

A Bibliometric Approach for Department-Level Disciplinary Analysis and Science Mapping of Research Output Using Multiple Classification Schemes

Pitambar Gautam ¹

This study describes an approach for comparative bibliometric analysis of scientific publications related to (i) individual or several departments comprising a university, and (ii) broader integrated subject areas using multiple disciplinary schemes. It uses a custom dataset of scientific publications (ca. 15,000 articles and reviews, published during 2009-2013, and recorded in the Web of Science Core Collections) with author affiliations to the research departments, dedicated to science, technology, engineering, mathematics, and medicine (STEMM), of a comprehensive university. The dataset was subjected, at first, to the department level and discipline level analyses using the newly available KAKEN-L3 classification (based on MEXT/JSPS Grants-in-Aid system), hierarchical clustering, correspondence analysis to decipher the major departmental and disciplinary clusters, and visualization of the department-discipline relationships using two-dimensional stacked bar diagrams. The next step involved the creation of subsets covering integrated subject areas and a comparative analysis of departmental contributions to a specific area (medical, health and life science) using several disciplinary schemes: Essential Science Indicators (ESI) 22 research fields, SCOPUS 27 subject areas, OECD Frascati 38 subordinate research fields, and KAKEN-L3 66 subject categories. To illustrate the effective use of the science mapping techniques, the same subset for medical, health and life science area was subjected to network analyses for co-occurrences of keywords, bibliographic coupling of the publication sources, and co-citation of sources in the reference lists. The science mapping approach demonstrates the ways to extract information on the prolific research themes, the most frequently used journals for publishing research findings, and the knowledge base underlying the research activities covered by the publications concerned.

Keywords: bibliometry, science map, correspondence analysis, clustering, research field, cross-disciplinarity, co-word, co-citation, KAKEN-L3

¹ Creative Research Institution (CRIS), Hokkaido University, N21 W10, Kita-Ku, Sapporo, Hokkaido 001-0021, Japan. e-mail: pgautam@cris.hokudai.ac.jp

Introduction

Research performance at the level of various research entities (e.g., researcher, research unit, university or research institute as well as country or even regional geopolitical unit) is primarily judged using the basic indicators such as the number of publications as a quantity and the citation impact as a quality or the degree of influence on the relevant research community (Glänzel, 2003; Moed, 2010; Gautam, 2016). For research benchmarking such as university rankings, a much more extensive set of indicators, with varying weights, either as raw data or after standardization for the publication period, research area, document type, the size of research entities, funding data, etc. is used (Moed, 2017; Gautam, 2017).

Likewise, complex bibliometric indicators, such as citation percentiles (typically, percentage of publications within the top 1% and top 10% percentile windows), citation indices (h-index and its variants), journal influence factors like the journal quality indicators and category-normalized citation indicators are being increasingly used for research evaluation or to respond to the needs of both governmental and non-governmental funding bodies (Williams & Bornmann, 2014; Gautam, 2017). Also, citation data related to research publications worldwide comprise the ‘big data’ requiring network analysis to map the co-citation and bibliographic coupling relationships that provide information on the intellectual knowledge base and research strengths (e.g., research fronts, research competencies) of the research entities (Börner & Polley, 2014; van Eck & Waltman, 2014; Zupic & Čater, 2014; Jalali & Park, 2018).

Disciplinary classifications (research fields, subject areas, subject categories, etc.) that appropriately address the subject orientations of the targets need to be considered during collection, calculation and interpretation of the research performance indicators at the level of research entities. It is also important to pay attention to the fact that citation-based indicators are normally aggregated over the subject categories for which data availability is technically dependent on the databases, which are managed by vendors or organizations that prefer specific disciplinary classification scheme(s) (Rafols, 2014). The bibliometric indicators, often with due consideration of citations, are being increasingly used in the research performance assessments and the university research administrators (URAs) engaged in formulation, planning and designing the research strategies led by the university executives, play a vital role in generating them. For these reasons, it is high time to develop methodologies for accurate and effective analysis of the scientific publications with due account of all possible factors including the disciplinary or cross-disciplinary (including multidisciplinary, transdisciplinary and interdisciplinary) characteristics of the research entities (Wagner et al., 2011; Leydesdorff, 2013; Lee, 2013; Gautam et al., 2014).

