# CONSTRUCTION OF TYPOLOGY OF SUB-DISCIPLINES BASED ON KNOWLEDGE INTEGRATION

Qiuju Zhou  Fuhai Leng

*zhouqj@mail.las.ac.cn, lengfh@mail.las.ac.cn*
National Science Library, Chinese Academy of Sciences, Beijing, 100190, People's Republic of China

## Abstract
This investigation recalled the frameworks proposed by Stirling (2007), Rafols & Meyer (2010), Liu et al., (2012) and Zhou et al., (2012). The bibliometric methodology presented here provides an overview of scientific sub-disciplines, with special attention to their interrelation. This work aims to establish a tentative typology of disciplines and research areas according to their degree of knowledge integration. Knowledge integration is measured through diversity based on the Subject Categories mapping from the references of the articles set in sub-disciplines. The similarity-weighted cosine was used to measure the interrelations between sub-disciplines.

## Conference Topic
Visualisation and Science Mapping: Tools, Methods and Applications (Topic 8)

## Introduction
In a recent article, Rafols and Meyer (2010) presented an analytic framework for the study of interdisciplinarity. And the framework was enriched by Rafols et al. (2012), Wagner et al. (2011) and Leydesdorff and Rafols (2011). The two main factors of this framework are diversity and coherence. Zhou et al. (2012) – inspired by Rafols and Meyer (2010), Stirling (2007) and related ecological research (Nei & Li 1979; Shriver et al. 1995) on the relationship of diversity within populations and the similarity between populations – proposed a generalize framework to study systems' diversity and the similarity (homogeneity) of systems. Zhou et al. (2012) then applied it to the research profile of countries to present the unbalanced and concentrated disciplinary structure of 32 countries.
Liu et al. (2012) synthesized the main points of the Rafols-Meyer approach, and showed how these ideas can be applied to knowledge diffusion and knowledge integration.
In this article, we want to apply the framework proposed by Zhou et al. (2012) to knowledge integration, following Rafols and Meyer (2010), Leydesdorff and Rafols (2011) and Liu et al. (2012). We further aim to analyze the sub-disciplinary structures based on knowledge integration. For case studies we use articles from the various sub-disciplines of ecology and analyze their knowledge integration, as revealed through their references.

The objectives of this paper are threefold.
1) Give an overview of the three analytical frameworks.
2) Find the theoretical foundations for using knowledge integration to explain the framework proposed by Zhou et al. (2012) in the construction of a disciplines structure.
3) Apply the framework proposed by Zhou et al. (2012) to analyze the knowledge integration of selected sub-disciplines in ecology and use this to evaluate the homogeneity of sub-disciplinary structure.
The article is organized as follows. In Section 2 we recall the main points from the three frameworks proposed by Rafols and Meyer (2010), Liu et al. (2012) and Zhou et al. (2012). Section 3 gives the details for a study of knowledge integration within the framework proposed by Zhou et al. (2012). Section 4 provides a case study related to sub-disciplinary structures based on knowledge integration and discusses the results. Section 5 concludes by summarizing the results and discussing their implications and limitations within current research and also discusses issues for further research.

**Overview of the three analytical frameworks**

In this section we mainly describe the work by Stirling (2007), Rafols and Meyer (2010), Zhou et al. (2012), and Liu et al. (2012).

*Overview of the framework for the study of diversity and coherence proposed by Rafols and Meyer (2010) and Liu et al.（2012）*

Stirling (2007) proposed a framework of diversity for understanding any system of science and technology. Rafols and Meyer (2010) further developed this by proposing a framework for understanding interdisciplinary through diversity and coherence. Their understanding of diversity was as a measure of the variety of categories used, while coherence explains the interrelatedness of categories and topics. Rafols and Meyer (2010) first applied this framework to a single article. Then, Rafols et al. (2012) applied the concepts to whole groups of related articles, using as case studies an entire university's and a department of a university's output (Rafols et al., 2012). In Rafols and Meyer (2010) diversity measures are based on the JCR categories of the references-of-references, while coherence is understood through the strength of bibliographic coupling in the network of references.
Figure 1 show the Conceptualisation of interdisciplinarity in terms on knowledge integration.

Liu et al. (2012) introduce a general framework for the analysis of knowledge integration and diffusion using bibliometric data. They considered a framework that consists of three entities: (1) the source; (2) the intermediary set (IM) derived from the source; and (3) a target set.
The specific operationalization of diversity and coherence may differ in these empirical studies (due to their different goals, focus, sample size, etc.), but having

a conceptually well-defined framework is important for the sake of clarity and in order to be able to compare cases.

