sweeney, 1985a; Small *et al.*, 1985b, Börner *et al.*, 2003 and references therein; Boyack, 2004 and references therein; Shiffrin and Börner, 2004, Mane, K. K. and Börner, 2004). Observation of the changing nature of science based on the mapping would be helpful to understand how science is evolving over time.

The map was created by a gravity model (or a force-directed-placement method). In this model, each research area is treated like an atom in a molecule. A research area feels attractive and repulsive forces that are caused by interactions with other research areas. An attractive force between a pair of research areas is evaluated by the following formula:

$$N_{ij}^{\text{norm}} r$$

where N_{ij}^{norm} is the normalised co-citation frequency between research areas i and j , and r is a physical distance between research areas i and j in the map. A repulsive force between a pair of research areas is evaluated by $N_{\text{max}}^{\text{norm}} / r^2$, where $N_{\text{max}}^{\text{norm}}$ is the largest normalised co-citation frequency among all pairs of research areas. This force is a necessity to obtain a stable configuration. A net force for a certain research area is calculated by taking a sum of repulsive and attractive forces for all pairs of research areas. Research areas with similar topics tend to locate closely according to the definition of the attractive force.

In the calculation, first research areas are put randomly in the two-dimensional space.² After that, research areas are moved incrementally in the direction of the net force until the most stable configuration is achieved. Positions of research areas are calculated by the velocity Verlet algorithm (Swope *et al.*, 1982). In the present analysis, the optimisation is continued until the net forces decrease less than a threshold.³ The mapping relies on the attractive and repulsive forces among research areas, thus only relative location of research areas is important. The authors developed Excel VBA programs for the mapping.

The same method was also utilised to generate the map which represents co-authorship among different institutions.

3. Mapping of current scientific activities

One hundred and thirty-three research areas were obtained through the co-citation analysis. A map showing the relative relation among them was generated by the gravity model. The map provides a bird's eye view of current science activities and future analyses will shed light on how research areas are evolving over time.

The map of 133 research areas is shown in Figure 3. Each circle represents a research area and the size of the circle is proportional to the number of citing papers. The bigger circles represent the bigger research areas. Numbers alongside the circles represent ID codes of research areas. Shaded circles represent research areas showing high activities, *i.e.* hot research areas. The relative location of circles

-
2. The mapping is also possible in three-dimensional space, but due to a difficulty of visualisation, the mapping presented here was conducted in the two-dimensional space. There are some attempts to visualise results in three-dimensional space via the Virtual Reality Modelling Language (Chen, 1999; Chen and Paul, 2001).
 3. Threshold is 1.0^{-6} . This value was determined empirically based on the former study in NISTEP (NISTEP, 2005a).

reflects how strongly they are linked to one another. The strongest link stemming from an individual research area is depicted by a solid line, if its normalised co-citation frequency exceeds $0.1 \times N_{\max}^{\text{norm}}$.

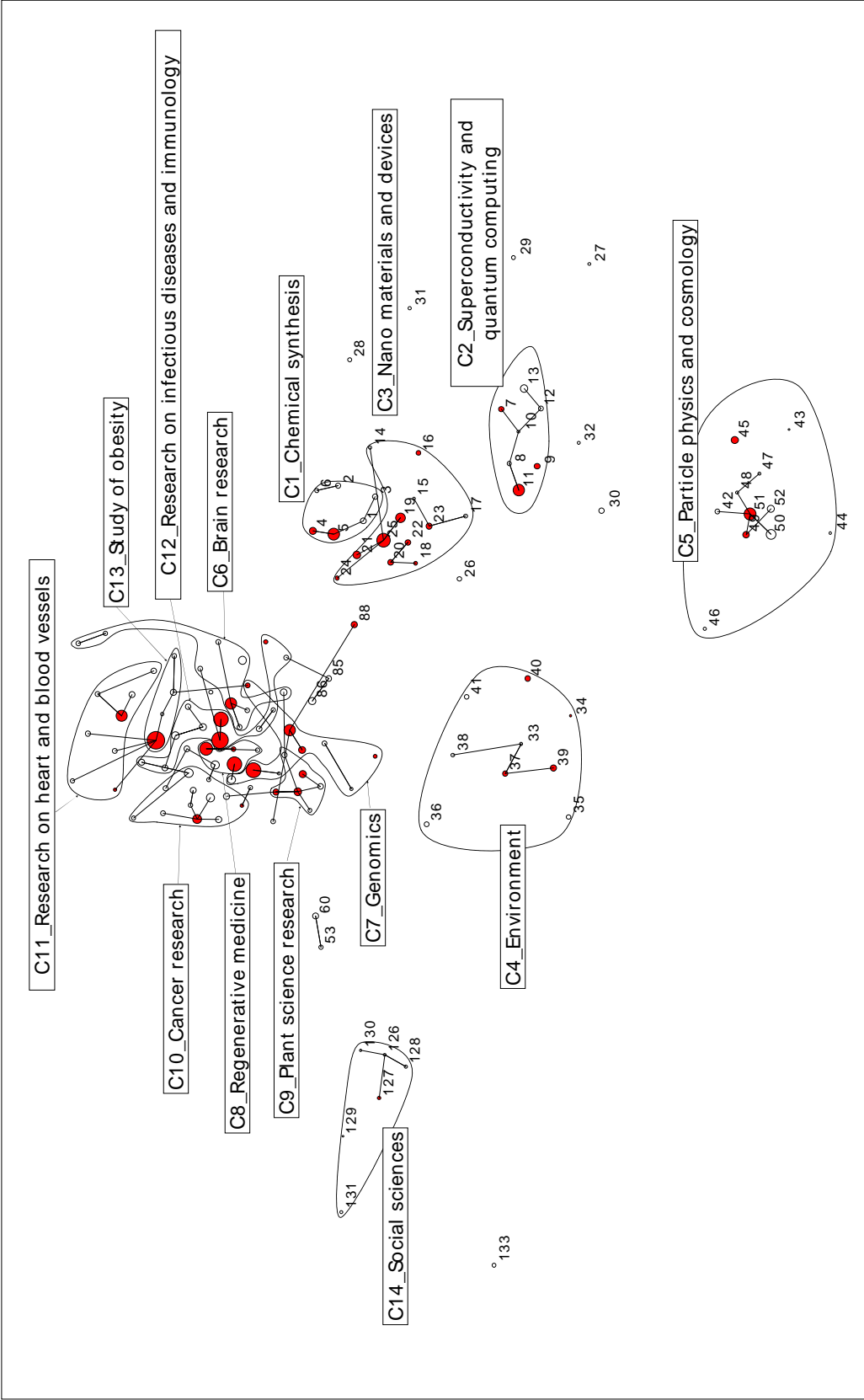
Based on the map and referring results of content analysis to experts, it was found that 133 research areas could be categorised into 14 categories, shown in Table 1. Approximately half of them are related to bioscience or healthcare, reflecting the database structure. Research areas in the same category are indicated by a dotted circle in the map. Since research areas consisting of a substantial number of core papers were selected through clustering, small research areas are not observed in this analysis, *e.g.* those related to mathematics. Their contribution, however, can be observed through analysis of the distribution of mathematics papers in the extracted research areas (see section 4).

Research areas located on the top central part of the map are related to bioscience or healthcare. An enlarged map of this region is shown in Figure 4. Clear separation of bioscience and healthcare is difficult, because research areas show complicated mutual interactions. Roughly though, it could be said that the upper half corresponds to research areas relating to healthcare and the lower half to bioscience.

Thirty-eight research areas fall into bioscience and they make up four research categories. They are “Brain research”, “Genomics”, “Regenerative medicine”, and “Plant science research”. Healthcare consists of 35 areas and 4 categories: “Cancer research”, “Research on infectious diseases and immunology”, “Study of obesity”, and “Research on heart and blood vessels”.

The research categories “Genomics” and “Research on infectious diseases and immunology” are located at the centre of the map, like the trunk of a tree, and other areas are spread widely from it, like branches. This intriguing configuration likely reflects the fact that these categories are going to be the foundations of other categories. These categories also link with “Nano materials and devices”; this point will be discussed further later.

Figure 3. A bird's eye view of current scientific activities: result of mapping 133 research areas



Note: Each circle represents a research area and the size of the circle is proportional to the number of citing papers. Among the circles, shaded ones represent research areas showing high activity. Research areas in the same category are indicated by a dotted circle in the map. The strongest link stemming from individual research areas is depicted by a solid line, if its normalised co-citation frequency exceeds a threshold.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

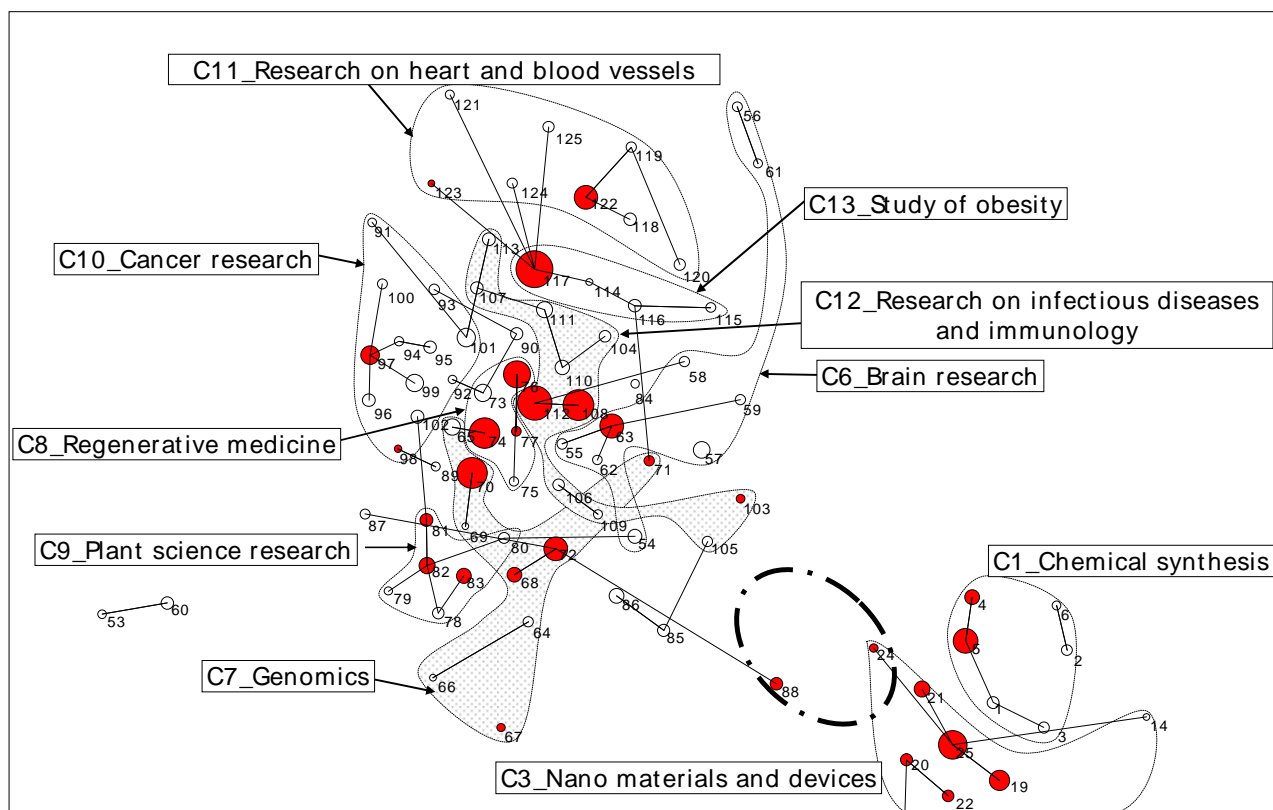
Table 1. A list of research areas by category

Category		ID	Research area
Nanoscience and materials	Chemical synthesis	1	Dendrimer research
		2	High performance catalysis for olefin polymerization
		3	Research on living free-radical polymerization
		4	Asymmetric synthesis
		5	Organic synthesis and its application to a sustainable society
		6	Metal-organic complex and its catalytic activity
	Supeconductivity and quantum computing	7	Basic and application research on MgB2
		8	Quantum computing devices
		9	Spintronics
		10	Physics in high-temperature superconductor junctions
		11	Quantum electronics
		12	Superconductivity in heavy-fermion system
		13	Bi-based high-temperature superconductors
	Nano materials and devices	14	Catalytic activity of gold clusters
		15	Formation of nanostructures based on block copolymers
		16	Novel layer silicate nano-composite
		17	Photonic crystal and devices
		18	Research on high performance organic thin film transistor
		19	Basic and application research on carbon nanotubes
		20	High-efficiency dye-sensitised solar cell
		21	Organic-Metal hybrid meso-porous material
		22	Research on high efficiency organic LED
		23	Synthesis of various nano-structures by self-organization
		24	Research on anion sensing
		25	Research on nanowire, molecular device, and molecular wires
	Others	26	Stability and vitrification of supercooled liquid
		27	Research on meshless finite element method
		28	Research on plastic deformation in nano-crystals
		29	Research on nitride compound semiconductor
		30	Basic and application research on ultra-short-pulse laser
		31	High-dielectric gate insulating technology for semiconductor integrated circuit
		32	Development and application of proton-exchange membrane fuel cell
Environment		33	Carbon cycle in south pacific ocean
		34	Environment pollution and risk of persistent organic halide pollutants
		35	North atlantic oscillation and climate change
		36	Study on biodiversity in plants
		37	Research on global carbon cycle
		38	Research on methane and carbon-dioxide
		39	Effects of aerosol and air pollutant on climate and atmospheric circulation model
		40	Research on spectroanalysis
Particle physics and cosmology		41	Aquatic pollution by toxic chemical compounds
		42	Noncommutative geometry and quantum field theory
		43	Research on novel baryon
		44	Relativistic astronomy and gravity wave
		45	Quantum chromodynamics
		46	Mars exploration
		47	Cosmic microwave background radiation and super string theory
		48	Quantum cosmology
		49	Neutrino study
		50	X-ray astronomy
		51	Cosmic microwave background fluctuation and inflationary cosmology
		52	Supersymmetry and CP violation
Biosciences	Brain research	53	Neuroscience and cognitive-psychology
		54	Glutamate receptors in plasticity brain
		55	Research on neurodegenerative mechanism in Huntington's disease based on transgenic mice
		56	Clinical trial for social phobia and social anxiety disorder
		57	Diversity of mammalian TRP receptors and mechanism of their action
		58	Therapy of multiple sclerosis
		59	Research on prion diseases
		60	Research on visual stimulation and learning
		61	Early diagnostics and therapy of schizophrenia
		62	Mechanism of molecules involved in formation of brain
		63	Research on Alzheimer's disease and Parkinson's disease

Category			ID	Research area
Biosciences	Gnomics	Genome and bioinformatics	64	Genome analyses of <i>Helicobacter pylori</i> and <i>Mycobacterium tuberculosis</i>
			65	Molecular mechanism of DNA damage and repair
			66	Utility of agrobacterium-mediated genetic engineering and the genomic character
			67	Research on isoprenoid biosynthesis pathway in malaria parasite
			68	Network analysis and its application to genome, social-network, and infection transmission
			69	Development of statistics method for microarray data analysis
			70	Research on epigenetic transcriptional regulation
	Proteomics		71	Dug discovery research
			72	Research on proteome
	Regenerative medicine		73	Molecular mechanism of PI3/Akt signal transduction pathway
			74	Research on molecular mechanism in apoptosis
			75	Research on stem cell therapy and its side effects
			76	Stem cell therapy on nervous, hematopoietic, and cardiovascular system
			77	Research on multipotency and differentiation mechanism of stem cell in cardiovascular system, cancer, and embryo
	Plant science research		78	Plant genome research
			79	Functional analysis on abscisic acid, a plant hormone
			80	Dynamics and regulation of cytoskeleton
			81	Functional analysis on auxin, a plant hormone
			82	Stress response in plants
			83	Study of biological clock
	Others		84	Research on signal transduction in cell proliferation
			85	Material transport mechanism in organism
			86	Structures and functions of G-protein-coupled receptor
			87	Nucleocytoplasmic traffic and cell function
			88	Research on microanalysis of biochemical substances
			89	Mechanism of control of life-span
			90	Signal transduction in metabolic pathway
Healthcare	Cancer research		91	Research on medical examination of colorectal cancer
			92	Function study of mammalian TOR
			93	Research on molecular mechanism of sex hormone receptor against cancer
			94	Cancer therapy with angiogenesis inhibitor
			95	Research on multiple myeloma therapy
			96	Research on therapy of B-cell neoplasms
			97	Research on cancer therapy (Molecular target drug for cancer)
			98	Cancer therapy with histone deacetylase inhibitor
			99	Research on leukemia therapy
			100	Research on breast cancer therapy
			101	Effects of COX-2 inhibitor against cancer
			102	Hypoxia-inducible factor and tumorigenesis
	Research on infectious diseases and immunology		103	Research on SARS and avian influenza
			104	Research on allergy therapy
			105	Research on countermeasure to bioterrorism
			106	Signal transduction in immune system
			107	Research on osteoclastic mechanism
			108	Research on infection mechanism and therapy of HCV and HIV
			109	Research on infection mechanism of HIV
			110	Genetic diagnosis and therapy of Crohn's disease
			111	Side effects of medicine for autoimmune disease
			112	Research on immune system
			113	Clinical research on COX-2 inhibitor as anti-inflammatory drug
	Study of obesity		114	Functional analysis of peroxisome proliferator-activated receptor
			115	Ghrelin; its mechanism of action
			116	Research on appetite and metabolism
			117	Functional analysis on peroxisome proliferator-activate receptor
	Research on heart and blood vessels		118	Clinical treatment of severe sepsis and septic shock
			119	Implantable cardioverter defibrillator (ICD) therapy
			120	Various responsible gene of cardiac arrhythmia
			121	Development of imaging techniques for cardiovascular system
			122	Clinical trial of therapeutic agent for cardiovascular disease
			123	Effect of dietary intake of hormone on cardiovascular health
			124	Therapy of arterial stenosis with drug-eluting stents
			125	Research on venous thromboembolism therapy
Social sciences		126	Behavioral economics	
		127	Research on corporate governance	
		128	Political power and human right	
		129	Research on venture capital	
		130	Research on intellectual property right problems	
		131	Study on local economy	
Others		132	Application of cryptographic technologies to digital information distribution	
		133	Research on modulation schemes for ultra-wide band communication	

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

Figure 4. Enlarged map of bioscience and healthcare



Note: Each circle represents a research area and the size of the circle is proportional to the number of citing papers. Shaded circles represent research areas showing high activity. Research areas in the same category are indicated by a dotted circle in the map. The strongest link stemming from an individual research area is depicted by a solid line, if its normalised co-citation frequency exceeds a threshold.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

In the middle right region in Figure 3, there are 32 research areas related to “Nanoscience and materials”. The region can further be divided into three categories. In the middle, there is a research category related to “Nano materials and devices”. There are 12 research areas in this category. It seems that “Research on nanowire, molecular device, and molecular wires (ID25)” plays a central role in these areas; its central location in the map and four direct links to other areas rationalise its importance.