Many bibliometric/scientometric studies have compared the publication records in two or more of the databases comprising mainly Web of Science (WoS), Scopus, Google Scholar and Dimensions at the level of various research entities or aspects (regions such as Latin America and the Caribbean: Santa & Herrero-Solana, 2010; national science system such as Russian: Moed et al., 2018; universities: Vieira & Gomes, 2009; institutes of technologies in India: Prathap, 2013; business schools: Mingers & Lipitakis, 2010; nursing schools: De Groote & Raszewski, 2012; large interdisciplinary project: Meho & Rogers 2008; specific research

community such as soil researchers: Minasny et al., 2013; senior academics: Harzing & Alakangas, 2016; humanities and social science programs: Prins et al., 2016; specific academic fields such as computer science (Franceschet, 2010), oncology (López-Illescas et al., 2008), and food science (Thelwall, 2018); article: de Winter et al., 2012; journal titles and coverage: Gavel & Iselid, 2008, Mongeon & Paul-Hus, 2016 and so on). These studies address many aspects of the research performance indicators both quantitative and qualitative, coverage of journals or other publication sources in terms of areal (regional, national or local) as well as broad fields of science (natural sciences; engineering; biomedical sciences; social sciences, arts and humanities), and database-specific differences from the field-specific and national citations indexes. Despite the discipline-related concerns, about the degree of representation of publications of research entities in specific fields and department-level research performance indicators, raised in these and other studies (e.g., Nederhof et al., 1993; Bartol et al., 2016), analysis of the same set of publications using various disciplinary schemes available through different databases has not yet received much attention.

This study is dedicated to the appraisal of differences arising from the analysis of scientific publications of the same research entity using multiple disciplinary classification schemes offered by different databases. For this purpose, it describes an approach of analysis of a 5-yr set of the peer-reviewed core journal publications (articles and reviews) from a comprehensive research-intensive university using disciplinary schemes used by WoS and Scopus. It is a known issue that scientific publications from social sciences and humanities departments, especially from universities in countries with the primary working language other than English, are under-represented in the most frequently used standard databases such as WoS and Scopus (e.g., Harzing & Alakangas, 2016, and references therein). This inference was found to be largely true by the author's own observations of research output of the university concerned. Therefore, only the publications produced by research entities (graduate schools/faculties/schools, research institutions, and research centers – loosely referred hereafter as departments) engaged in science, technology, engineering, mathematics, and medicine (STEMM), as verified by the affiliation of at least one (co-)author from them, are considered for further analysis.

The following sections describe mainly four different aspects of the bibliometric analysis and science mapping combined: (i) Breakdown of publications to departments and the KAKEN-L3 subject categories scheme (Clarivate Analytics, 2015a; JSPS, 2015), which was recently added to the InCites - a research analytics tool for analysis of the WoS content; both WoS and InCites are managed currently by Clarivate Analytics and accessible under licensed subscriptions; (ii) Clustering and correspondence analyses to decipher the department-discipline relationships in terms of the KAKEN-L3 subject categories; (iii) Reorganizing the publications considering the organizational affiliations and subject orientations, especially important for departments executing research that transcend the traditional research fields, and visualization with two-dimensional diagrams; (iv) Creation of subsets for broader disciplinary areas related to specific groups of departments, parsing them using several disciplinary schemes (Essential Science Indicators (ESI) 22 research fields, SCOPUS 27 subject areas, OECD Frascati 38 subordinate research fields, and KAKEN-L3 subject categories) and comparative disciplinary analysis; and,

(v) Illustration of the example subset obtained following the previous step using sophisticated mapping and visualization tools.

Table 1

Departments and related academic disciplines (adapted from Gautam (2016))

Department	Disciplinary Focus	Department	Disciplinary Focus
AGR	Agriculture	IGM	Genetic medicine
CIIS	Isotope science	ILTS	Low temperature science
CRC	Catalytic chemistry	LFSC	Life science
CRIS	Transdisciplinary sciences	MEDH	Medicine & Hospital
DENT	Dentistry	MUSE	Museum
EES	Environmental science	PHARM	Pharmacology & pharmacy
ENGG	Engineering	RCIQE	Quantum electronics
FISH	Fisheries	RCZC	Zoonoses
FLSC	Field science	RIES	Electronic science
GCSE	Chemical sciences & engineering	SCI	Natural sciences
GIST	Information science & technology	VETM	Veterinary medicine
HLTS	Health sciences		

Table 2

Four types of disciplinary classifications (subject areas, fields, subordinate fields, categories) used to assign scientific publications