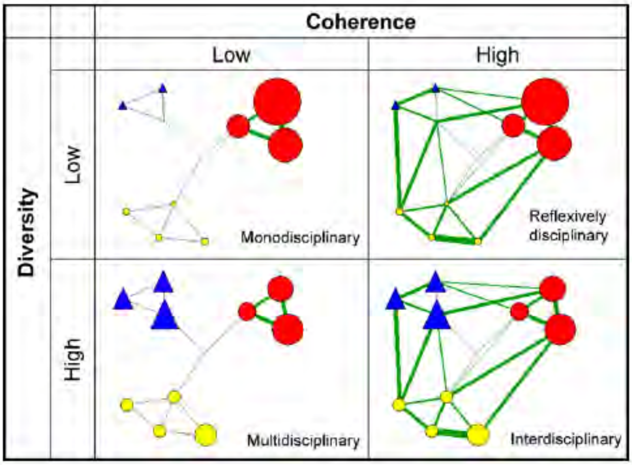


**Figure 1. Overview of the proposed framework for the study of diversity and coherence**

*Overview of the framework for the study of diversity and similarity proposed by Zhou et al. (2012)*

Recalling the framework proposed by Zhou et al. (2012), the key innovation was the addition of a measure of similarity between systems. They were inspired by Rafols and Meyer (2010), Stirling (2007) and related ecological research (Nei & Li, 1979; Shriver et al., 1995) on the relationship of diversity within populations and the similarity between populations. They proposed a general framework to study systems' diversity and the similarity of systems.

**Diversity within systems** can be described, using the number of categories (number of species in ecology; number of scientific fields in scientometric studies), as variety and evenness; or one can go one step further and include a measure for the *disparity* between categories. So, diversity can be said to contain variety, eveness, and disparity. Many early cases used a classical measure, such as the Gini evenness index or the Simpson diversity measure. If there exists any disparity between categories, it is more appropriate to use the Rao-Stirling diversity measure proposed by Rao (1982) and Stirling (1998, 2007).

**Similarity between systems** can be studied by using a classical similarity measure such as Salton's cosine measure or a weighted form taking category similarity into account. For this purpose Zhou et al. (2012) proposed the similarity-weighted cosine measure. All this is presented in Table 6. This is an expansion on the ecological work of Nei & Lei (1979) and Shriver et al. (1995). It shows that their framework can be applied to any system.

**Table 1. Overview of the framework for the study of diversity and similarity proposed by Zhou et al. (2012)**

| | **Diversity within systems** | **Similarity between systems** |
|---|---|---|
| With a given number of **independent categories** in the system | Simpson diversity measure, $L_X = 1 - \sum_{i=1}^{N} p_{Xi}^2$ | Salton's cosine measure, $\lambda(X,Y) = \frac{\lambda_{X,Y}}{\sqrt{\lambda_X . \lambda_Y}} = \frac{\sum_{i=1}^{N} p_{Xi} p_{Yi}}{\sqrt{(\sum_{i=1}^{N} p_{Xi}^2)(\sum_{i=1}^{N} p_{Yi}^2)}}$ |
| Taking **disparity (or similarity) between categories** into account | Stirling-Rao measure, $SR_X = \sum_{i,j=1}^{N} p_{Xi} p_{Xj} D_{ij} = 1 - \sum_{i,j=1}^{N} p_{Xi} p_{Xj} S_{ij}$ | similarity-weighted cosine measure, $\varphi(X,Y) = \frac{\varphi_{X,Y}}{\sqrt{\varphi_X \varphi_Y}} = \frac{\sum_{i,j=1}^{N} p_{Xi} p_{Yj} S_{ij}}{\sqrt{(\sum_{i,j=1}^{N} p_{Xi} p_{Xj} S_{ij}) . (\sum_{i,j=1}^{N} p_{Yi} p_{Yj} S_{ij})}}$ |

## Our understanding of knowledge integration and disciplinary structure

*Knowledge integration*

Knowledge integration can be described as a property of an article (Porter et al., 2007) or a set of articles (Liu et al., 2012). In this research, we want to focus on the analysis of specific areas (a set of articles). The case study is based on the publications in SCI database of several sub-disciplines of ecology.

What are the knowledge integration's breadth and intensity? From how many SCs is the knowledge of a certain sub-discipline derived? Which SCs contribute the most to the sub-discipline? What is the relationship between sub-disciplines? Can we construct the typology of sub-disciplines based on knowledge integration? We are trying to answer these questions with the frameworks proposed by Rafols and Meyer (2010), Liu et al. (2012) and Zhou et al. (2012).