There are six research areas associated with “Chemical synthesis”. Exploring efficient catalyses is a major objective of these areas. The majority of core papers in this category fall into chemistry. Seven research areas at the bottom are associated with “Superconductivity and quantum computing”. Most of the core papers in this category fall into physics.

The location of “Nano materials and devices” between “Chemical synthesis” and “Superconductivity and quantum computing” is apparently an illustration of multi-disciplinary character. The mapping only relies on co-citation relations among research areas, thus the intermediate location implies that this research category is materialised through active mutual interaction between physics and chemistry. It is worth mentioning that “Nanoscience and materials” is slightly attracted towards “Bioscience” as shown in Figure 4. This presumably indicates the existence of weak interactions between them; the interactions are not strong enough to substantiate visible research areas. The blank region between them, which is indicated by a thick dashed-dotted circle in the map, may be a ground for an emerging area, *e.g.* bio-nanoscience. There is a research area on “Research on microanalysis of biochemical substances (ID88)”. This area aims

to develop tiny biochemical reaction fields like a Lab-on-chip, which is expected to be a key technology in personalised medical care in the future. Furthermore, some research areas in “Nano materials and devices” have multi-disciplinary research topics, such as “Research on anion sensing (ID24)”. Many research areas related to “bio-nanoscience” could be observed, if a subsequent analysis were conducted based on the latest dataset.

The category “Particle physics and cosmology” lies in the bottom right of the map in Figure 3. Its isolated location suggests weak interaction with other categories. In other words, it can be said that direct knowledge flows from “Particle physics and cosmology” to other categories are relatively small, and the knowledge in this category is materialised thorough participation of researchers from restricted traditional disciplines.

“Environment” is located at centre of the map: there are nine research areas in this category. It mainly consists of research on climate changes, chemical pollutants, and biodiversity. The central location likely shows active knowledge flows between “Environment” and other research, and is a reflection of its notable multi-disciplinary character.

Though this study mainly focused on natural sciences, some research areas related to social sciences were found. They are located at the middle left in the map. It is worth mentioning that two research areas, “Neuroscience and cognitive-psychology (ID53)” and “Research on visual stimulation and learning (ID60)”, are found in the intermediate location between bioscience and social sciences. It could be said that collaboration between researchers from bioscience and social sciences is beneficial to the development of these research areas.

4. Measurement of the multi-disciplinary character of research

The importance of multi-disciplinary science and technology fields has been recognised in recent years. Interaction among researchers from different fields is expected to bring many occasions for new discoveries and innovation. Measurement of the multi-disciplinary character of science and technology fields would be beneficial in order to understand the characteristics of scientific activities in multi-disciplinary fields.

The ESI’s 22 fields were further reclassified into 10 fields, for analysing the distribution of scientific fields. These 10 fields are: chemistry; material sciences; physics; space science & geosciences; computer science & mathematics; engineering; environment/ecology; clinical medicine; basic life sciences; and others. Basic life sciences include eight⁴ and others include four⁵ ESI fields. Such analysis would show how many scientific fields are involved in the progress of research. The results would be a proxy for measuring the multi-disciplinary character of research categories.

Information entropy was also calculated in order to measure the multi-disciplinarily character of research categories. The entropy was proposed by Shannon (Shannon, 1948). It is usually used to measure randomness (or uncertainty) in information. It is defined by:

-
4. These are Agricultural Sciences, Biology & Biochemistry, Immunology, Microbiology, Molecular Biology & Genetics, Neuroscience & Behaviour, Pharmacology & Toxicology, and Plant & Animal Science.
 5. These are Economics & Business, Multidisciplinary, Psychiatry/Psychology, and Social Sciences, general.

$$H(x) = -\sum_{i=1}^n p(i) \ln p(i) \quad (1)$$

where $p(i)$ is the share (in other words probability) of field i in core or citing papers in a certain category. If core or citing papers are composed of a single field, Eq. (1) equals zero. Equation (1) reaches a maximum when all fields have equal shares, *i.e.* $p(i)=1/10$.

Results of this kind of exercise strongly depend on the selection of a basic set of fields. For example, Eq. (1) tends to overestimate the multi-disciplinary character of basic life sciences-related categories, if the analysis is conducted based on the ESI's 22 fields, because classification of bioscience-related fields are sub-divided in the ESI classification. Therefore, eight ESI fields related to basic life sciences are treated as one field in this analysis for assessing the multidisciplinary character across the research categories obtained.

Information entropies are shown in Table 2, where $H(\text{Core})$ and $H(\text{Citing})$ represent the information entropy in core and citing papers, respectively. It clearly shows the high multi-disciplinary character of “Nano materials and devices”, “Environment”, and “Genomics”.

Table 2. Information entropy in core and citing papers by category

	Category	Research areas	Core paper	Citing paper	H(Core)	H(Citing)	H(Citing)-H(Core)
Nanoscience and materials	Chemical synthesis	6	728	15906	0.12	0.31	0.19
	Superconductivity and quantum computing	7	650	17119	0.10	0.44	0.34
	Nano materials and devices	12	1312	28614	1.15	1.22	0.07
	Particle physics and cosmology	11	1460	23021	0.52	0.74	0.22
	Environment	9	626	11302	1.38	1.66	0.28
Bioscience	Brain research	11	630	22343	1.03	0.90	-0.13
	Genomics	9	763	28017	1.40	1.33	-0.07
	Regenerative medicine	5	527	26375	0.78	0.86	0.08
	Plant science research	6	561	11574	0.20	0.55	0.35
Healthcare	Cancer research	12	573	22818	0.48	0.77	0.29
	Research on infectious diseases and immunology	11	966	40444	0.76	0.90	0.14
	Study of obesity	4	525	20314	0.68	0.81	0.13
	Research on heart and blood vessels	8	365	15041	0.17	0.46	0.29

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

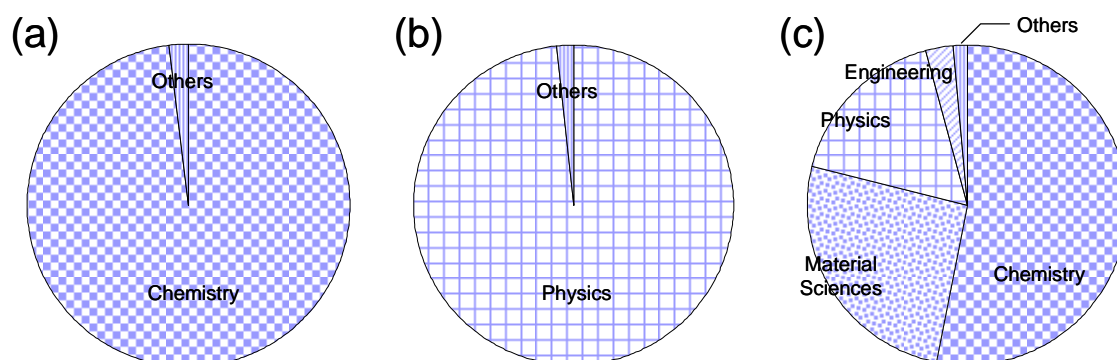
4.1 Nanoscience and materials

Small information entropies in “Chemical synthesis” and “Superconductivity and quantum computing” indicate dominance of a single field. Knowledge in “Chemical Synthesis” mainly comes from chemistry. Physics is dominant in “Superconductivity and quantum computing”. It could be said that these single-field-dominated categories are those mainly based on the traditional scientific fields [see Figures 5(a) and (b)].

The information entropy in “Nano materials and devices” exceeds unity, indicating a high degree of multi-disciplinary character. Actually the category is composed of a combination of knowledge from

various fields. Chemistry has the largest share, followed by material science and physics, as shown in Figure 5(c). It is worth noting that biology & biochemistry has a small share (about 1%) in this category. Application of DNA molecules to nanowires and synthesis of DNA-metal particles hybrid materials are examples of research topics involving biology & biochemistry. Characteristic of information entropies is almost the same between core and citing papers. This leads to a small differential in the information entropy as shown in Table 2.

Figure 5. **Distribution of fields in core papers; (a) Chemical synthesis, (b) Superconductivity and quantum computing, and (c) Nano materials and devices**



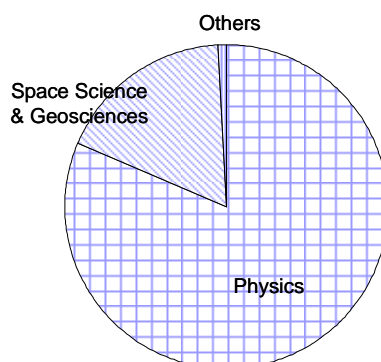
Note: Fields whose share in core papers is less than 2% are included in "others" in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

4.2 Particle physics and cosmology

Research on "Particle physics and cosmology" is constituted by core papers from physics and space science, and it leads to a small information entropy reflecting dominance of these fields. The tendency is identical in both core and citing papers. Knowledge of "Particle physics and cosmology" mainly spreads within physics and space science.

Figure 6. **Distribution of fields in Particle physics and cosmology (Core papers)**



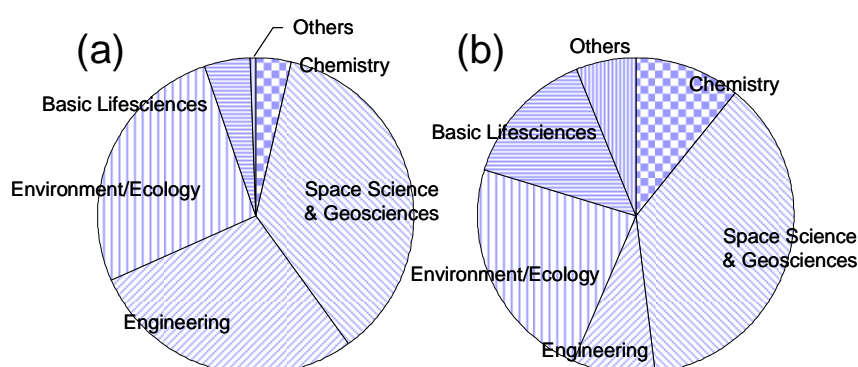
Note: Fields whose share in core papers is less than 2% are included in others in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

4.3 Environment

Large information entropy in “Environment” clearly shows the multi-disciplinary character of this research category. Geosciences have the largest share followed by engineering and environment/ecology in the core papers, as shown in Figure 7. Core and citing papers have different field distributions. Shares of chemistry and basic life sciences, especially plant and animal sciences, show a remarkable increase in citing papers. In contrast, the share of engineering goes down significantly. This seems to be attributable to different characteristics in research in core and citing papers. According to the content analysis of core papers, it was revealed that core papers in this research category are mainly dedicated to the development of research equipment, *i.e.* basic techniques to measure properties of materials. Meanwhile, citing papers are mainly focusing on application of these techniques to various targets like organic halide pollutants.

Figure 7. Distribution of fields in (a) core and (b) citing papers in Environment



Note: Fields whose share in core papers is less than 2% are included in others in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

4.4 Bioscience

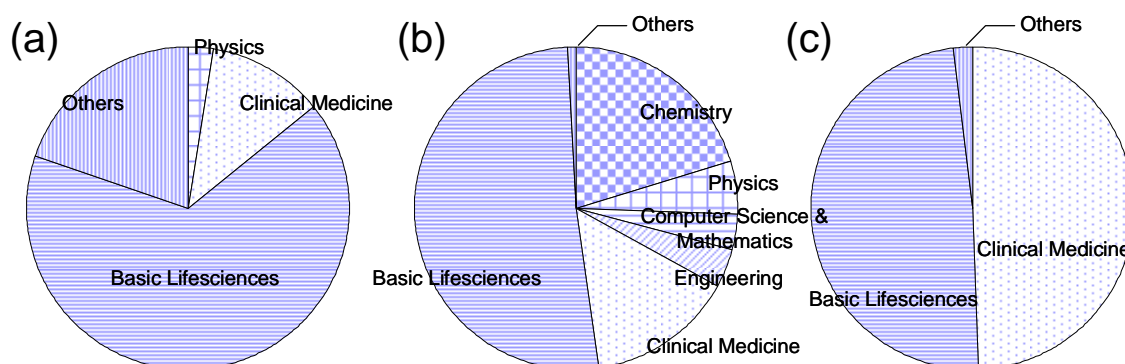
The information entropy of “Brain research” is the fourth largest in core papers. Core papers consist of basic life sciences (neuroscience & behaviour and biology & biochemistry), clinical medicine, and others, especially psychiatry/psychology. A large share of psychiatry/psychology is a unique characteristic in “Brain research”.

The largest information entropy in core papers *i.e.* 1.40, indicates the high multi-disciplinary character of “Genomics” research. Basic life sciences (microbiology & genetics and biology & biochemistry), chemistry, and clinical medicine are three major fields. Physics (5.5%), engineering (3.9%), computer science and mathematics (3.5%) have a non-negligible share in core papers. Their summation accounts for about 10% of total core papers. Their presence in citing papers is smaller than that in core papers. These results suggest that knowledge of non-life sciences has a crucial role in “Genomics”, especially in highly-cited research.

In “Regenerative medicine”, clinical medicine and basic life sciences account for 98% of total core papers. Molecular biology & genetics, biology & biochemistry, and neuroscience & behaviour are major fields in basic life sciences. Comparable shares of clinical medicine and basic life sciences suggest the importance of close co-operation between basic and application research.

Information entropy in “Plant science research” is very small, reflecting the dominance of basic life sciences in its field distribution. Detailed analysis, however, shows qualitative differences in distribution between core and citing papers. Core papers in basic life sciences mainly come from plant & animal science. In contrast, shares of other scientific fields show a remarkable increase in citing papers. Shares of biology & biochemistry and molecular biology & genetics increase by about 10%. These results indicate the wide dissemination of discoveries in “Plant science research” to research in basic life sciences. A typical example of such wide knowledge dissemination is the RNA interference. The phenomenon was first discovered in petunia plants. After the initial observations, extensive research on *C. elegans* were conducted and its effectiveness as a tool for controlling gene expressions was revealed (Fire *et al.*, 1998). Now the phenomenon attracts wide recognition from scientists working on genomics and drug discovery research.

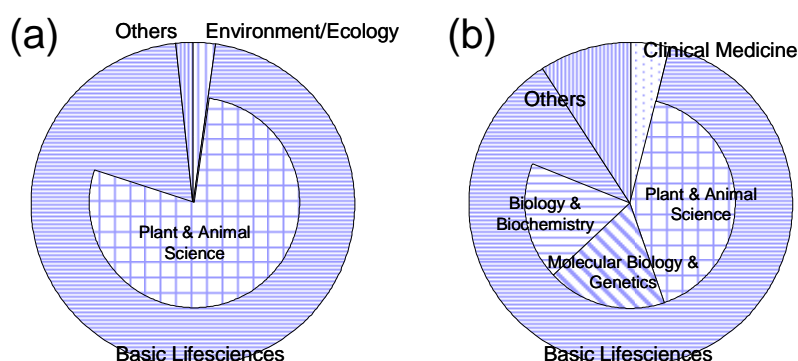
Figure 8. **Distribution of fields in core papers; (a) Brain research, (b) Genomics, and (c) Regeneration medicine**



Note: Fields whose share in core papers is less than 2% are included in others in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

Figure 9. **Distribution of fields in (a) core and (b) citing papers in Plant science research**



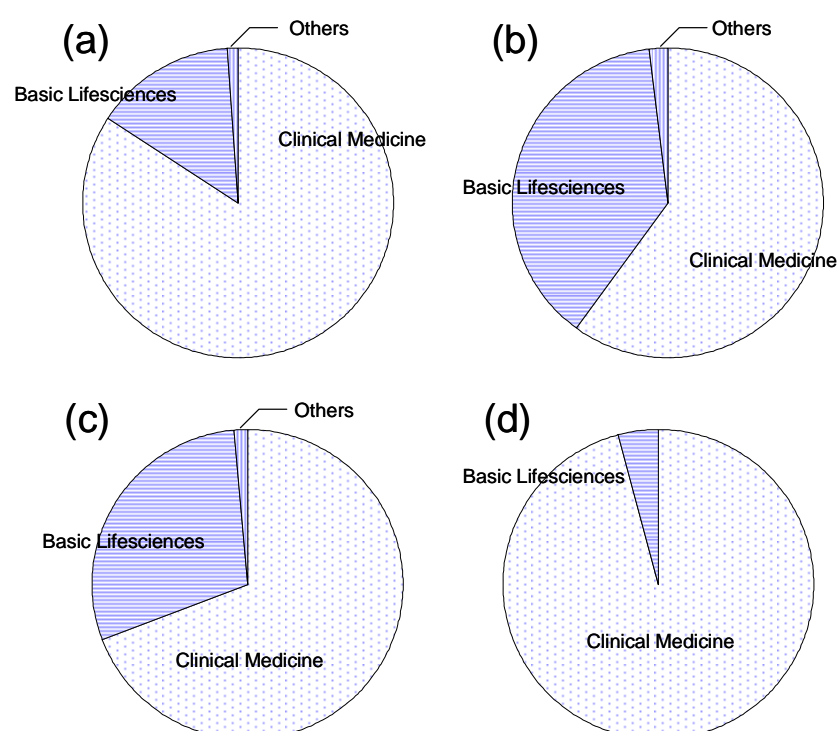
Note: Fields whose share in core (citing) papers is less than 2% are included in others in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

4.5 Healthcare

Information entropies in healthcare-related categories are not as high as those in bioscience. Clinical medicine dominates core papers in “Cancer research” and “Research on heart and blood vessels”. Basic life sciences account for 38% in “Research on infectious diseases and immunology” and 29% in “Study of obesity”, indicating the crucial role of basic life sciences in development of these categories. Immunology, microbiology, and biology & biochemistry are major basic life sciences fields in “Research on infectious diseases and immunology”. Biology & biochemistry and pharmacology & toxicology cover a large part of core papers from basic life sciences in “Study of obesity”.