*Diversity as one attribute of knowledge integration*

Diversity is the property of how the elements of a system are apportioned into categories (Stirling, 2007). As an aspect of knowledge integration, diversity is now based on the image of the categories-mapping (Liu et al., 2012). As explained by Rafols and Meyer (2010) the best approach is to take the three aspects of diversity – i.e. variety, balance and disparity – into account. If a distance or dissimilarity measure exists in the categories, this suggests the need to use the Rao-Stirling diversity measure.

We can consider systems to be the sub-disciplines, the elements to be the references of the sub-disciplines, and the categories to be the SCs to which the references belong. This way the three aspects of diversity can measure the different aspects of the level of knowledge integration in contributing sets:

- Knowledge Integration Breadth (Variety)

Variety is the number of categories of elements, in this case, the SCs into which publications or references can be partitioned. We followed Liu and Rousseau (2010) and Liu et al. (2012) calling variety "SCs knowledge integration breadth", the number of SCs (or ESI fields) in which a set of articles is cited.

- Knowledge Integration Intensity (Balance) – The distribution of the 172 Subject Categories (172SCs). The 172SCs are defined by the JCR system.

We define SCs knowledge integration intensity as the distribution of the 172SCs.

- Knowledge Integration Intensity (Balance + Disparity) – The distribution of the 14 disciplines (or 22 broad fields).

Within the Essential Science Indicators (ESI) the science system is divided into 22 broad fields in such a way that every journal belongs to exactly one field. It is implied that the 22 broad fields are independent. Unlike the 22 broad fields, the ISI subject categories are not disjointed or hierarchically organized, but interconnected, because more than one category is often attributed to a journal (Leydesdorff and Rafols, 2009). Furthermore, these are more specific and therefore contain more information. As we use a method that is designed especially to take similarity between categories into account, we prefer the more detailed approach with more than 170 subdivisions.

Leydesdorff and Rafols (2009) applied factor analysis based on the citation matrix of 172 Subject Categories, suggesting a 14-factor solution with a minimum loss of information. Factor analysis is used to identify clusters of inter-correlated variables (here mean Subject Categories). Factors consist of relatively homogeneous variables. That is to say factor analysis taking category similarity into account. Considering this we confirm that Leydesdorff's 14 disciplines can interpret the results of the field similarity-weighted cosine similarity.

Another definition is SCs knowledge integration intensity as the distribution of 14 disciplines. The 14 disciplines can also help us to distinguish which discipline is contributed more in the process of knowledge integration.

*Coherence and similarity are the attributes of knowledge integration at different levels*

Knowledge integration is not only about how diverse the knowledge is, but also about making connections between the various bodies of knowledge drawn upon.

Coherence is another attribute of interdisciplinary knowledge integration (Liu et al., 2012). It is the property describing how the elements of a system are related to each other. Rafols et al. (2012) ensured that this measure of coherence is orthogonal to diversity.

Our interest lies in using the framework proposed by Zhou et al. (2012) to track knowledge integration. In the process of knowledge integration, if two sets of articles derive the ideas from the same SCs, we can say they are close in relationship. If two sets of articles seldom share ideas from the same SCs, we can say they are relatively distant in relation. So we can use the cosine measure and

the similarity-weighted cosine measure to explore the relation (homogeneity) between different systems (sets of articles or sub-disciplines).
It is clear the coherence is critical in the relationship between elements of a system. But similarity between systems is also very important in the process of knowledge integration.

*Disciplinary structure*

Mapping of documents has been a discussion topic in scientometric research for a number of years (Boerner et al., 2003). In general, the procedure follows a three-step process (Sternitzke & Bergmann, 2009).

- First, bibliographic coupling, co-citation analysis, co-word analysis, co-authorship, and semantic structures of texts are the common methodologies to select the (bibliographic) data from documents (Kessler, 1963, Small, 1973, Marshakova, 1973, Rip & Courtial, 1984, Callon et al., 1991 and Tsourikov et al., 2000).
- In the second step, similarities are computed based on the above-mentioned data. Measures such as the Pearson correlation coefficient, Salton's Cosine formula, the Jaccard Index, or the Inclusion Index are possible methods of normalizing this data (Hamers et al., 1989; Peters et al., 1995; Qin, 2000; Ahlgren et al., 2003).
- Finally, in the third step, the previously computed data is visualized by means of multivariate analyses such as cluster analysis or multidimensional scaling (MDS) (see e.g. Leydesdorff, 1987) or social network analysis (see e.g. Leydesdorff and Rafols, 2009).