Figure 10. **Distribution of fields in core papers; (a) Cancer research, (b) Research on infectious diseases and immunology, (c) Study of obesity, and (d) Research on heart and blood vessel**



Note: Fields whose share in core papers are less than 2% are included in others in the chart.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

5. Measurement of countries' share

Shares in core and citing papers are an indicator for measuring countries' specialisation in a certain area of research. The shares of the OECD countries, the EU15 (European Union)⁶, the ASEAN+3 (Association of South East Asian Nations)⁷, the BRICs (Brazil, Russian Federation, India, and China), and Chinese Taipei were investigated. In this study, the share of core paper is used as a proxy to measure the quality of research, while the share of citing papers is considered a sign of catch up, especially in non-OECD countries. It is known that researchers in "catching-up" countries tend to cite more papers than researchers in prestigious countries. They need to get more accumulated knowledge in order to catch up on the state-of-the-art research.

The shares were analysed based on the whole count method. In this method, the contribution of a paper from a certain country is counted as a unity, if at least one address in that country is included in the address list. Thus the sum of countries' shares exceeds 100%. Countries' contributions are evaluated based on the addresses of institutions, not the nationality of authors. Another commonly used method is the fractional count method. Contribution of an institution is weighted by its share in all contributions. This means that if there are ten institutions and two of them are from a target country, the presence of the country is calculated as 0.2. Therefore, the sum of countries' shares always equals 100%. There are several examples in the literature showing the differences between these two methods (Rousseau 1992, Gauffriau and Larsen 2005). The methodological difference is usually observed as the difference in the share, but it rarely changes country's order.

5.1 Nanoscience and materials

Figure 11 shows countries' share in three categories in "Nanoscience and materials". The largest shares of the United States in all three categories in core papers indicate its leading role. US dominance is, however, weakened in citing papers. The EU15 have the largest share in "Chemical synthesis" and "Superconductivity and quantum computing". Competition between the US, the EU15, and the ASEAN+3 is tough in "Nano materials and devices". The ASEAN+3 have the largest share in citing papers. The United States seems to have some advantage in terms of quality of papers.

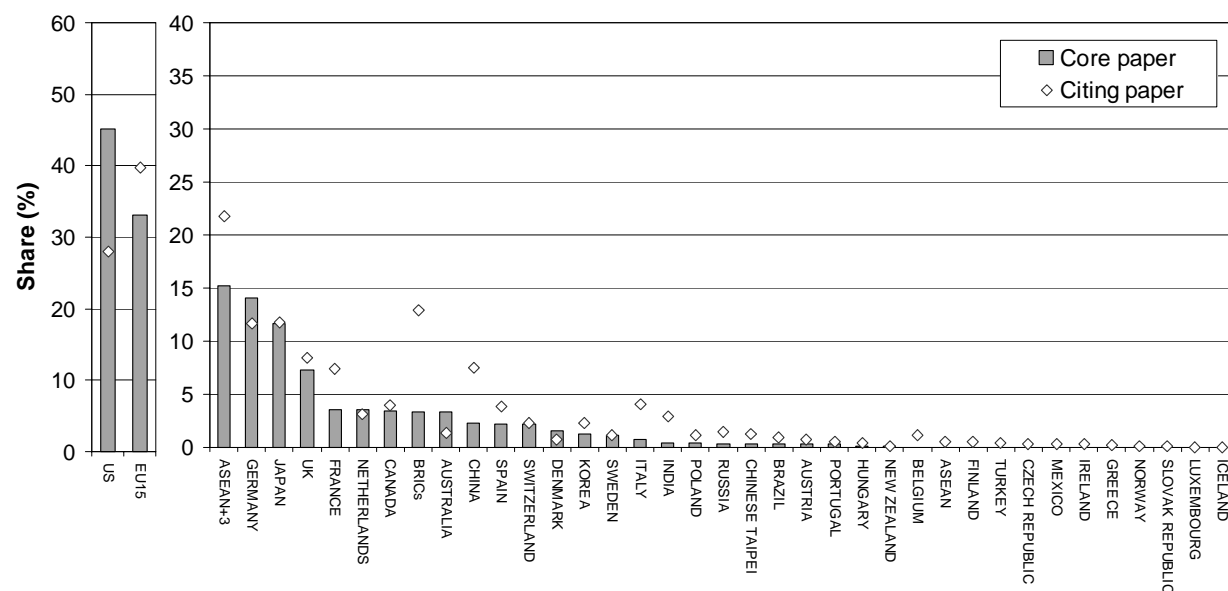
The BRICs have the fourth largest share in citing papers in all categories. Contributions come mainly from China. The number of citing papers is 2 ~ 4 times larger than those of core papers. This indicates that BRICs are now in the intense catching-up stage in nanoscience and materials research.

At the country level, Japan has the second largest share in core papers in "Superconductivity and quantum computing" and "Nano materials and devices". Germany has the second largest share in core papers in "Chemical synthesis". China is already the top-ranking producer of citing papers, with the second largest share in "Nano materials and devices". This suggests that the role played by China should be considered when examining knowledge creation in "Nanoscience and materials". This observation is consistent with other literature (Kostoff, 2004; Zhou and Leydesdorff, 2006).

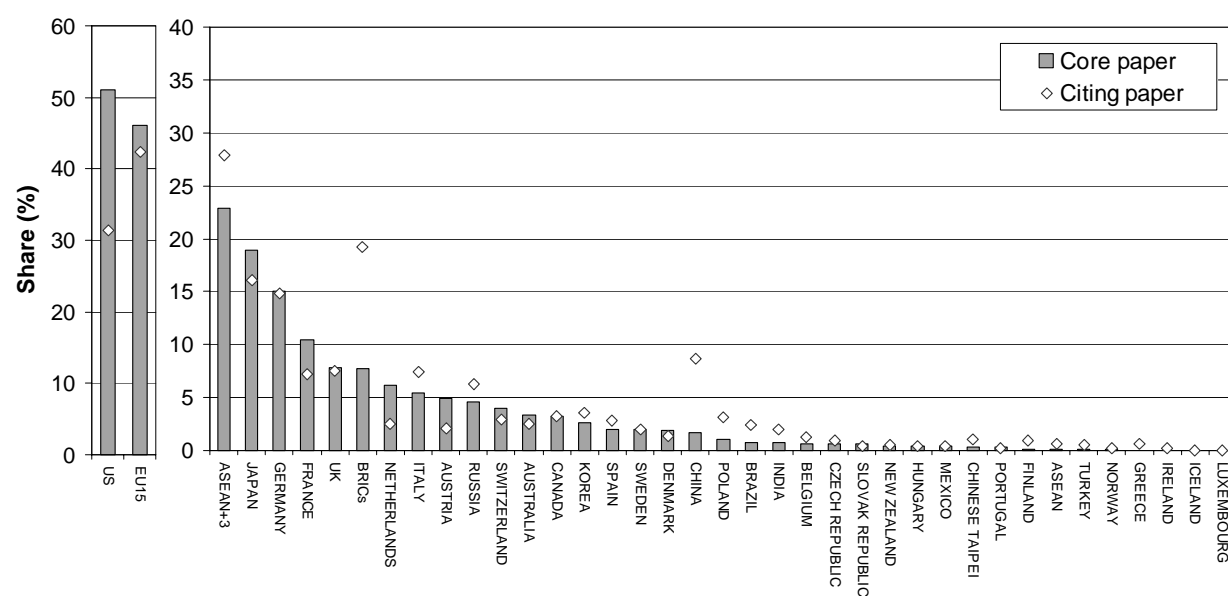
-
6. The EU15 countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain, Sweden, and the United Kingdom.
 7. The ASEAN countries: Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, and Vietnam. +three countries: China, Japan, and Korea.

Figure 11. Countries' share in core and citing papers by research categories; (a) Chemical synthesis, (b) Superconductivity and quantum computing, and (c) Nano materials and devices

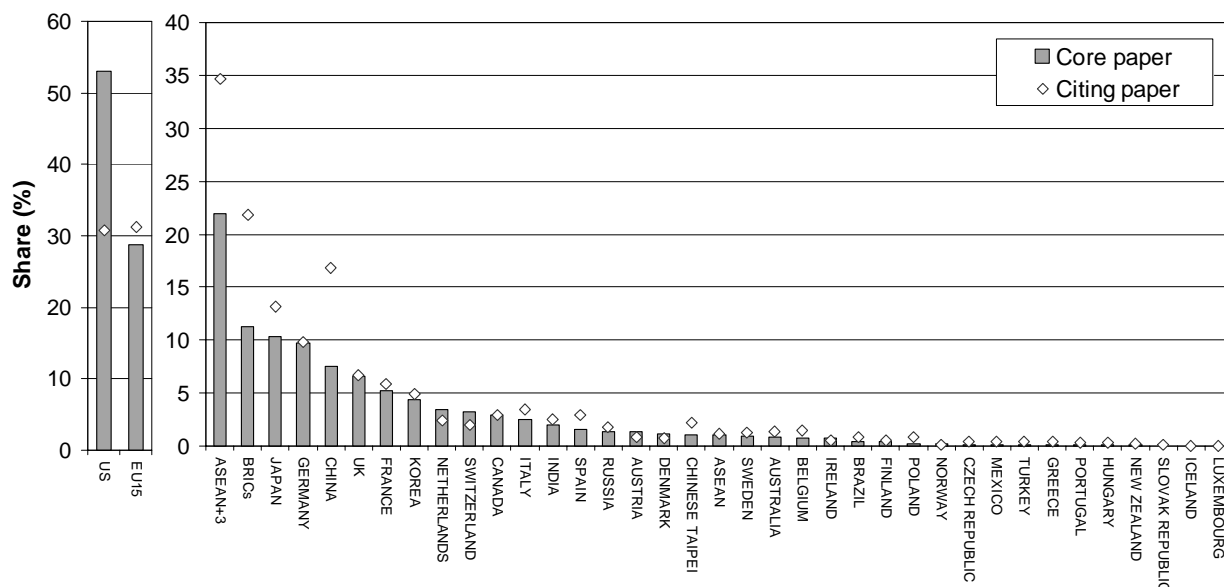
(a) Chemical Synthesis



(b) Superconductivity and quantum computing



(c) Nano materials and devices



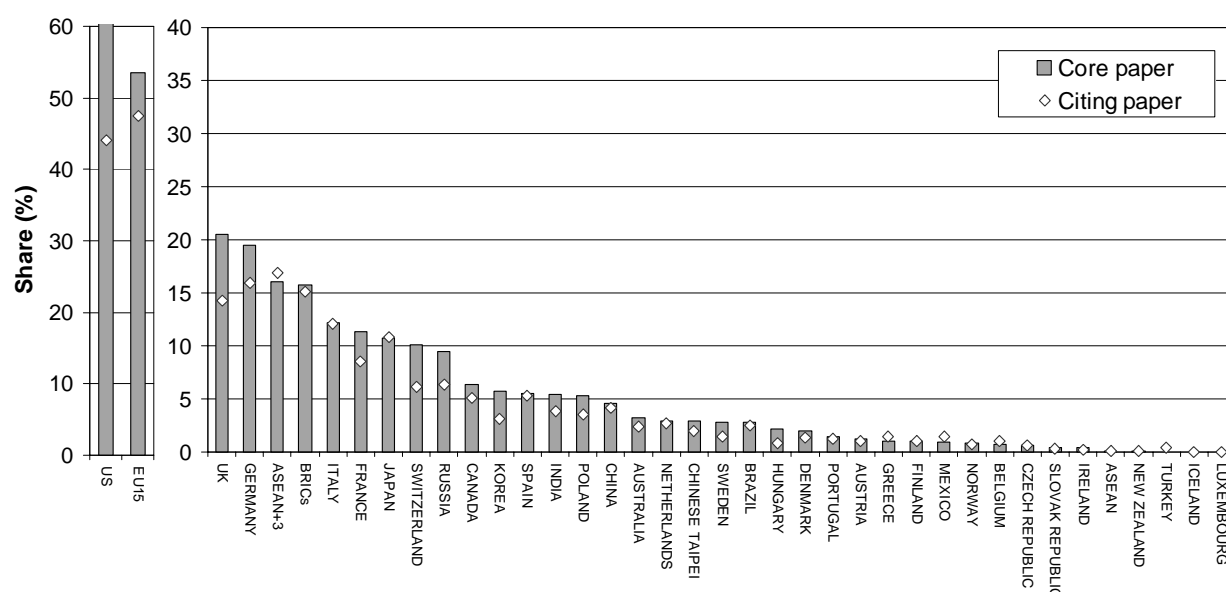
Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

5.2 Particle physics and cosmology

The United States and the EU15 complete each other in “Particle physics and cosmology”. The EU15 has the largest share in citing papers. Switzerland’s large share is attributable to the CERN (European Organisation for Nuclear Research), the world largest research centre on high energy particle physics studies. The CERN has the largest share at the institutional level analysis. Switzerland’s shares in “Neutrino study (ID49)” and “Supersymmetry and CP violation (ID52)” account for nearly 20% and 30% of total core papers. Eastern European countries, such as Poland, Russia, Hungary, Slovak Republic, also produce extremely larger numbers of scientific papers compared to the country average.

Figure 12. Countries' share in core and citing papers in Particle physics and cosmology



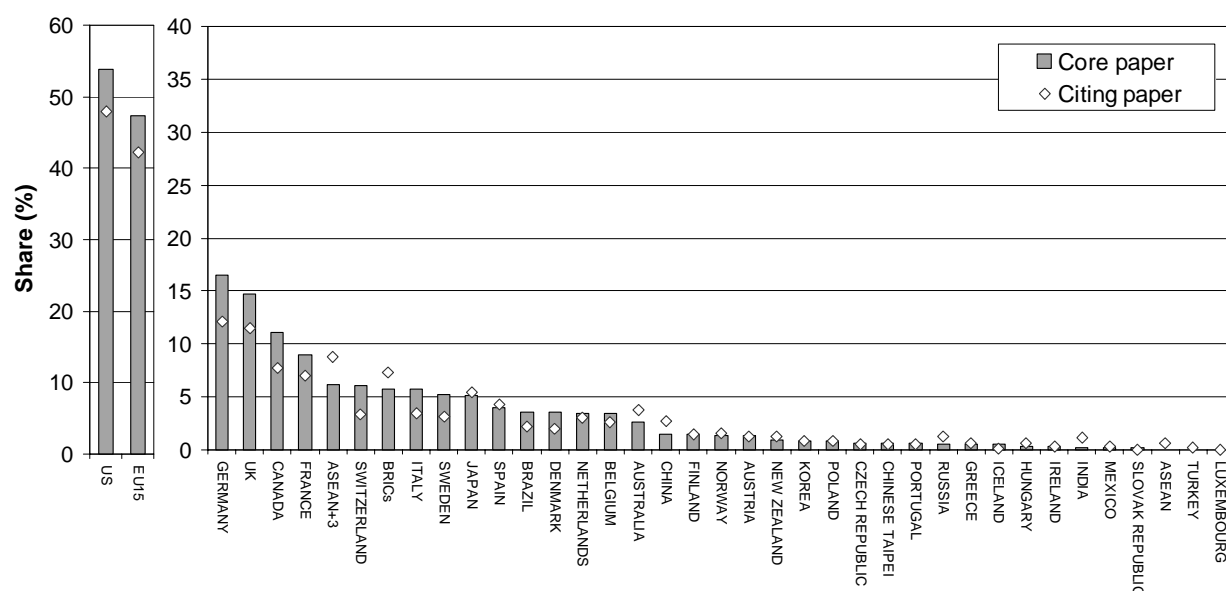
Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

5.3 Environment

The United States has the largest share, followed by the EU15. The shares of the ASEAN+3 and the BRICs except Brazil are not so large at present. Nordic countries, Brazil, Canada, and New Zealand have a large share both in core and citing papers compared to the country average.