Bibliometric methods have been used in the study of interdisciplinarity, especially those based on the "maps of science," built upon co-word, co-authorship, or co-citation analysis, which aim to identify structural relations between various subfields and to show them in graphical representations (Tijssen, 1992; Kessler, 1963, Small, 1973, Marshakova, 1973, Rip & Courtial, 1984; Callon et al., 1991] and Tsourikov et al., 2000).
We sort out the frameworks of Zhou et al. (2012) into co-classification methods, which can also identify structural relations between various subfields.

*Similarity as a measure to construct discipline structure based on knowledge integration*

Two closely related individuals have a lot of genetic information in common (biology, ecological point of view). Kessler (1963) suggested the use of the references contained in papers, given that documents with the same references are regarded as very similar in nature. This approach is known as bibliographic coupling. The same for two set of articles (such as two sub-disciplines), if they have references from the same SCs, they are also considered related. This is Reference Co-Classification (RCC).

So we can say bibliographic coupling and Reference Co-Classification (RCC) are similar in approach in capturing similarity in the process of knowledge integration.
The Salton's cosine index is a commonly used similarity measure, the greater the relationship between two given sets of articles, the higher the similarity between them. Its value ranges from 0 (no relation at all) to 1 (maximum relation).
The Similarity-Weighted cosine measure, proposed by Zhou et al. (2012) considers the similarity between the SCs, which can also identify structural relations between various sub-disciplines.
In the following sections we show the results of an empirical study. To refine similarity results based on the cosine index we will use the similarity-weighted cosine index. Diversity within sub-disciplines is measured using the Rao-Stirling diversity. An investigation of the similarities between sub-disciplines leads to a general view on the homogeneity of the group of sub-disciplines under study.

**The empirical study**

*Data*

Data are extracted from the Thomson Reuters database. 7 sub-disciplines of ecology were chosen: GLOBAL CHANGE BIOLOGY, LANDSCAPE ECOLOGY, MICROBIAL ECOLOGY, WILDLIFE BIOLOGY, MOLECULAR ECOLOGY, RESTORATION ECOLOGY, and SOIL BIOLOGY. For each sub-discipline, sample bibliometric records come from the following journals: *GLOBAL CHANGE BIOLOGY, LANDSCAPE ECOLOGY, MICROBIAL ECOLOGY, WILDLIFE BIOLOGY, MOLECULAR ECOLOGY, RESTORATION ECOLOGY, and EUROPEAN JOURNAL OF SOIL BIOLOGY*.
The references of the publications of each sub-discipline were counted and grouped into 172 SCs.

*Method*

We use a case study of sub-disciplines in ecology to analyze interdisciplinarity as revealed through the set of references. The analysis yields quantitative measures of: (1) the level of knowledge integration in contributing sub-disciplines; (2) the strength of knowledge integration relations between these sub-disciplines. A topological structure based on disciplinary similarity of sub-disciplines is constructed.
Since categories are scientific fields we use a field-similarity weighted cosine measure. In order to obtain the Rao-Stirling diversity index, a field-similarity matrix $S_{ij}$ in the cited dimension provided by Leydesdorff and Rafols (2009) is chosen (see Table 3). The $S_{ij}$ describes the similarity in the citation patterns for each pair of SCs in 2006.

**Table 2. Basic sample data information**

| No. | Short Name of System | System<br>*Sub-discipline* | Sample<br>*Journal* | Element<br>*Record* | Intermediary Set<br>*Reference* |
|---|---|---|---|---|---|
| 1 | GLOB | GLOBAL CHANGE BIOLOGY | *GLOBAL CHANGE BIOLOGY* | 2174 | 70295 |
| 2 | LANDE | LANDSCAPE ECOLOGY | *LANDSCAPE ECOLOGY* | 1157 | 36821 |
| 3 | MICE | MICROBIAL ECOLOGY | *MICROBIAL ECOLOGY* | 2359 | 62268 |
| 4 | WILDB | WILDLIFE BIOLOGY | *WILDLIFE BIOLOGY* | 506 | 15188 |
| 5 | MOLE | MOLECULAR ECOLOGY | *MOLECULAR ECOLOGY* | 4882 | 122398 |
| 6 | TROE | TROPICAL ECOLOGY | *JOURNAL OF TROPICAL ECOLOGY* | 1416 | 28549 |
| 7 | SOIB | SOIL BIOLOGY | *EUROPEAN JOURNAL OF SOIL BIOLOGY* | 721 | 20443 |

The similarity between sub-disciplines is shown with the Reference Co-Classification (RCC) which indicates the breadth of the basis of knowledge in common between sub-disciplines.
From the ISI Web of Science we downloaded full bibliometric records for the publications. These records were processed using the bibliometric program TDA, the statistical packet SPSS (2007), the network analysis software Ucinet, and Excel.
The Ucinet was used for multidimensional scaling techniques (MDS). And the sub-disciplines were grouped according to their normalized disciplinary similarity through hierarchical clustering analysis (SPSS, Ward Method).