Figure 13. Countries' share in core and citing papers in Environment



Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

5.4 Bioscience

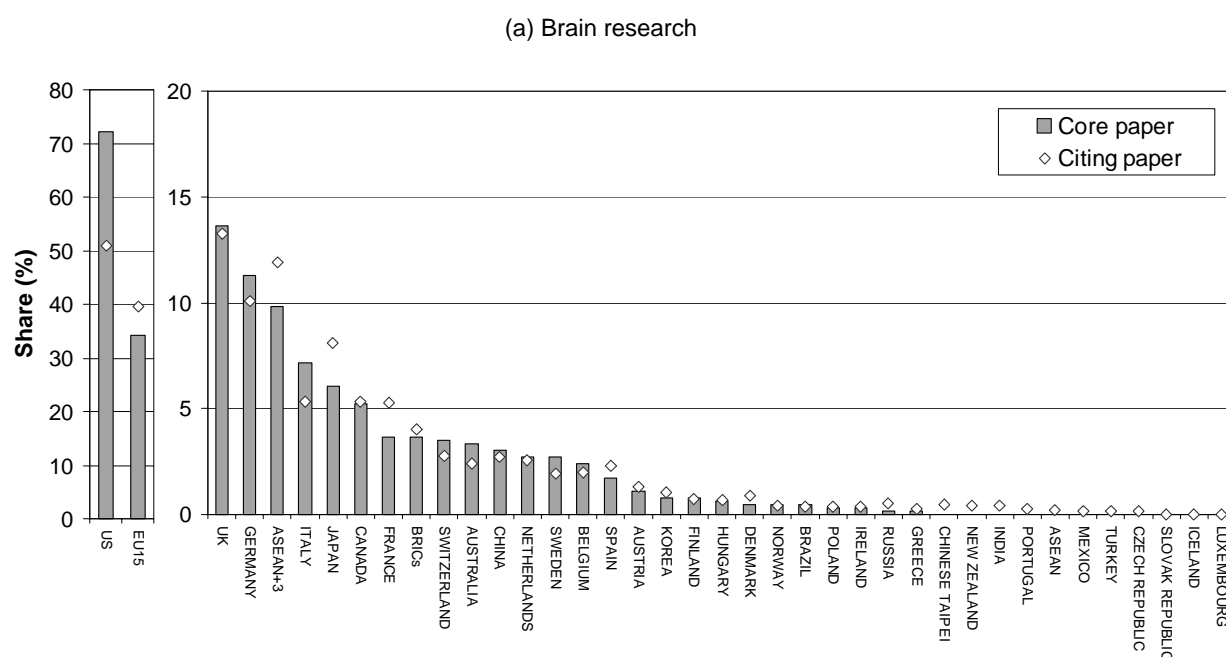
The United States has the largest share both in core and citing papers in all categories, followed by the EU15. The shares account for about 70% of core papers, indicating the prominent position of the United States. The US dominance is, however, weakened in citing papers. At the country level, the United Kingdom has the second largest share in “Brain research” and “Genomics”, Germany in “Plant science research”, and Japan in “Regenerative medicine”.

Denmark and Austria have a remarkably high share of core papers in “Genomics”. Italy has a large share in research areas on “Neuroscience and cognitive-psychology (ID53)” and “Therapy of multiple sclerosis (ID58)” leading to a relatively large share in the research category “Brain research”.

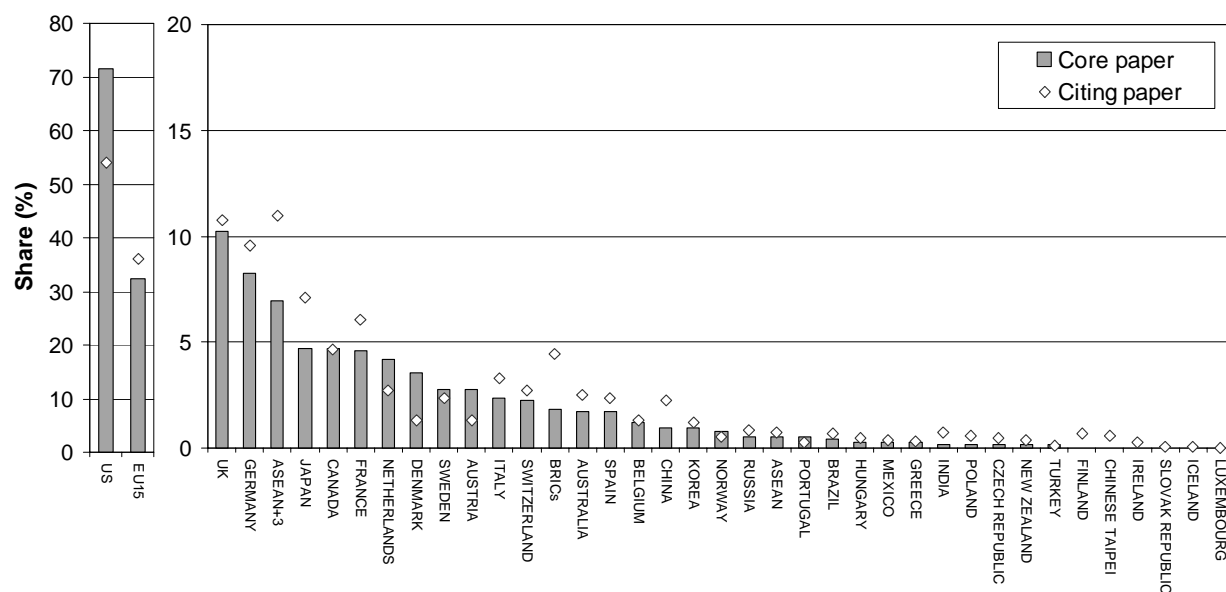
Dominance of the United States in “Plant science research” is not as remarkable as in other categories. Germany, the United Kingdom, and Japan compete with each other in this category. Korea also has a relatively large share: its core papers share is larger than the country average in four out of six research areas related to “Plant science research”.

The BRICs’ share is still small in both core and citing papers. It seems that it would take a while until the BRICs influence knowledge creation and flows in “Bioscience” as in “Nanoscience and materials”.

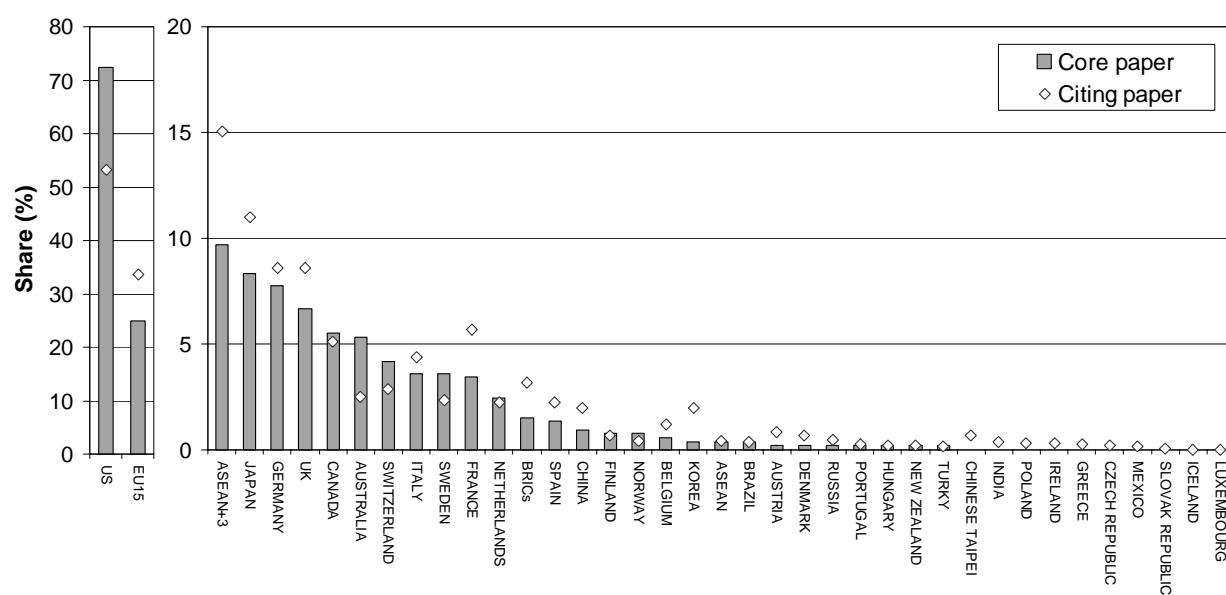
Figure 14. **Countries’ share in core and citing papers by research categories; (a) Brain research, (b) Genomics, (c) Regenerative medicine, and (d) Plant science research**



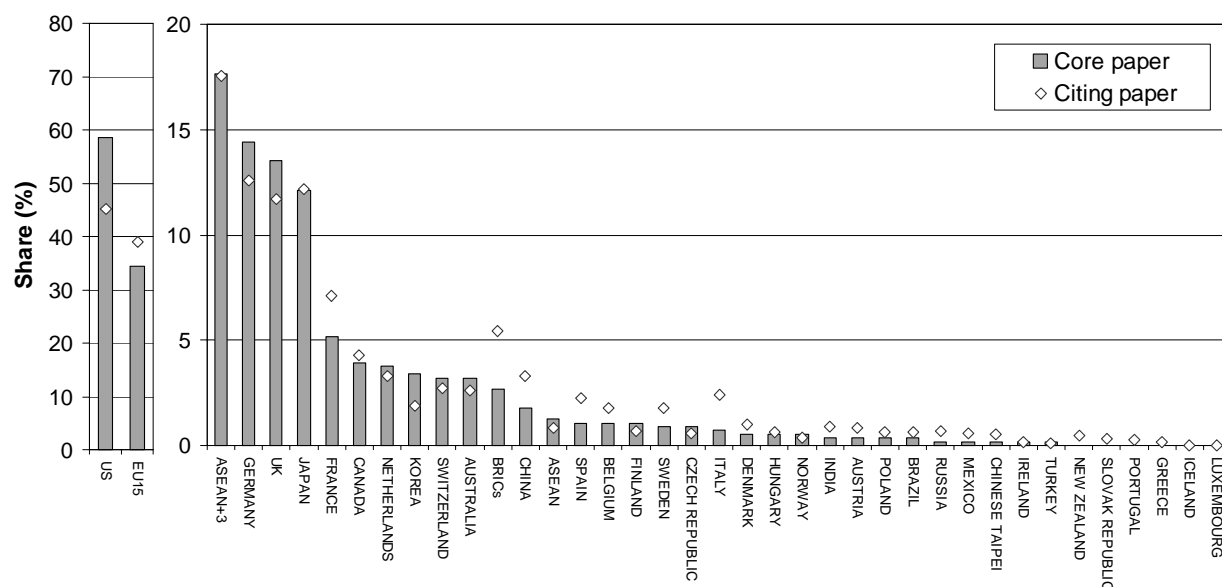
(b) Genomics



(c) Regenerative medicine



(d) Plant science research



Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

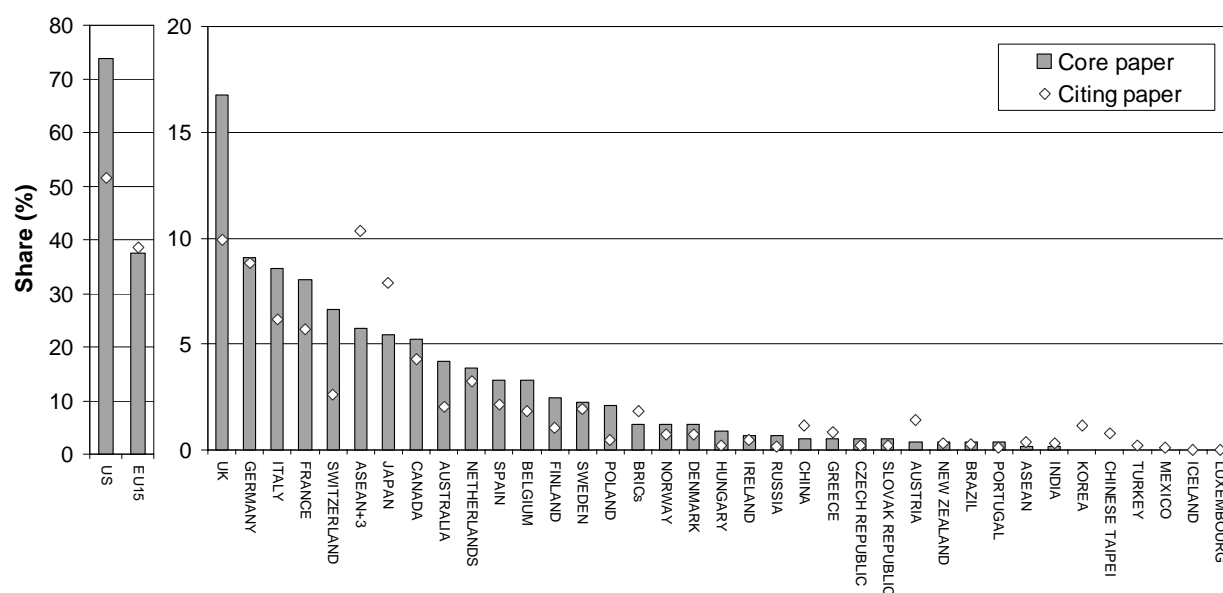
5.5 Healthcare

The United States has the largest share in all four categories, followed by the EU15. In terms of average share in healthcare, Canada has the fourth largest share in core papers at the country level. This is due to a relatively large share in “Research on heart and blood vessels”. The Netherlands, Sweden, Belgium, and Denmark also have comparably large shares in this research category.

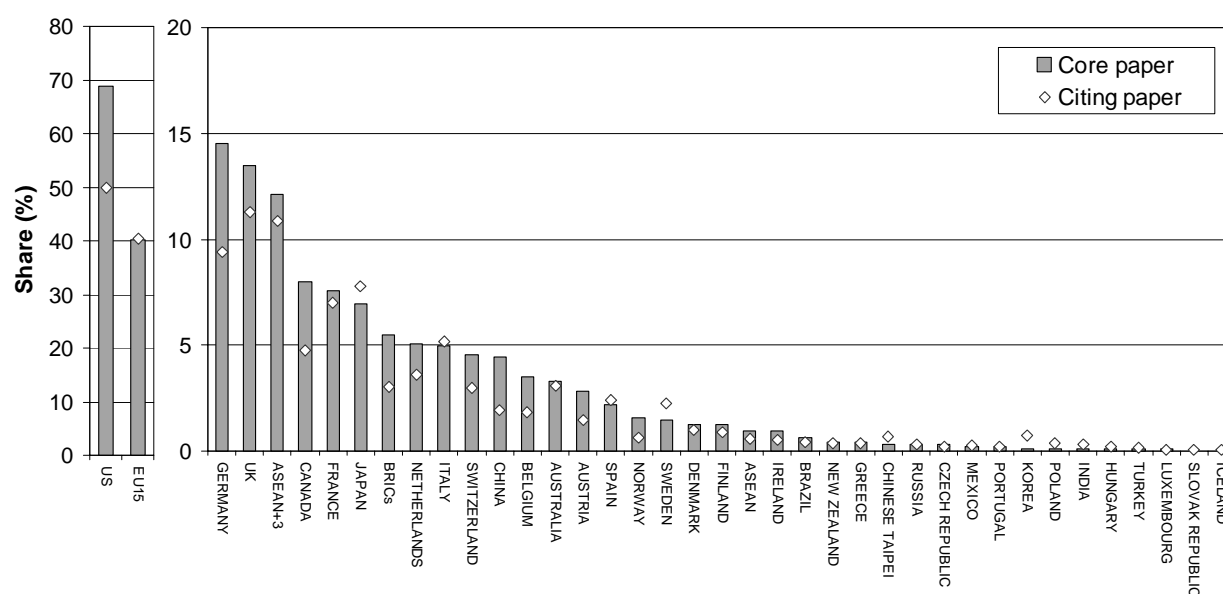
The BRICs have a large share in “Research on infectious diseases and immunology”. This is due to China’s large share in “Research on SARS (Severe Acute Respiratory Syndrome) and avian influenza (ID103)”, currently it holds about a 46% share in core papers. It is worth mentioning that the ASEAN countries also have a 7% share. This figure is remarkably high in contrast to their share in total core papers (about 0.5%). China’s large share may be explained by the geometrical factor, because the SARS epidemics started from China in 2002. But it also indicates China’s capability in healthcare which enables it to quickly respond to an unexpected epidemic of infection diseases.