*Results and Discussion*

(1) The level of knowledge integration in contributing sub-disciplines:

- Knowledge Integration Breadth (Variety)

MOLE (Molecular Ecology), MICE (Microbial Ecology) and GLOB (Global Change Biology) have the highest Knowledge Integration Breadth; the knowledge of the three sub-disciplines comes from 128, 125, and 119 SCs.

- Knowledge Integration Intensity (Balance + Disparity) – The distribution of the 14 disciplines and 3 sup-disciplines.

The 14 disciplines can also help us to distinguish which discipline contributes more to the process of knowledge integration (see Table 5). We give some example of this index to explain the diversity measure and the relationship between sub-disciplines.

- The profile of knowledge integration--Simpson diversity and Rao-Stirling diversity.

Diversity is a combined index. It can give us the profile of the knowledge integration. From Table 6, we can see that GLOB has the highest Knowledge

Integration Intensity. It shows the highest values (0.9066 in Simpson diversity and 0.6408 in Rao-Stirling diversity, respectively).

Three sub-disciplines WILDB, TROE (Tropical Ecology) and MOLE show the lowest level of interdisciplinarity, since the disciplinary diversities within these sub-disciplines are the lowest (Table 6).

**Table 3. Similarity matrix Sij of 172 SCI subject categories – partim**

| Number | 172 SCs | 1 | 2 | 3 | 4 | 5 | 6 | … |
|---|---|---|---|---|---|---|---|---|
| 1 | Biochemistry & Molecular Biology | 1.0000 | 0.9760 | 0.9489 | 0.0226 | 0.3372 | 0.1406 | … |
| 2 | Biophysics | 0.9760 | 1.0000 | 0.9041 | 0.0375 | 0.3444 | 0.1947 | … |
| 3 | Cell Biology | 0.9489 | 0.9041 | 1.0000 | 0.0126 | 0.2492 | 0.0928 | … |
| 4 | Thermodynamics | 0.0226 | 0.0375 | 0.0126 | 1.0000 | 0.1738 | 0.2883 | … |
| 5 | Chemistry, Applied | 0.3372 | 0.3444 | 0.2492 | 0.1738 | 1.0000 | 0.5053 | … |
| 6 | Chemistry, Physical | 0.1406 | 0.1947 | 0.0928 | 0.2883 | 0.5053 | 1.0000 | … |
| … | …… | …… | …… | …… | …… | …… | …… | … |

**Table 4. Distribution of Top SCs for 7 sub-disciplines based on the Knowledge Integration Intensity**

| **14 disciplines** | **172SCs** | MOLE | MICE | GLOB | SOIB | TROE | LANDE | WILDB |
|---|---|---|---|---|---|---|---|---|
| Ecology | Ecology | 4.73% | 3.67% | ***17.00%*** | ***19.66%*** | ***47.34%*** | ***31.46%*** | ***19.18%*** |
| Ecology | Forestry | 0.68% | 0.40% | 7.90% | 1.72% | 6.17% | 6.10% | 1.65% |
| Ecology | Marine & Freshwater Biology | 3.76% | ***10.96%*** | 2.36% | 0.96% | 1.08% | 1.74% | 0.54% |
| Ecology | Oceanography | 0.97% | 9.78% | 2.24% | 0.27% | 0.36% | 0.57% | 0.29% |
| Ecology | Evolutionary Biology | ***34.81%*** | 1.54% | 1.58% | 1.30% | 6.61% | 4.79% | 5.21% |
| Ecology | Zoology | 5.62% | 0.46% | 1.04% | 3.45% | 8.15% | 5.59% | ***49.01%*** |
| Ecology | Ornithology | 0.88% | 0.07% | 0.44% | 0.01% | 1.44% | 1.91% | 6.50% |
| Biomedical Sciences | Multidisciplinary Sciences | 9.44% | 5.25% | ***10.83%*** | 2.33% | 5.06% | 3.85% | 2.26% |
| Biomedical Sciences | Biochemistry & Molecular Biology | 3.35% | 3.97% | 0.59% | 2.26% | 0.23% | 0.08% | 0.35% |
| Biomedical Sciences | Biotechnology & Applied Microbiology | 1.36% | ***21.82%*** | 0.47% | 5.41% | 0.07% | 0.02% | 0.03% |
| Biomedical Sciences | Genetics & Heredity | ***13.81%*** | 0.54% | 0.07% | 0.33% | 0.20% | 0.35% | 0.74% |
| Agriculture | Plant Sciences | 3.92% | 3.71% | ***10.77%*** | 4.96% | 6.50% | 1.93% | 0.19% |
| Agriculture | Soil Science | 0.13% | 4.57% | 5.99% | **25.83%** | 1.31% | 0.75% | 0.01% |
| Environmental Sciences | Environmental Sciences | 2.18% | 1.89% | ***13.49%*** | 6.61% | 5.93% | ***17.22%*** | 6.87% |
| Geosciences | Geosciences, Multidisciplinary | 1.04% | 1.73% | 8.83% | 1.20% | 1.09% | ***11.94%*** | 0.50% |
| Infectious Diseases | Microbiology | 0.65% | ***13.61%*** | 0.42% | 4.24% | 0.05% | 0.01% | 0.00% |
| … | … | … | … | … | … | … | … | … |

(2) The strength of knowledge integration relations between these sub-disciplines: A topological structure based on disciplinary similarity of sub-disciplines is constructed. It is a comparison between the results obtained with the cosine and with the field-similarity weighted cosine index (Shown in Table 1).

Figure 2 shows the final dendrograms using Ward's Method based on the cosine index and field-similarity weighted cosine index. Figure 3 shows the results of the MDS analysis. The distribution of the 14 disciplines and 3 sup-disciplines of 7 sub-disciplines are used to explain Figure 2.

Since Ward's method is a bottom-up agglomerative clustering method (each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy) we compare the results of the two methods in a "bottom-up" fashion.

**Table5. The distribution of the 14 disciplines and 3 sup-disciplines of 7 sub-disciplines**

| | | MOLE | MICE | GLOB | SOIB | TROE | LANDE | WILDB |
|---|---|---|---|---|---|---|---|---|
| **14 disciplines** | Biomedical Sciences | ***31.20%*** | ***35.29%*** | ***15.81%*** | ***14.81%*** | 7.94% | 7.11% | 5.82% |
| | Clinical Sciences | 0.04% | 0.27% | 0.02% | 0.05% | 0.04% | 0.04% | 0.11% |
| | Neuro-Sciences | 0.22% | 0.02% | 0.04% | 0.01% | 0.07% | 0.06% | 0.16% |
| | Infectious Diseases | 2.19% | ***17.57%*** | 0.60% | 6.37% | 0.55% | 0.12% | 1.22% |
| | Gen. Medicine; Health | 0.43% | 0.46% | 0.14% | 0.18% | 0.25% | 0.53% | 0.74% |
| | Materials Sciences | 0.01% | 0.18% | 0.13% | 0.41% | 0.01% | 0.04% | 0.01% |
| | Chemistry | 0.20% | 1.15% | 0.40% | 1.21% | 0.23% | 0.08% | 0.05% |
| | Computer Sciences | 1.14% | 0.26% | 0.09% | 0.18% | 0.09% | 0.66% | 0.05% |
| | Engineering | 0.04% | 0.04% | 0.05% | 0.03% | 0.04% | 0.22% | 0.08% |
| | Physics | 0.01% | 0.08% | 0.07% | 0.01% | 0.03% | 0.18% | 0.02% |
| | Ecology | ***54.74%*** | ***27.55%*** | ***33.57%*** | ***29.68%*** | ***73.58%*** | ***53.62%*** | ***82.84%*** |
| | Environmental Sciences | 2.24% | 4.17% | ***14.68%*** | 8.43% | 6.36% | ***18.29%*** | 7.10% |
| | Geosciences | 1.38% | 2.53% | ***14.16%*** | 1.54% | 1.70% | ***15.64%*** | 0.62% |
| | Agriculture | 6.16% | ***10.40%*** | ***20.24%*** | ***37.04%*** | 9.11% | 3.43% | 1.16% |
| **3 sup-disciplines** | **Life Sciences** | ***34.08%*** | ***53.61%*** | ***16.60%*** | ***21.42%*** | 8.85% | 7.85% | 8.05% |
| | **Physical Sciences** | 1.39% | 1.71% | 0.74% | 1.85% | 0.39% | 1.17% | 0.21% |
| | **Environmental Sciences** | ***64.52%*** | ***44.64%*** | ***82.65%*** | ***76.70%*** | ***90.75%*** | ***90.97%*** | ***91.72%*** |