Figure 15. Countries' share in core and citing papers by research categories; (a) Cancer research, (b) Research on infectious diseases and immunology, (c) Study of obesity, and (d) Research on heart and blood vessels

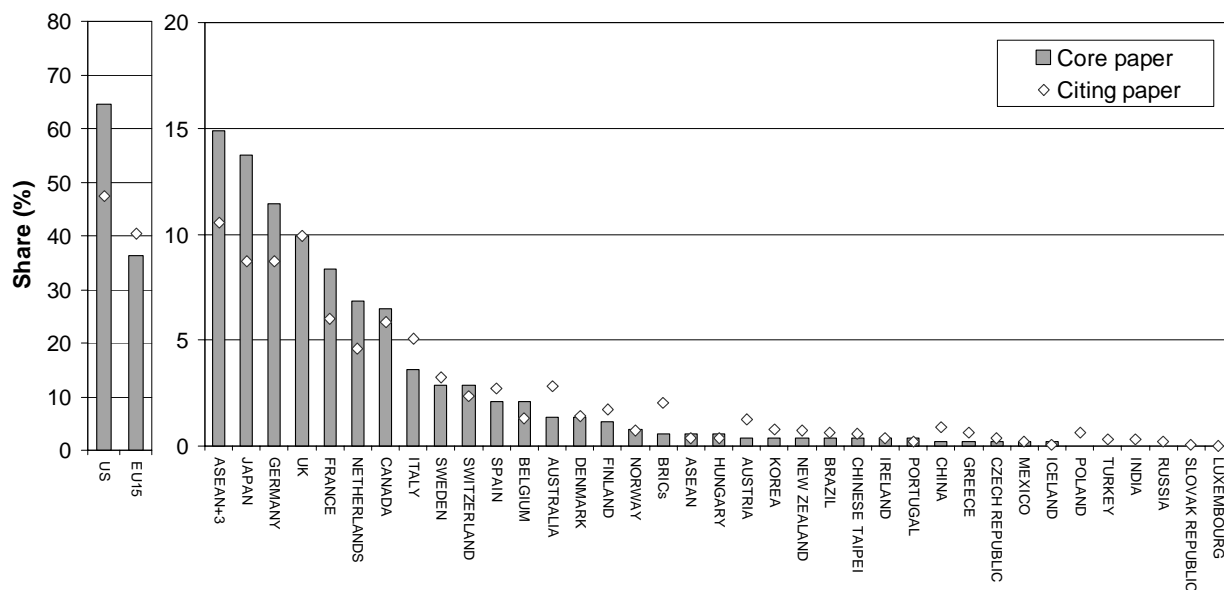
(a) Cancer research



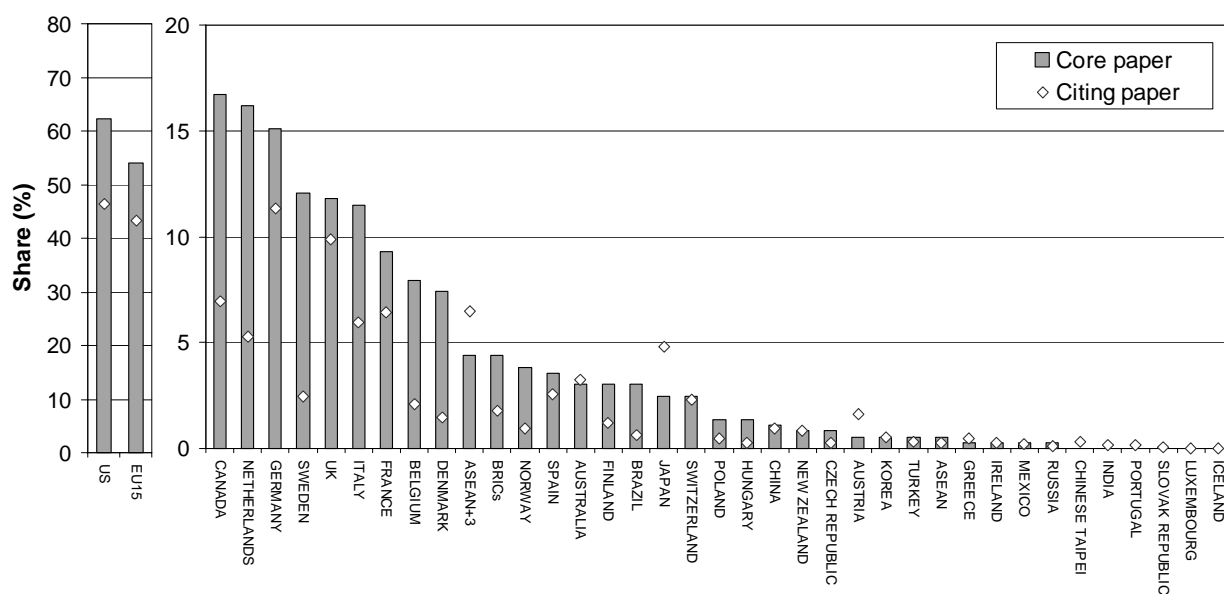
(b) Research on infectious diseases and immunology



(c) Study of obesity



(d) Research on heart and blood vessels



Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

6. Characteristics of international research in citing papers

Indicators of international co-authorship are among measures of cross-border knowledge flows. Here, the ratio of international co-authorship, which shows how knowledge is shared or diffused among countries, is analysed. The ratio of international co-authorship in a country was defined as the ratio of multi-country authored papers to all papers, where both include the target country.

There may be some indications about language barriers and geographical factors in international co-authorship. Nevertheless this kind of exercise would provide important insight into international collaborative work, because English is one of the major international languages commonly used among researchers nowadays. Furthermore it seems to be reasonable to think that physical distance between researchers has some correlation with ratio of co-authorship, though the influence of the development of information and communication technology on knowledge flows is of interest.

6.1 *Nanoscience and material sciences*

As shown in Figure 16, the ratio of international co-authorship varies depending on categories. “Superconducting and quantum computing” has the largest ratio, followed by “Nano materials and devices” and “Chemical Synthesis”. High international co-authorship indicates the existence of a dense knowledge network in “Superconducting and quantum computing”. Asian countries tend to have a smaller ratio of international co-authorship in comparison to European countries, although they have a relatively high share in this research category. This result indicates that the Asian countries compete actively with each other and knowledge flow is likely localised within their borders.

The ratio of international co-authorship in the EU countries is higher than the Asian countries, especially in countries producing a small number of papers. The ratio of international co-authorship between the EU15 and other countries is as low as Asian countries. It seems that the EU countries attain the diversity of researchers due to international co-operation within the EU countries and the knowledge flow seems to be confined to EU countries.

The ratio of international collaboration in research in the United States is also small. This is because the United States can obtain a diversified base of researchers within the country itself due to its large size, as it can be seen in the next section.

6.2 *Particle physics and cosmology*

“Particle physics and cosmology” has the highest ratios of international co-authorship among all categories, showing wide spill over of acquired knowledge over the world. It is quite in contrast to the fact that it has a very weak interaction with other research areas (see Figure 3). Cross-regional co-operation among the United States, the EU15, and the ASEAN+3 also seems to be active. The relatively high share of Eastern European countries presumably shows active international co-operation. Turkey has the smallest ratio of international co-authorship here.

6.3 *Environment*

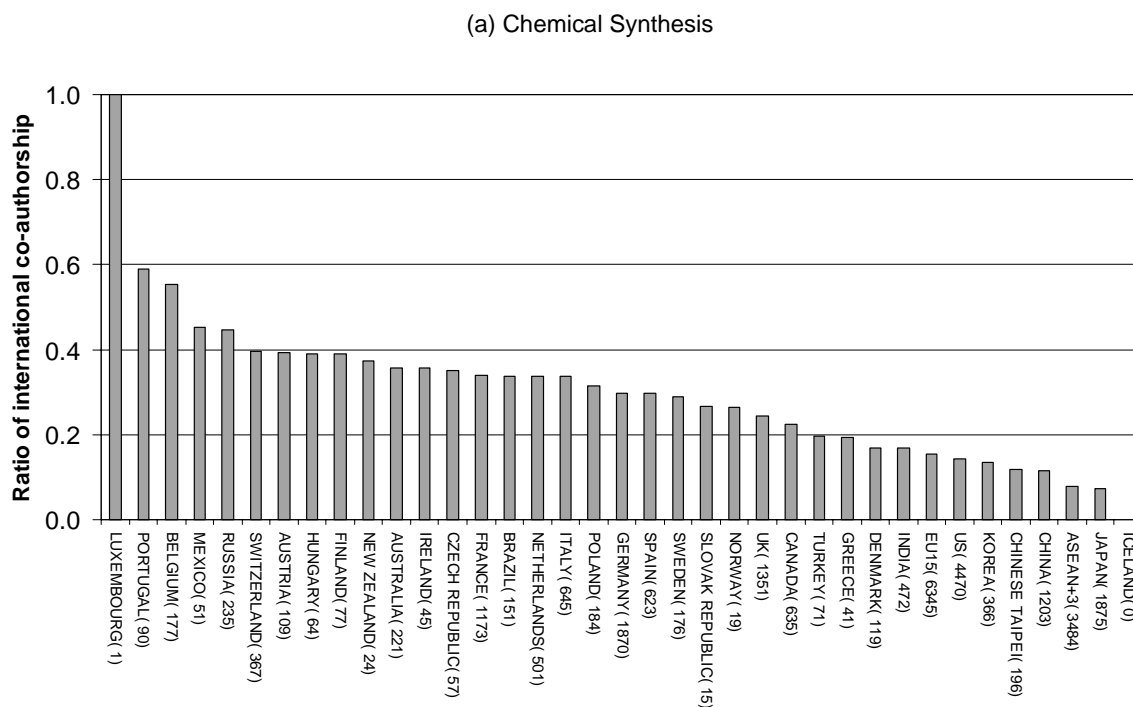
The ratio of international co-authorship in “Environment” is as large as “Superconductivity and quantum computing”. Korea and China conduct more international co-operative research here than other categories. Approximately 60% of papers from Korea and China are multi-country authored. Spain has a relatively small ratio of international co-authorship.

6.4 Bioscience and healthcare

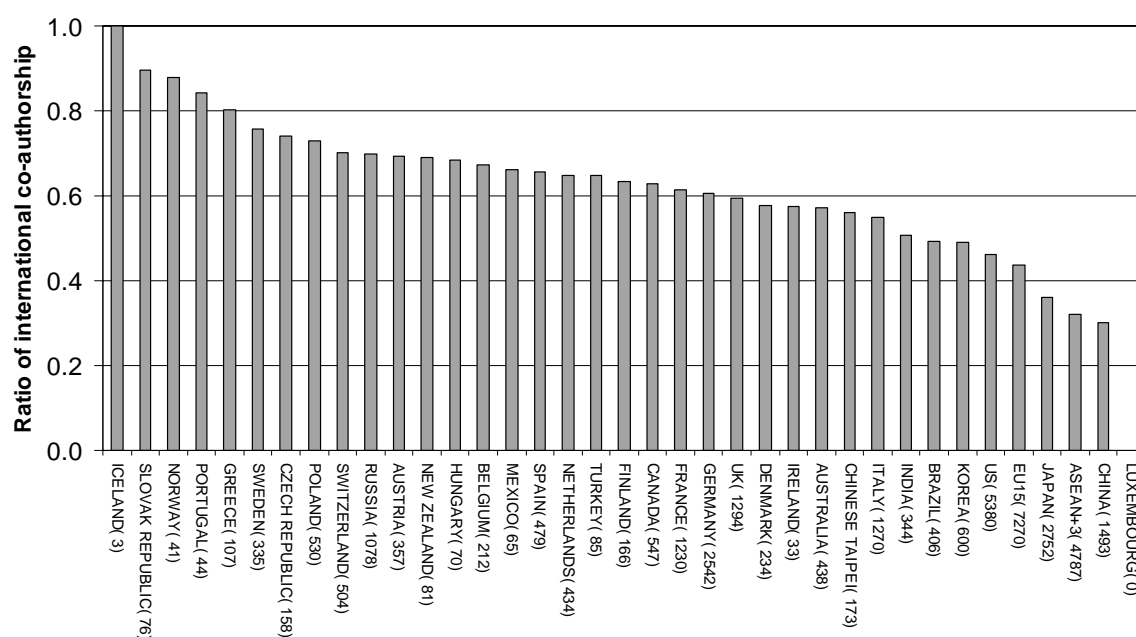
The ratio of international co-authorship in bioscience is larger than in healthcare. Asian countries tend to have a smaller ratio of international co-authorship in comparison to European countries, similar to the tendency observed in “Nanoscience and materials”. Knowledge flows in the Asian countries seem to be confined domestically. China has a slightly larger ratio of international co-authorship in healthcare.

Ratios of international co-authorship in the EU countries are higher than the Asian countries, especially in Eastern European countries. The ratio in the EU15 is as low as in Asian countries, indicating lower interaction with countries outside the EU15. Turkey’s relatively high ratio of international co-authorship in bioscience attributes is due to the high international co-authorship in “Plant science research”.

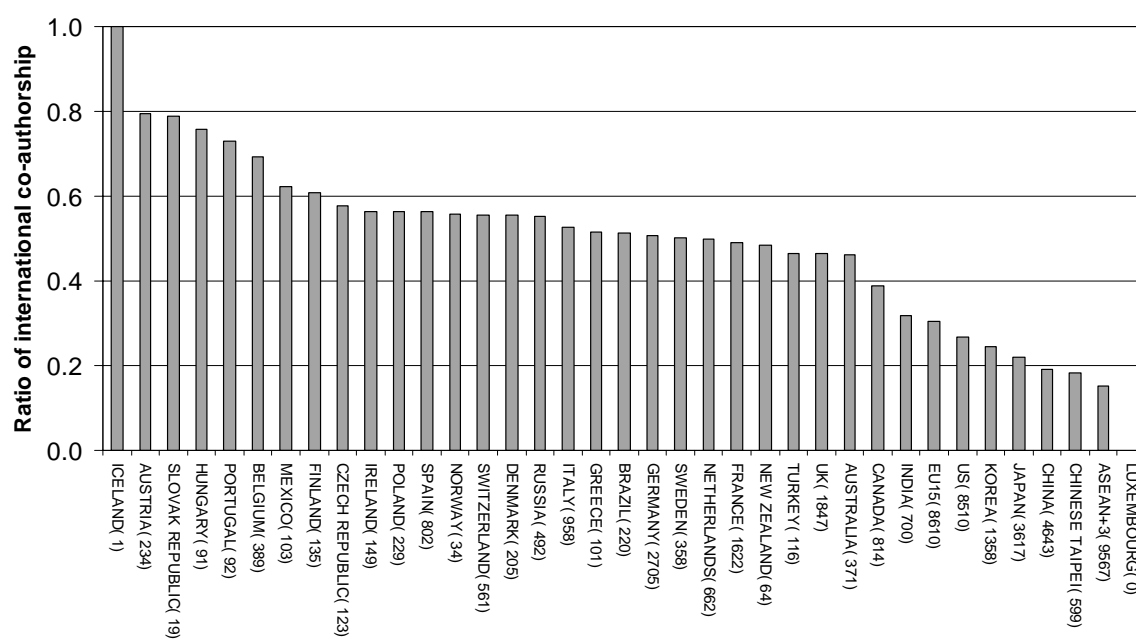
Figure 16. Ratio of international co-authorship in citing papers; (a) Chemical synthesis, (b) Superconductivity and quantum computing, (c) Nano materials and devices, (d) Particle physics and cosmology, (e) Environment, (f) Bioscience, and (g) Healthcare



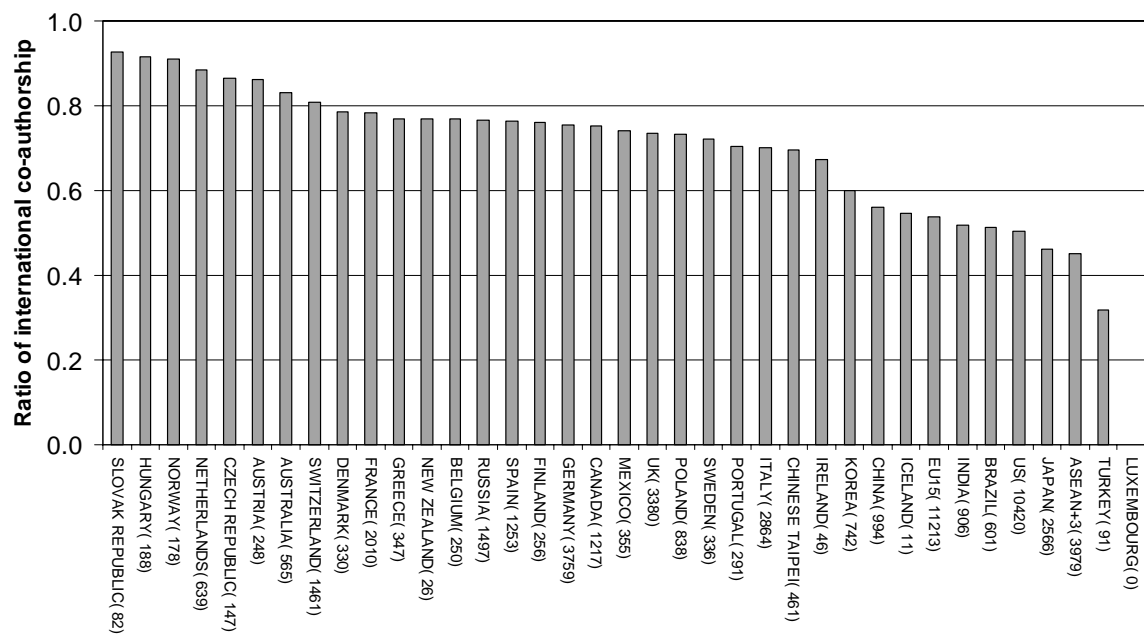
(b) Superconductivity and quantum computing



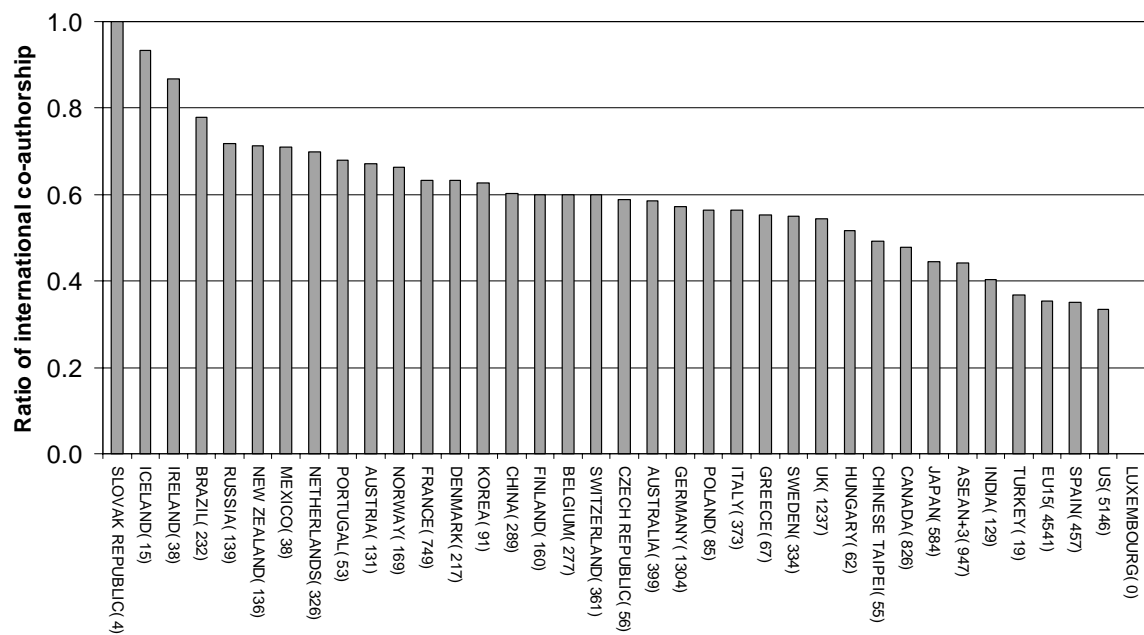
(c) Nano materials and devices



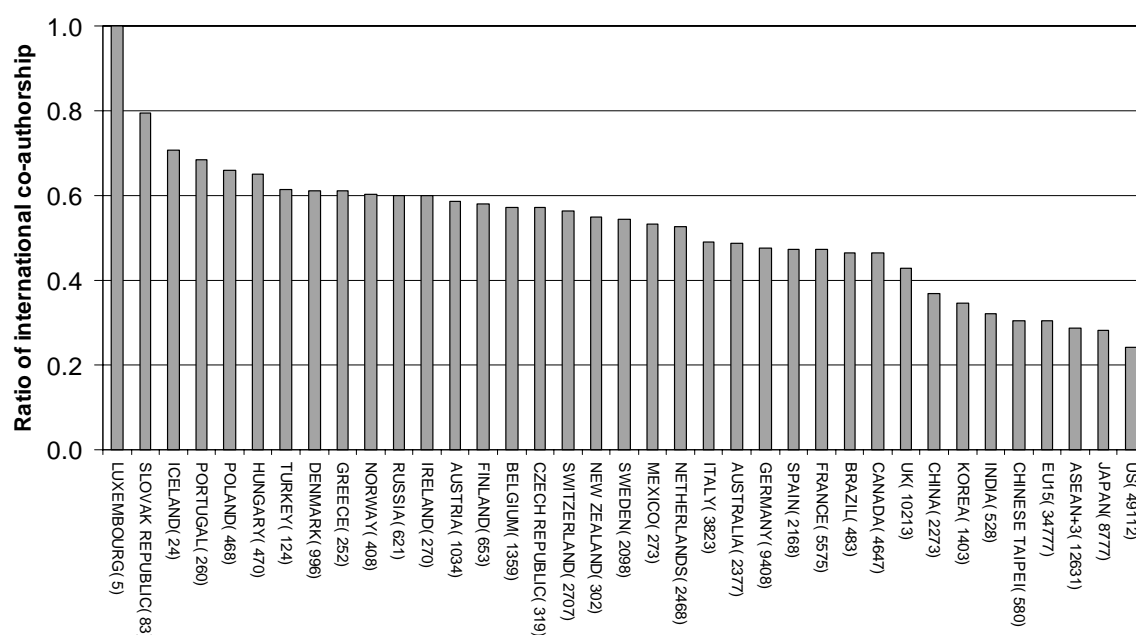
(d) Particle physics and cosmology



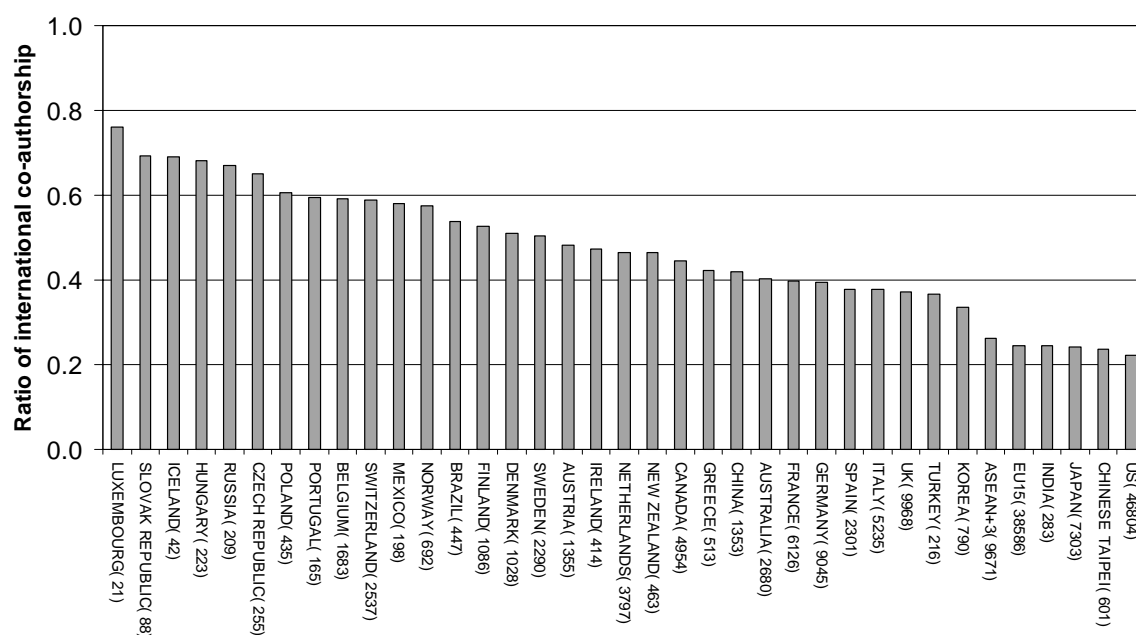
(e) Environment



(f) Bioscience



(g) Healthcare



Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators and Science Citation Index on CD-ROM [1999-2004]; provided by Thomson Scientific; analysed by authors.

7. Characteristics of institutional diversity

Diversity and openness in research networks are likely key factors for countries to be innovative. One of the main objectives of the EU Framework Programme is enhancing co-operation among the EU countries, aiming to establish a European research area (European Commission, 2000). The United States ensures its diversity through co-operative work within its borders. Each country or region tries to enhance or ensure its diversity in research in different ways. Analysis of research networks at the institutional level helps a more precise understanding of knowledge flows.

Methodology which is commonly used in social network analysis was applied to examine co-authorship networks at the institutional level. Social network analysis provides various useful indicators for measuring characteristics of the networks.

Three research categories in “Nanoscience and materials” and “Particle physics and cosmology” were analysed. In order to understand the characteristics of co-authorship networks at the institutional level, three typical indicators, density, degree centrality (Wasserman and Faust, K., 1994; Carrington *et al.*, 2005) and PageRank (Page *et al.*, 1998) in the top 20% institutions in core paper share were analysed. These are useful indicators for measuring the intensity of knowledge flow among institutions and for assessing which institutions play a crucial role in the knowledge flow. An explanation of these indicators is found in Box 2.

Diagrams showing co-authorship network spreads around high PageRank institutions⁸ were also created for providing an intuitive picture of the co-authorship network. In order to create the diagrams, co-authorship linkages within the high PageRank institutions and linkages between the high PageRank institutions and the top 20% institutions in core paper (except for the high PageRank institutions) were analysed. The former represents linkages between core institutions in the co-authorship network and the latter represents linkages between core and periphery institutions. Strength of linkages was measured by the number of co-authorships.

In order to find out the most crucial linkages among complex co-authorship networks, the pathfinder network method (Schvaneveldt *et al.*, 1989) was applied.⁹ This methodology was originally developed in the field of psychology. Since this methodology is suitable to find the most essential cognitive map from complicated relations, there are several applications of this method to co-authorship networks (Chen, 1999; Chen and Paul, 2001). Linkages obtained from the pathfinder network were visualised via the gravity model. An attractive force between a pair of linked institutions was set to a constant value, so that the diagrams purely show the topological structure of the co-authorship network.

Errors in institutions’ names were modified manually. A more sophisticated and systematic methodology for correcting errors in institutions’ names (Wooding *et al.*, 2006) is needed for large-scale analysis, such as analysis of citing papers. It should also be noted that this analysis does not intend to evaluate individual institutions; its purpose is to seek possible application of the social network analysis to the co-authorship network in order to understand characteristics of knowledge creation and flows.

8. The high PageRank institutions are the top 10% of institutions in the PageRank among the top 20% of institutions in core paper share.

9. In this study, two parameters r and q are set to ∞ and $N-1$, respectively, where N is the number of institutions in a co-authorship network. The network obtained is identical to the sum of multiple minimum cost graphs.

Box 2. Density, degree centrality, and PageRank

Density

The density of an institutional network is calculated by the ratio of the number of co-authorships among actors to its maximum possible. The density goes from zero, if there are no co-authorships, to unity, if all possible co-authorships are present. It is a proxy showing how strongly and mutually institutions interact. High density is indicative of a high degree of knowledge flow among actors.

Degree Centrality

The degree centrality is an indicator measuring the prominence of an institution. It is the most simplified version of the centrality; meanwhile there are various expressions of centrality. The degree centrality of an institution is evaluated by counting the number of co-authorship linkages stemming from the institution. The number of linkages is normalised by the total number of institutions in the network. If one institution has a co-authorship relation to all other institutions, the degree density equals unity. Many linkages represent the prestige of those involved.

PageRank

PageRank is the core algorithm in the Google search. Its basic concept comes from an iterative and simple idea; web pages which receive many links from prestigious web pages would also be prestigious. The prestige of a given web page is analysed based on linkage structure, *i.e.* hyperlink structure. In addition to simple counting of the number of received-links, like evaluation of the degree of centrality, the algorithm also weights the links based on the prestige of the linking web pages. PageRank can be calculated using a simple iterative algorithm.

7.1 Overview of institutional networks

Table 3 shows some characteristics of co-authorship networks in four research categories. There are 296 institutions and 179 links in a co-authorship network of “Chemical synthesis”. The density of the network is 0.004. It means each institution is supposed to have approximately 1.2 links to others. Among three categories in “Nanoscience and materials”, “Chemical synthesis” possesses the smallest density. The density in “Superconductivity and quantum computing”, *i.e.* 0.014, is the largest and its average number of links is 5.5. “Nano materials and devices” lies in between them. The tendency is consistent with the results of international co-authorship, discussed in section 6.

The density in “Particle physics and cosmology” is almost four times greater than that in “Superconductivity and quantum computing”. The average link reaches 53.5. This distinctive feature represents the unique characteristics of Big Science (Irvine and Martin, 1985). It often requires state-of-the-art facilities and a huge budget for conducting the research and an immense number of researchers are involved.

Table 3. **Characteristics of co-authorship network in four research categories**

Research category	Institutions	Links	Density	Avg. Link
Chemical synthesis	296	179	0.004	1.2
Superconductivity and quantum computing	406	1121	0.014	5.5
Nano materials and devices	549	948	0.006	3.5
Particle physics and cosmology	931	24914	0.058	53.5

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

7.2 Nanoscience and materials

Co-authorship networks in “Superconductivity and quantum computing” and PageRank, degree centrality, and share in core papers are shown in Figure 17(a) and Table 4(a), respectively. Eighty-one institutions were selected for the analysis. Among them 29 are from the United States. Twenty-seven institutions come from the EU15 and fifteen from the ASEAN+3.

Results obtained by PageRank and degree centrality are well correlated with each other. There is no remarkable deviation between them. It was found, however, that the large share in core papers does not mean high ranking in PageRank and degree centrality. The French Atomic Energy Commission (Commissariat à l’Energie Atomique - CEA, France) has the second largest PageRank, though its share in core papers ranked ninth. There are several institutions which show a similar tendency.

The dense map indicates active and complicated interaction among institutions as shown in Figure 17(a). The result is consistent with the high density of the network. An interesting point is that high density leads to a fusion of co-authorship networks from different region, *i.e.* the United States, the European Union, and Asia (especially Japan). This would be a consequence of intense knowledge flows across borders ever since the discovery of high-temperature superconductivity in 1986. The map clearly has characteristics of core-periphery structures, a typical example can be seen around NIST (National Institute of Standards and Technology, the United States). Institutions with high PageRank form the core of the research network and other institutions exist on the periphery, indicating considerable presence and influence of prestigious institutions in the research network.

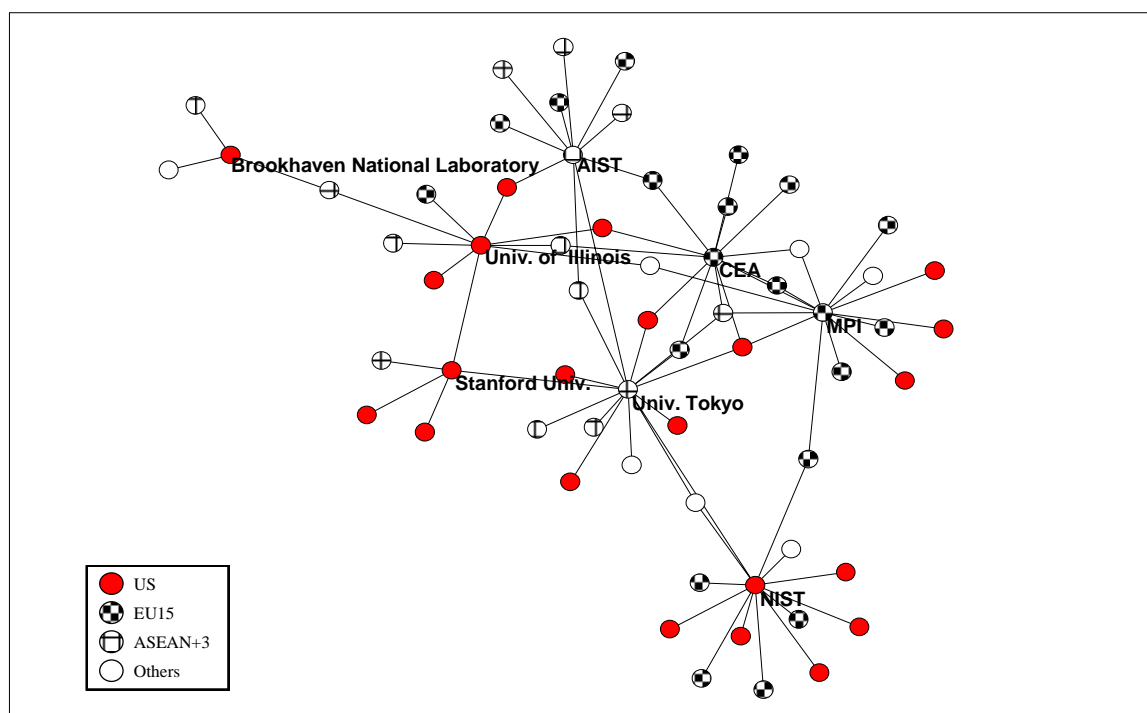
Co-authorship networks in “Nano materials and devices” and PageRank, degree centrality, and share in core papers are shown in Figure 17(b) and Table 4(b), respectively. Deviation between PageRank and core paper shares also exists here. Ninety-eight institutions were selected for the analysis; Forty-eight from the United States and twenty-one institutions from the EU15. There are 21 institutions from the ASEAN+3, among which 6 are from China. This is another clue showing the recent dramatic rise of China in nanoscience. Two Indian institutions and one Russian institution are also in the network.

Characteristic fusion of co-authorship is not observed in “Nano materials and devices”. Each region tends to form its own co-authorship network within the region. Among them the MPI (Max Planck Institute, Germany) acts as a bridge between European and Asian institutions. No distinctive core-periphery structure is observed in the US research network (the right part of the map). Meanwhile, the CNRS (National Centre of Scientific Research, France), and Tohoku University (Japan) play a role in knowledge flow processes in terms of the network structure, although their core paper share is not remarkably high. IBM co-operation which has the tenth largest PageRank, is involved in the network. This is rationale for the existence of knowledge flows involving the private sector.