**Table 6. Simpson and Rao-Stirling diversity values for seven sub-disciplines**

| Sub-disciplines | Variety | Simpson diversity | Rao-Stirling diversity |
|---|---|---|---|
| GLOB | 119 | 0.9066 | 0.6408 |
| SOIB | 89 | 0.8751 | 0.6284 |
| MICE | 125 | 0.9000 | 0.6218 |
| LANDE | 109 | 0.8434 | 0.5632 |
| MOLE | 128 | 0.8389 | 0.4642 |
| TROE | 86 | 0.7488 | 0.4349 |
| WILDB | 79 | 0.7095 | 0.3707 |

Corresponding to the divisions of the dendrogram is the cluster-enhanced MDS map below:

In the MDS analysis based on the cosine index (Figure 3), WILDB belongs to the same cluster as MOLE and MICE (Cluster 1). Taking field similarity into account, based on field-similarity weighted cosine index, WILDB joins TROE, LANDE (Landscape Ecology), and the others in Profile 2.

From Table 5, we can see WILDB has 82.84% knowledge derived from Ecology. The percentage is at the same level as TROE (73.58% knowledge derived from Ecology).

The data shows that the 54.74% and 27.55% of the knowledge of MOLE and MICE comes from Ecology, while the 31.20% and 35.29% of the knowledge of MOLE and MICE is integrated from Biomedical Sciences. WILDB has the same knowledge basis as TROE, and LANDE, not MOLE and MICE.

At the level of 3 sup-disciplines, “Life Sciences” contributes 34.08% and 53.61% knowledge to MOLE and MICE. While the three sub-disciplines (TROE, LANDE and WILDB) inherit 90% of their knowledge from "Environmental Sciences". It can be seen that the MDS map based on the field-similarity weighted cosine index gives us the correct description of the relationships between 7 sub-disciplines.

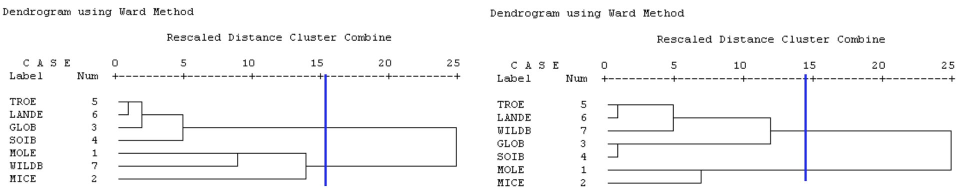


**Figure 2. Dendrograms using Ward's clustering analysis of sub-disciplinary structure. Left: based on the cosine index; Right: based on the similarity-weighted cosine index**

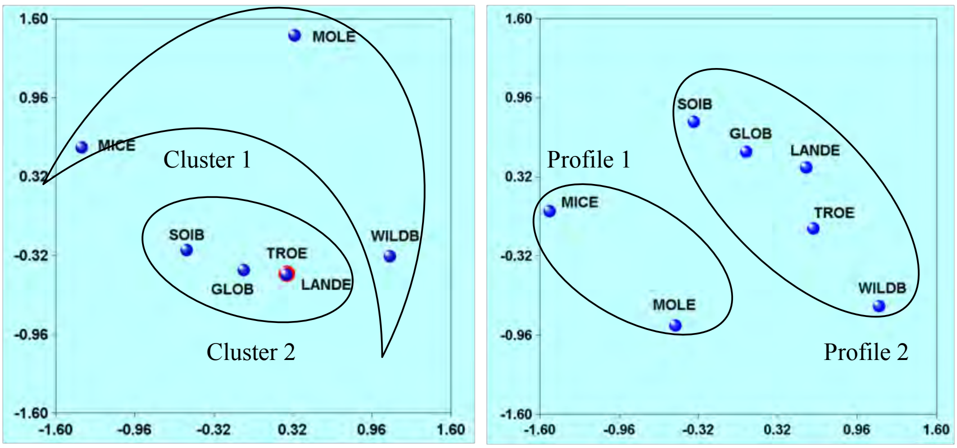


**Figure 3. Maps resulting from MDS analysis of sub-disciplinary structure. Left: based on the cosine index; Right: based on the similarity-weighted cosine index**

## Conclusion and further research

A key point in the study is determining to what extent our indicators are based on categories' distribution of references into measured knowledge integration. From a conceptual standpoint, we consider the indicators proposed to be valid.

As to the most relevant results of our study, we would like to stress the following:

● Disciplinary diversity at the level of knowledge integration is observed, with GLOB at the upper range of the scale. The Rao-Stirling diversity can give the whole profile of the level of the knowledge integration.

The three aspects of diversity can measure the different aspects of the level of knowledge integration in contributing sets:

Variety – the SCs into which publications can be partitioned can measure the Knowledge Integration Breadth of a set of articles.