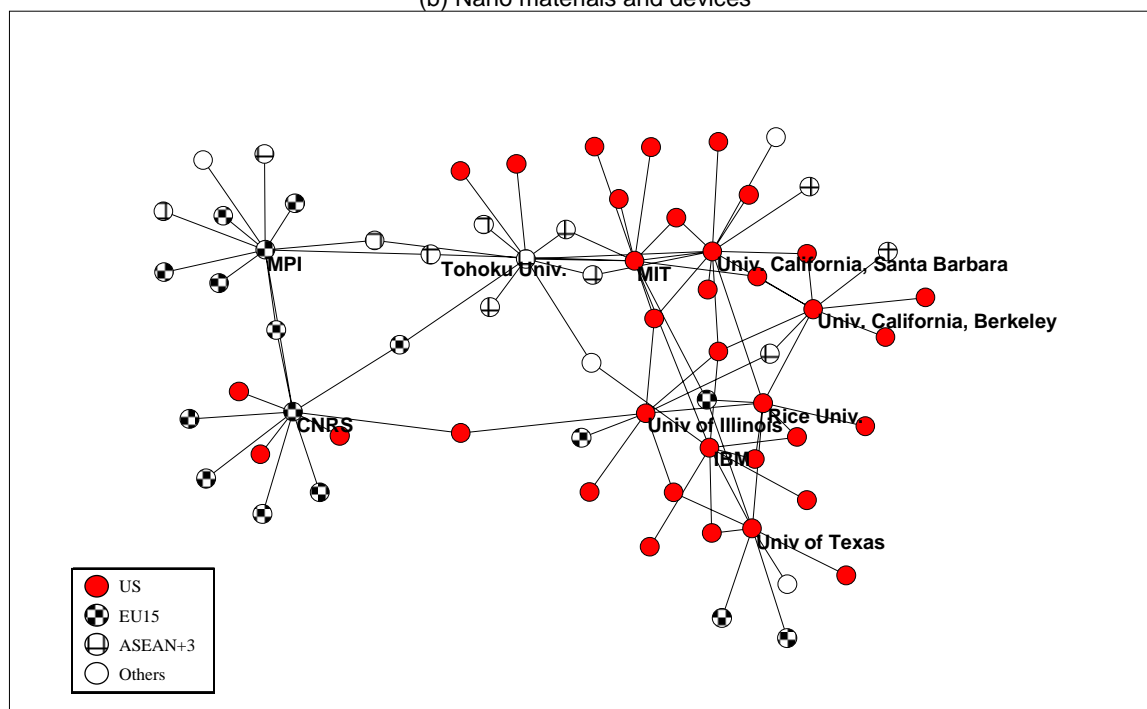
Mapping of co-authorship networks in “Chemical Synthesis” does not work, because the linkages are too sparse to generate the map. This indicates that knowledge of this research category tends to be localised individual institutions or diffuses through another medium, instead of co-authorship.

Figure 17. Co-authorship network in (a) Superconductivity and quantum computing and
(b) Nano materials and devices

(a) Superconductivity and quantum computing



(b) Nano materials and devices



Note: Each circle represents an institution. Co-authorship network spread from the high PageRank institutions is shown. The linkages among institutions were obtained by the pathfinder network method.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

Table 4. **PageRank, degree centrality, and share in core paper.**
(a) Superconductivity and quantum computing and (b) Nano materials and devices

(a) Superconductivity and quantum computing

	PageRank	Degree Centrality	Share (%)
University Tokyo (Japan)	0.034 (1)	0.35 (1)	5.5 (1)
French Atomic Energy Commission (France)	0.033 (2)	0.33 (2)	3.2 (9)
National Institute of Standards and Technology (US)	0.030 (3)	0.29 (3)	5.5 (1)
Max Planck Institute (Germany)	0.030 (4)	0.26 (4)	5.4 (3)
National Institute of Advanced Industrial Science and Technology (Japan)	0.025 (5)	0.25 (5)	2.8 (13)
University of Illinois (US)	0.024 (6)	0.24 (6)	2.6 (14)
Brookhaven National Laboratory (US)	0.023 (7)	0.24 (6)	1.8 (25)
Stanford university (US)	0.022 (8)	0.21 (8)	4.8 (5)
Japan Science and Technology Agency (Japan)	0.021 (9)	0.21 (8)	3.2 (9)
Argonne National Laboratory (US)	0.020 (10)	0.20 (10)	2.6 (14)

(a) Nano materials and devices

	PageRank	Degree Centrality	Share (%)
University of California, Santa Barbara (US)	0.029 (1)	0.14 (1)	2.2 (9)
Max Planck Institute (Germany)	0.026 (2)	0.13 (2)	4.3 (1)
Massachusetts Institute of Technology (US)	0.025 (3)	0.13 (2)	3.0 (2)
National Center for Scientific Research (France)	0.024 (4)	0.11 (5)	1.3 (23)
Tohoku University (Japan)	0.023 (5)	0.12 (4)	1.7 (13)
University of Illinois (US)	0.022 (6)	0.11 (5)	1.2 (26)
Rice University (US)	0.021 (7)	0.11 (5)	2.8 (3)
University of Texas (US)	0.019 (8)	0.10 (8)	2.3 (7)
University of California, Berkeley (US)	0.019 (9)	0.09 (9)	2.4 (5)
IBM (US)	0.018 (10)	0.09 (9)	2.0 (11)

Note: Top 10 institutions in the PageRank are shown in the table.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

7.3 Particle physics and cosmology

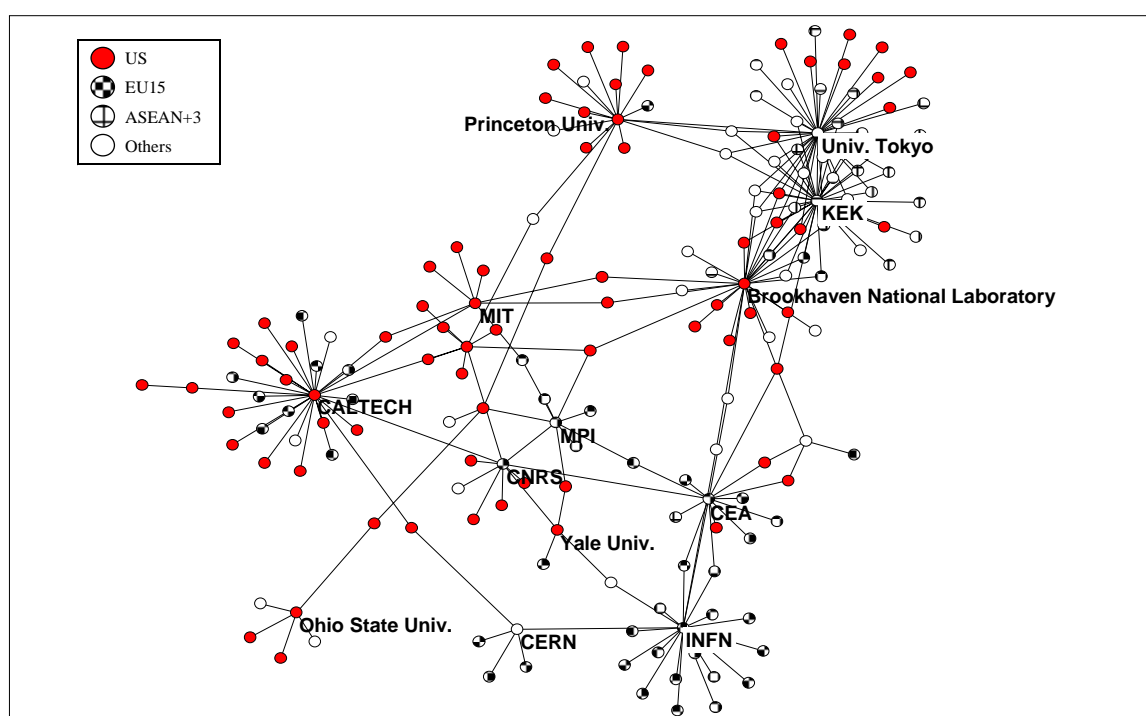
A map of “Particle physics and cosmology” shows an extremely dense co-authorship network – see Figure 18. One-hundred and seventy-eight institutions were selected in order to conduct the analysis. They are the top 20% of institutions in core paper share: 75 from the United States, 48 from EU15 and 26 from ASEAN+3, respectively. Twenty-nine institutes fall into “others”, among which eight institutes from Russia and five from India.

The map clearly shows the existence of two types of institutions which have different functions in the research network. Clear core-periphery research networks are found around Princeton University (United States), the University of Tokyo (Japan), the KEK (High Energy Accelerator Research Organisation, Japan), the Brookhaven National Laboratory, the CALTECH (California Institute of Technology, United

States), the INFN (National Institute for Nuclear Physics, Italy), and the CEA (French Atomic Energy Commission, France). The dense linkages to these institutions show their notably important role in the research network. In contrast, the MIT (Massachusetts Institute of Technology, United States), the CNRS, and the MPI are acting as bridges connecting different research groups. Elimination of these institutions from the network results in weakening linkages among different groups. These institutions likely have an important role in knowledge flows between different groups.

The CERN is located at the edge of the map; nevertheless it has the largest core paper share, *i.e.* 8%. Its degree centrality is 0.63, about three quarters of MIT's. This is probably because of the characteristics of huge international institutions like the CERN. Since the CERN can attract researchers from all over the world and control a huge budget, it probably preserves its diversity and strength without conducting intensive international collaborative research.

Figure 18. Co-authorship network in Particle physics and cosmology



Note: Each circle represents an institution. Co-authorship network spreads from the high PageRank institutions are shown. The linkages among institutions were obtained by the pathfinder network method.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

Table 5. **PageRank, degree centrality, and share in core papers in Particle physics and cosmology**

	PageRank	Degree Centrality	Share (%)
Massachusetts Institute of Technology (US)	0.011 (1)	0.86 (1)	5.3 (9)
University Tokyo (Japan)	0.011 (2)	0.82 (2)	5.3 (9)
National Center for Scientific Research (France)	0.010 (3)	0.75 (3)	6.1 (7)
California Institute of Technology (US)	0.009 (4)	0.73 (4)	7.3 (2)
French Atomic Energy Commission (France)	0.009 (5)	0.72 (5)	4.1 (14)
Max Planck Institute (Germany)	0.009 (6)	0.71 (7)	6.9 (3)
Princeton University (US)	0.009 (7)	0.72 (6)	6.9 (3)
Ohio State University (US)	0.009 (8)	0.71 (7)	3.5 (22)
Yale University (US)	0.009 (9)	0.71 (9)	2.6 (38)
National Institute for Nuclear Physics (Italy)	0.009 (10)	0.67 (11)	6.4 (5)

Note: Top 10 institutions in the PageRank are shown.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

8. Conclusion and discussion

This document presents a bibliometric method that can be used to analyse the changing nature of science. The analysis shows that science is widening its horizons not only across disciplines but also across countries or regions.

8.1 *Perpetual evolution of scientific knowledge*

The analysis clearly shows that new scientific research, such as “Nano materials and devices”, “Environment”, and “Genomics”, are more multi-disciplinary in comparison to traditional research such as “Chemical synthesis” and “Particle physics and cosmology”. In these multi-disciplinary research subjects, scientific papers, which constitute research areas, are composed of a combination of journals from various traditional fields. It could be said that these areas of research are developed via interaction among researchers who used to work on traditional fields and are acting as catalysts leading to co-operative work. Of course, it should be kept in mind that this kind of exercise strongly depends on the choice of initial fields. Even though it could be said that “Nano materials and devices” still has a multi-disciplinary character, because of its location in the map. It is located in between research areas related to physics and chemistry. Since the mapping only relies on co-citation, it is a proxy showing interaction between physicists and chemists.

A precursor of the emergence of nano-bioscience is also observed in the map. The research area is “Research on microanalysis of biochemical substances”. Furthermore, a small number of papers in “Nanoscience and materials” come from biology & biochemistry. Application of DNA molecules to nano-wires and synthesis of DNA-metal particles hybrid materials are examples of research topics developing involving biology & biochemistry. Research on nano-bioscience has not advanced enough matured to be observed in the current stage, but it would be established as a single independent research topic like “Nano materials and devices” in the future. The United States, the European Union, and Asian countries seem to have different tactics to tackle this emerging research, because they have their own characteristics in the portfolio of scientific research (see Appendix A). The difference in tactics would have some influence on future competition between the United States, the European Union and Asian countries. Nano-bioscience is a good example to observe the changing nature of science involving different fields and countries and regions.

It is pointed out that chemistry departments in UK Universities have been facing closure or merger with other departments in recent years (*Nature materials* editorial, 2006). The position of “Chemical synthesis” in the map likely reflects such a tendency in current scientific activities. “Chemical synthesis” interacts strongly with “Nano materials and devices” and also weakly interacts with “Biosciences”. It could be said that continuous interaction between chemistry and other disciplines has been blurring the border of chemistry. This could be attributed to the universal character of chemistry. Since chemistry is the science of matter in the atomic to molecular scale, its knowledge can disseminate widely beyond the disciplines and be embedded in newly emerging research. For example, chemists could exclusively enjoy the beauty of Fullerenes (Kroto *et al.*, 1985), spherical molecules made by carbon atoms, in the mid-1980s, but they are a key material in nanoscience and nanotechnology nowadays and fierce competition in research and development is taking place all over the world. Blurred borders presumably also indicate spill-over of human resources from chemistry to newly emerging research. Systematic re-organisation of knowledge in chemistry, in line with the evolution of science, and attracting talented human resources seem to be a necessity for chemistry to maintain its position in an evolving scientific-world.

“Bioscience” and “Healthcare” have the largest share in 133 research areas, indicating intense dedication to such research, though characteristics of the database should be taken account. It was also found that non-life sciences research plays a crucial role in the development of “Genomics”. “Plant science research” is a research area whose knowledge is expected to widely spread out to other fields. Since Asian countries have a comparably large share in it: it could be said that the application of the accumulated knowledge in “Plant science research” to other research is a possible solution to enable them to expand their knowledge into other bioscience-related research.

“Particle physics and cosmology” has a substantial share in scientific activities. Commitment to the international project seems to contribute to an increasing presence of Eastern European and Asian countries. Its dense knowledge network indicates active knowledge flows among institutions across borders, meanwhile huge international institutions, such as the CERN, can assure its diversity within a single institution. This is an intriguing characteristic of Big Science. Weak interaction between “Particle physics and cosmology” and other research results in its rather isolated position on the map. Effective management of talented researchers, for example their utilisation in other research, would be promising to enhance countries’ potential in scientific and technology activities, eventually in innovative activities. For example, eight out of ten research areas in which India has a large share in citing papers are related to “Particle physics and cosmology”.

Although this study mainly focused on natural sciences, some research areas related to social sciences such as “Research on intellectual property right problems” and “Study on local economy” were identified. It is worth mentioning that two research areas, “Neuroscience and cognitive-psychology” and “Research on visual stimulation and learning”, were observed in an intermediate location between bioscience and social sciences. Collaboration of researchers from bioscience and social sciences is beneficial to the development of these research areas.

8.2 Characteristics of knowledge creation and knowledge flow by countries or regions

The prominent presence of the United States in science is documented throughout this study. The largest share in almost all research areas is evident, though one should be aware of the bias of the database (skewed towards American scientific literature). The US dominance is particularly remarkable in “Bioscience” and “Healthcare” and would partly result from the budget doubling campaign in the National Institutes of Health. Meanwhile, its advantage in “Nanoscience and materials” is not so remarkable because of the large presence of other actors. The analysis of international co-authorship and characteristics of institutional diversity revealed that the United States ensures its diversity thorough

domestic co-operative work. Securing or attracting excellent human resources within the country will keep playing a crucial role for retaining its advantage, especially in research areas related to non-life sciences.

The share of individual European countries is not as large as the United States' share. An accumulated share of EU15 is, however, comparable to the United States. Furthermore the portfolio of the EU15 is almost identical to the world average (see Appendix A), though each country has its own characteristics in the portfolio. This demonstrates that EU countries compensate each other in terms of the portfolio (NISTEP, 2005b). A high ratio of international co-authorship shows active knowledge exchange within the EU. This may be the positive results of continuous promotion of collaborated projects in Framework Programs. The knowledge flows seem to be localised within EU countries. Establishment of eminent centre of excellence (COE) also means a concentration of cutting-edge knowledge in specific countries. If the amount of knowledge directly limits the ability to innovate, it would be indicative of the localisation of its benefit within specific countries. This indicates the importance of attaining COE in individual countries, as well as preserving the well-balanced portfolio and accumulated knowledge in the EU.

The BRICs are now catching up rapidly. Their rise is especially remarkable in "Nanoscience and materials". This is a typical example showing how science is evolving not only across disciplines but also across countries or regions. Large gaps in shares between core and citing papers likely represents a lack of leading researchers, however. Since developed countries in Asia, *i.e.* Japan and Korea, also have their strength in the same research areas, the portfolio of Asian countries is distorted towards non-life sciences. Furthermore, a small ratio of international co-authorship indicates localisation of acquired knowledge in individual countries. This is quite a contrast to political endeavours to establish research networks in Europe. Enhancing international co-operative work would bring a higher presence of Asian countries in the knowledge network. To increase the ratio of life sciences in portfolios seems to be also important. Since they are one area of research considered to have large economic and social impacts in the future.

Finally, social network analysis casts light on how international co-authorship networks have been formed. The co-authorship networks show quite different structures depending on research. It was found that active knowledge exchange among institutions in some cases resulted in a fusion of co-authorship networks beyond geographical factors. A typical example is the co-authorship network in "Superconductivity and quantum computing". Such a network likely contributes to rapid global flows of knowledge. Meanwhile, it was also found that certain research had a sparse co-authorship network.

It was revealed that some institutions play a significant role in the network, though their shares in core papers are not large. This suggests that individual institutions have their own functions in the knowledge network. In recent years, the significance of steering and funding of research institutions has been receiving wide recognition. Financing excellence in public research is one of the crucial issues, because public research funding is considered to underpin advances in science and innovation (Mansfield and Lee, 1996; Salter and Martin, 2001; Feller *et al.*, 2002). Increasing demands for accountability require decision makers to assess outcomes and impacts of science and technology policy. The analysis shown here would contribute to assess functions of institutions in the research network. For example, it could be said that COEs should be selected from institutions which are in core positions rather than peripheral positions, if the promotion of COEs aims at extensive dissemination of knowledge. It makes for a much easier knowledge flow. Longitudinal observation would provide policy makers or administrative offices of institutions with information on how the function of the institution has been changing over time.

8.3 Future work

The changing nature of science was analysed based on co-citation analysis of scientific papers. The results presented here rationalised validity of the methodology; however, there are limitations and future work will enhance the understanding of evolution in science.

In this study only research areas which exceed certain thresholds in size and co-citation linkages were extracted and analysed. Because of these restrictions, some research areas, for example research areas related to mathematics, were not captured even if they played a crucial role in science. Adjustment of thresholds is required depending on the purpose of analysis. For example, the utilisation of different thresholds on the fields will be promising to obtain research areas related to mathematics. Furthermore, lower threshold in minimum cluster size and measurement of growth rate of the clusters will provide a potential list of research areas which are about to emerge (NISTEP, 2005a).

Longitudinal analysis will be promising, too. It would reveal how science has been changing its shape over time. The map of science obtained in this study will be a starting point for the longitudinal analysis. Evolution of nano-bioscience – a precursor of its emergence was observed in this study – is an issue of interest. The one common argument on modern science is that it is becoming increasingly inter-/multi-disciplinary. The longitudinal analysis will also reveal if the inter-/multi-disciplinary character of research areas obtained in this study is either increasing or decreasing over time.

Utilisation of scientific literature in analysis limits holistic understanding of science and technology activities. In some research, for example information and communication technology, their outputs or outcomes are not commonly published as scientific literature. Application of the methodology presented here to patents or other available data will shed light on the knowledge structure in another dimension (von Wartburg *et al.*, 2005; Mika *et al.*, 2006). Measurement of the linkage between scientific literature and patents will enable us to trace how scientific discovery ultimately blooms and impacts on the society.

Further development of social network analysis seems to be worth pursuing. One question that came up from this research is how mutual interactions among institutions change over time. Since research areas can maintain activity through continuous interaction among researchers or institutions, the analysis will cast light on evolution of research areas from an angle of research networks. It also may be promising to comprehend functions of institutions embedded in research networks.

APPENDIX A ANALYSIS OF CORE PAPERS

The ESI database was used as the information source for this study. The database includes 5 350 research fronts and these research fronts contain 21 411 core papers in all. In this appendix, characteristics of these core papers are briefly reviewed.

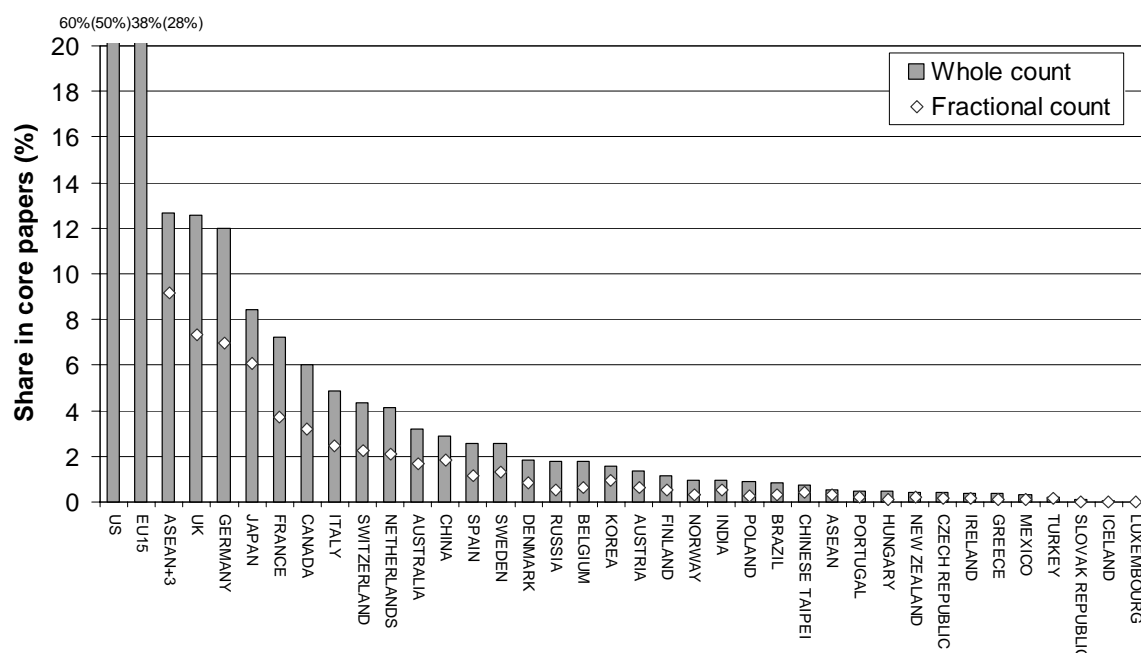
A.1 Countries' share in all core papers

Figure A-1 shows countries' share in all core papers. The United States has the largest share, followed by the EU15 and the ASEAN+3. Note that the share is calculated by the whole count method, thus the sum of shares in all countries exceeds 100%.

The results of the fractional count method are also shown, in Figure A-1. All shares are smaller than those in the whole count method. It seems that the reduction of the share in European countries is larger than in Asian countries such as Japan, China, and Korea.

About 94% of core papers come from the OECD countries in the fractional count method. This is one piece of evidence showing the OECD countries are major players in scientific research.

Figure A-1. Countries' share in all core papers



Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

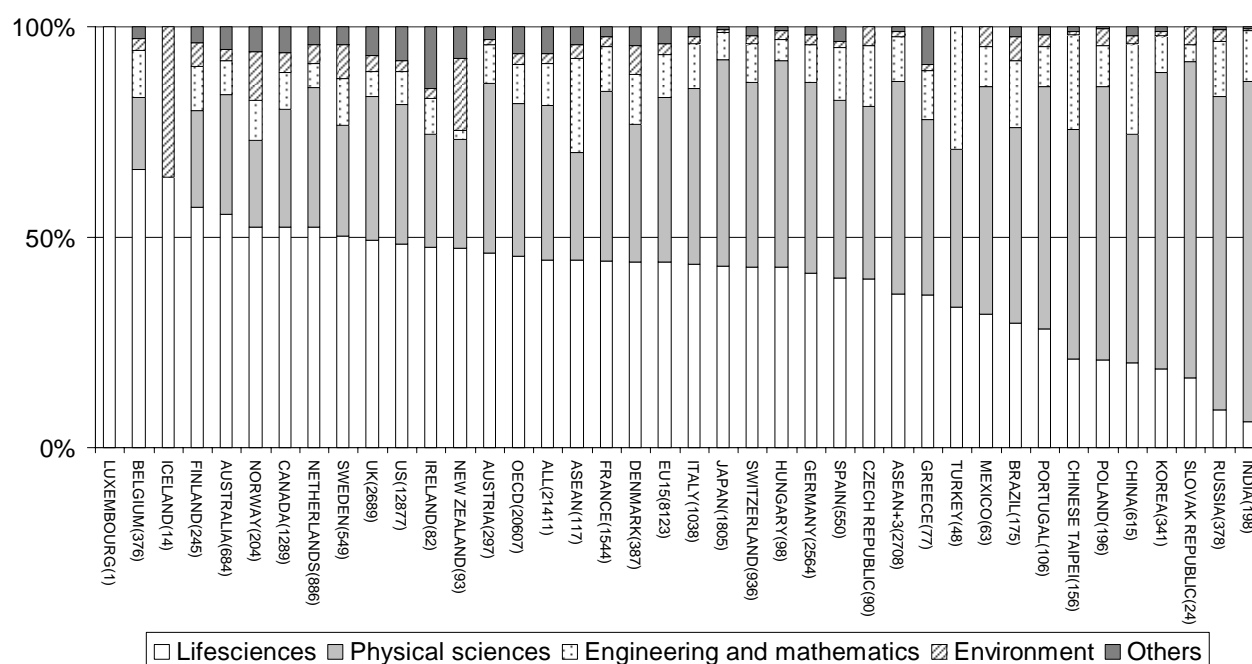
A.2 Characteristic of distribution of core papers by field of science

Core papers in the ESI database are classified into 22 fields based on the journal. Further classification was conducted here in order to examine shares of core papers between life sciences, physical sciences, engineering and mathematics, environment, and others.¹⁰ Figure A-2 shows distribution of core papers by fields of science in the OECD countries, the BRICs, the ASEAN+3, and Chinese Taipei. Note that several countries have few core papers, thus this data is a proxy showing the characteristics of scientific activities in terms of a macroscopic view.

Life sciences account for almost 50% of total core papers in the OECD portfolio. Share of physical sciences is rather high in Eastern Europe and Asian countries. Life sciences have large shares in Nordic countries. This result is in line with another macro-level analysis (OECD, 2005).

The BRICs tend to focus on non-life sciences. Portfolio in the BRICs varies in countries, though the number of core papers in physical sciences is the largest in all countries. Physics dominates portfolios in Russia and India. The share of engineering and mathematics is remarkably higher than the world average in China. Brazil has a rather high share of environment in the portfolio.

Figure A-2. Characteristic of distribution of core papers by fields of science



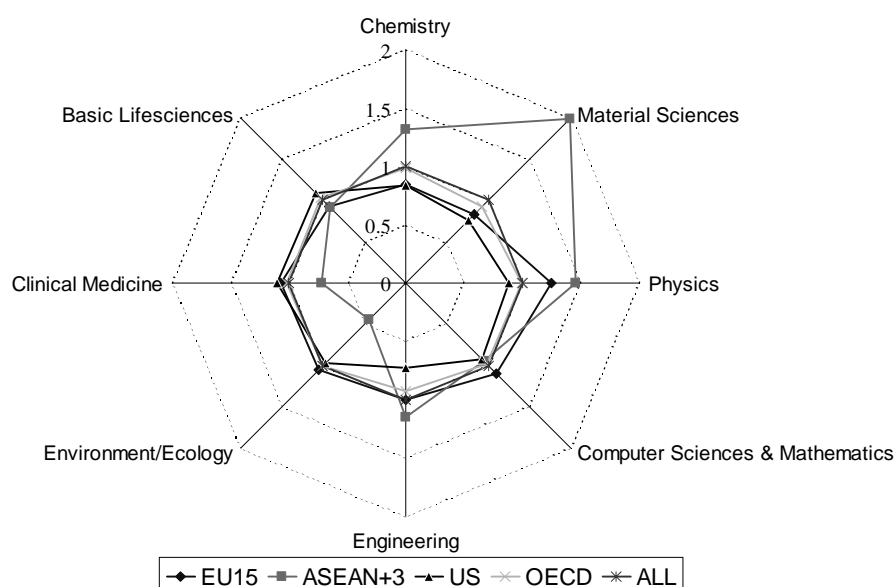
Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

10. Lifesciences include: Agricultural Sciences, Biology & Biochemistry, Clinical Medicine, Immunology, Microbiology, Molecular Biology & Genetics, Neuroscience & Behaviour, Pharmacology & Toxicology, and Plant & Animal Science. Physical sciences include Chemistry, Geosciences, Materials science, Physics, and Space Science. Engineering and mathematics include Engineering, Computer Sciences, and Mathematics. Environment includes Environment/Ecology. "Others" include Economics & Business, Multidisciplinary, Psychiatry/Psychology, and Social Sciences, general.

Figure A-3 shows the portfolio of the EU15, the ASEAN+3, the United States, and the OECD countries. In this analysis, the ESI's 22 fields were reclassified into 8 groups, *i.e.* chemistry, materials science, physics, computer science and mathematics, engineering, environment/ecology, clinical medicine, and basic life sciences. Basic life sciences include eight ESI fields.¹¹ Six ESI fields were omitted.¹²

The share of engineering is relatively smaller than the world average in the US portfolio. The EU15 has an almost circular portfolio, though each country has its own portfolio. The EU15 countries compensate each other in terms of portfolio. The portfolio of the ASEAN+3 distorts toward material sciences, chemistry, and physics.

Figure A-3. **Portfolio of the EU15, the ASEAN+3, the US, and the OECD countries**



Note: Paper counts are based on whole counts.

Source: Research Front dataset [1999-2004] extracted from Essential Science Indicators; provided by Thomson Scientific; analysed by authors.

-
11. These are Agricultural Sciences, Biology & Biochemistry, Immunology, Microbiology, Molecular Biology & Genetics, Neuroscience & Behaviour, Pharmacology & Toxicology, and Plant & Animal Science.
 12. These are Economics & Business, Geosciences, Multidisciplinary, Psychiatry/Psychology, Social Sciences, general, and Space Science.

BIBLIOGRAPHY

- Albert, R., Jeong, H., and Barabasi, A. (2000), "Diameter of the World-Wide Web", *Nature*, 406, 378-382.
- Börner, K., Chen, C., and Boyack, K. W. (2003), "Visualizing Knowledge Domains", *Annual Review of Information Science and Technology*, 37, 179-255.
- Boyack, K. W. (2004), "Mapping Knowledge Domains: Characterizing PNAS", *PNAS*, 101, 1, 5192-5199.
- Chen, C. (1999), "Visualising Semantic Spaces and Author Co-citation Networks in Digital Libraries", *Information Processing & Management*, 35, 401-420.
- Chen, C. and Paul, R.J. (2001), "Visualising a Knowledge Domain's Intellectual Structure", *Computer*, 34, 3, 65-71.
- Carrington, P. J., Scott, J., and Wasserman, S. (2005), *Models and Methods in Social Network Analysis*, Cambridge University Press.
- European Commission (2000), "Towards a European Research Area", COM(2000)6.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998), "Potent and Specific Genetic Interference by Double-stranded RNA in *Caenorhabditis Elegans*", *Nature*, 391, 806-811.
- Feller, I., Alies, C.P., and Roessner, J.D. (2002), "Impacts of Research Universities on Technological Innovation in Industry: Evidence from Engineering Research Centers", *Research Policy*, 31, 457-474.
- Garfield, E., Sher, I.H., and Torpie, R.J. (1964), "The Use of Citation Data in Writing the History of Science", Institute for Scientific Information, Philadelphia.
- Gaufriau, M. and Larsen, P. J. (2005), "Counting Methods are Decisive for Rankings Based on Publication and Citation Studies", *Scientometrics*, 64, 1, 85-93.
- Irvine, J. and Martin, B.R. (1985), "Evaluating Big Science: CERN's Past Performance and Future Prospects", *Scientometrics*, 7, 3-6, 281-308.
- Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. (1985), *Nature*, 318, 162-163.
- Kostoff, R. N. (2004), "The (Scientific) Wealth of Nations", *The Scientist*, 18, 18, 10.
- Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H. (2005) "Co-authorship Networks in the Digital Library Research Community", *Information Processing and Management*, 41, 6 (Dec. 2005), 1462-1480.
- National Institute of Science and Technology Policy (2005a), "The 8th Science and Technology Foresight Study on Rapidly-developing Research Area-", NISTEP Report No. 94 (in Japanese).

- National Institute of Science and Technology Policy (2005b), “Comprehensive Analysis of Science and Technology Benchmarking and Foresight”, NISTEP Report No. 99.
- National Institute of Science and Technology Policy (2007), “Bibliometric Analysis on Recent Trends in Scientific Research”, *forthcoming* (in Japanese).
- Nature materials editorial (2006), “Where Have all the Chemists Gone?”, *Nature materials*, 5, 671.
- Mane, K. K. and Börner, K. (2004), “Mapping Topics and Topic Bursts in PNAS”, *PNAS*, 101, 1, 5287–5290.
- Mansfield, E. and Lee, J.Y. (1996), “The Modern University: Contributor to Industrial Innovation and Recipient of Industrial R&D Support”, *Research Policy*, 25, 1047-1058.
- Mika, P., Elfring, T., and Groenewegen, P. (2006), “Applications of Semantic Technology for Social Network Analysis in the Sciences”, *Scientometrics*, 68, 1, 3-27.
- OECD (2005), *Science, Technology and Industry Scoreboard 2005*, OECD, Paris.
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1998), “The PageRank Citation Ranking: Bringing Order to the Web”, Proceedings of the Seventh International Conference on World Wide Web 7, 107-117.
- Rousseau, R. (1992), “Breakdown of the Robustness Property of Lotka’s Law: The Case of Adjusted Counts for Multiauthorship Attribution”, *Journal of the American Society for Information Science*, 43, 10, 645-647.
- Schvaneveldt, R.W., Durso, F.T., and Dearhold, D.W. (1989), “Network Structures in Proximity Data”, *The Psychology of Learning and Motivation*, 24, 249-284.
- Shannon, C. E. (1948), “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, 27, 379-423, 623-656.
- Shiffrin, R.M. and Börner, K. (2004), “Mapping Knowledge Domains”, *PNAS*, 101, 1, 5183-5185.
- Salter, A.J. and Martin, B.R. (2001), “The Economic Benefits of Publicly Funded Basic Research: a Critical Review”, *Research Policy*, 30, 509-532.
- Small, H. and Sweeney, E. (1985a), “Clustering the Science Citation Index using Co-citations. I. A Comparison of Methods”, *Scientometrics*, 7, 3-6, 391-409.
- Small, H., Sweeney, E., and Greenlee, E. (1985b), “Clustering the Science Citation Index using Co-citations. II. Mapping Science”, *Scientometrics*, 8, 5-6, 321-340.
- Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. (1982), “A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters”, *Journal of Chemical Physics*, 76, 637–649.
- von Wartburg, I., Teichert, T., and Rost, K. (2005), “Inventive Progress Measured by Multi-Stage Patent Citation Analysis”, *Research Policy*, 34, 1591-1607.
- Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

- Watson, J. D. and Crick, F. (1953), "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid", *Nature*, 171, 737-738.
- Wooding, S., Wilcox-Jay, K., Lewison, G., and Grant, J. (2006), "Co-author Inclusion: A Novel Recursive Algorithmic Method for Dealing with Homonyms in Bibliometric Analysis", *Scientometrics*, 66, 1, 11-21.
- Zhou, P. and Leydesdorff, L. (2006), "The Emergence of China as a Leading Nation in Science", *Research Policy*, 35, 83-104.