Balance_+ Disparity – The distribution of the 14 disciplines can measure the Knowledge Integration Intensity of a set of articles.
●The similarity of sets of articles (sub-disciplines) based on the Reference Co-Classification (RCC) has proven useful in providing a deeper understanding of the relations between sub-disciplines. Since the field-similarity weighted cosine is derived from Rao-Stirling diversity, so the aspects of Balance + Disparity (the distribution of the 14 disciplines) can be used to explain to what extent those two sub-disciplines are related based on knowledge integration.

In summary, we propose that the bibliometric methodology presented here provides a compelling overview of science's structure with a special focus on the inter-relationship between sub-disciplines that makes knowledge integration possible.

However, the results should be analyzed with caution since they are highly dependent on the ISI classification scheme, which is not perfect and has a metaphorically coarse granularity. Future research will include the dynamic structure of the literature over time.

**References**

Ahlgren, P., Jarneving, B. & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient, *Journal of the American Society for Information Science*, 54, 550–560.

Boerner, K., Chen, C. & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179–255.

Callon, M., Courtial, J. P. & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22, 155–205.

Hamers, L., Hemeryck, Y., Herweyers, G., Janssen, M., Keters, H., Rousseau, R. & Vanhoutte, A. (1989), Similarity measures in scientometric research: the Jaccard index versus Salton's cosine formula. *Information Processing and Management*, 25, 315–318.

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

Leydesdorff, L. (1987). Various methods for the mapping of science, *Scientometrics*, 11, 295–324.

Leydesdorff, L. & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*, 60 (2), 348–362.

Leydesdorff , L. & Rafols, I.(2011). Local Emergence and Global Diffusion of Research Technologies: An Exploration of Patterns of Network Formation. *Journal of the American Society for Information Science and Technology,* 62(5):846–860.

Liu, Y.X., Rafols I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion, *Journal of Documentation*, 68 ,(1),31 – 44.

Liu, Y.X. & Rousseau, R. (2010). Knowledge diffusion through publications and citations: a case study using ESI-fields as unit of diffusion”. *Journal of the American Society for Information Science and Technology*, 61(2), 340-51.
Marshakova, I. V. (1973). System of document connections based on references. *Scientific and Technical Information Serial of VINITI*, 6, 3–8.
Morillo, F., Bordons, M. & Gomez, I. (2003). Interdisciplinarity in science: a tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237-1249.
Peters, H., Braam, R. & Raan, A. (1995). Cognitive resemblance and citation relations in chemical engineering publications. *Journal of the American Society for Information Science*, 46, 9–21.
Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature. *Journal of the American Society for Information Science*, 51, 166–180.
Nei,M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
Nei, M. & Li W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Science USA* 76, 10, 5269-5273
Rafols, I. & Meyer, M. (2007). How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors. *Scientometrics,* 70(3), 633-650.
Rafols, I. & Meyer, M. (2010). Diversity and Network Coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics* 82, (2), 263-287.
Rafols,I., Leydesdorff,L. ,Hare,A., Nightingalea, P. & Stirling, A. (2012) .How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, 41, 1262-1282.
Rao, C.R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology,* 21(1), 24-43.
Rip, A. & Courtial, J. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6, 381–400.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265–269.
Spasserm, A. (1997). Mapping the terrain of Pharmacy: co-classification analysis of the International Pharmaceutical Abstract Database. *Scientometrics*, 39(1), 77-97.
Sternitzke, C. & Bergmann, I. (2009) . Similarity measures for document mapping: A comparative study on the level of an individual scientist. Scientometrics, 78, (1), 113–130.
Stirling, A. (1998). On the economics and analysis of diversity. SPRU Electronic Working

Paper.http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp28/sewp28.pdf Accessed 01-04-2006.
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interfa*ce, 4 (15), 707-719.
Solow, A., Polasky, S. & Broadus, J. (1993). On the measurement of biological diversity. – *J. Environ. Econ. Manage*, 24, 60–68.
Tijssen, R.J.W. (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research. *Research Policy*, 22, 27–44.
Tsourikov, V. M., Batchilo, L. S. & Sovpel, I. V. (2000). Document semantic analysis/selection with knowledge creativity capability utilizing subject-action-object (SAO) structures, United States Patent No. 6167370.
Zhou,Q.J., Rousseau, R., Yang, L.Y.,Yue,T. & Yang, G.L. (2012). A general framework for describing diversity within systems and similarity between systems with applications in informetrics. *Scientometrics*, 93( 3), 787-812.