1. ESI22 Fields	2. SCOPUS Subject Areas	3. OECD (Frascati) Subordinate Fields	4. KAKEN-L3 (Bunka3-H20) Categories
Agricultural sciences	Agricultural and biological sciences	Agriculture, forestry, fisheries	Agricultural Chemistry
Biology & biochemistry	*Arts and humanities	Animal and dairy science	Agricultural Science in Society and Economy
Chemistry	Biochemistry, genetics and molecular biology	*Art	Agro-engineering
Clinical medicine	*Business, management and accounting	Basic medical research	Animal Life Science
Computer science	Chemical engineering	Biological sciences	*Anthropology
*Economics & business	Chemistry	Chemical engineering	Applied Aquatic Science
Engineering	Computer science	Civil engineering	Applied Chemistry
Environment/ecology	*Decision sciences	Clinical medicine	Applied Physics
Geosciences	Dentistry	Computer and information sciences	Architecture and building engineering
Immunology	Earth and planetary sciences	Earth and related environmental sciences	*Area Studies
Materials science	*Economics, econometrics and finance	*Economics and business	Art Studies
Mathematics	Energy Engineering	*Educational sciences	Astronomy
Microbiology	Environmental science	Electrical engineering, electronic engineering	Basic Biology
Molecular biology & genetics	Health professions	Environmental engineering	Basic Chemistry
*Multidisciplinary	Immunology and microbiology	Health sciences	Basic Medicine
Neuroscience & behavior	Material science	*History and archaeology	Biological Science
Pharmacology & toxicology	Mathematics	Industrial biotechnology	Biomedical Engineering
Physics	Medicine	*Languages and literature	Boundary Agriculture
Plant & animal science	Multidisciplinary	*Law	Boundary Medicine
Psychiatry/psychology	Neuroscience	Materials engineering	Brain Sciences
*Social sciences, general	Nursing	Mathematics	Civil Engineering
Space science	Pharmacology, toxicology and pharmaceutics	Mechanical engineering	Clinical Internal Medicine
	Physics and astronomy	*Media and communication	Clinical Surgery
	Psychology	Medical engineering	*Cultural Anthropology
	Social sciences	Nano-technology	Culture Assets and Museology
	Veterinary	Other agricultural science	Dentistry
		Other engineering and technologies	Earth and Planetary Science
		*Other natural sciences	*Economics
		*Other social sciences	*Education
		*Philosophy, ethics and religion	Electrical and Electric Engineering
		Physical sciences and astronomy	Environmental Science
		*Political science	Forest and Forest Products Science
		Psychology	*Gender
		*Social and economic geography	Genome Science
		*Sociology	
		Veterinary science	

Note. ESI22 Fields classification is available in Web of Science and InCites, managed by Clarivate Analytics, whereas OECD Category to WoS Category Mapping 2012 and KAKEN-L3 assignments are available in InCites only. Scopus All Science Journal Classification (ASJC) scheme is used in Elsevier's SCOPUS database, and source title list with corresponding subject area information is openly available from the internet

(<https://www.elsevier.com/solutions/scopus/how-scopus-works/content>). Use of the first three classifications was described earlier by Gautam (2016).

*Subject Areas/Fields/Categories not considered in this study, which is limited to departments dealing with natural sciences, engineering, technology, mathematics, and medicine.

Data and methods of analysis

Bibliographic records for a total of 14,689 articles and reviews published during 2009-2013 by a comprehensive and research-intensive university were harvested from the WoS Core Collection (SCI-E, SSCI and A&HCI database modules) on February 23, 2015, as described by the author earlier (Gautam, 2015). The bibliographic records retrieved after searching the database for affiliations strings related to the concerned university were subjected to cleaning to verify whether they are correct and complete. A semi-automatic worksheet-based matching approach was used to identify any plausible affiliation strings related to the department(s) (i.e., research unit(s) such as graduate school/faculty/school, research institute or research center (Table 1) engaged independently in research and/or education programs), and each verified publication was assigned to one or several of them. If a publication had multiple affiliations, each of them was assigned full score using the whole-count method. Similarly, each publication was assigned to unique discipline (e.g., Biology & Biochemistry comprising ESI 22 fields) primarily based on the journal title following the master journal list (Clarivate Analytics, 2015b).

Due to the under-representation of the social sciences and humanities departments (e.g., economics, law, literature, etc.) mentioned above, further analysis is restricted to departments engaged in STEMM fields (Table 1). Furthermore, to ensure the representative size, departments that produced less than 10 papers a year were also excluded. A total of 13,537 publications, i.e. ca. 92% of the primary WoS-based dataset overlapping with Scopus, analyzed earlier additionally for SCOPUS and OECD Frascati schemes (Gautam, 2015 and 2017) have been further analyzed using the KAKEN-L3 scheme. The four different classification schemes used in the analysis are listed in Table 2.

Multivariate statistical analysis.

The percentage share of each discipline, as a size-independent measure, was used as raw data to perform hierarchical clustering and correspondence analysis. The purpose was to establish the similarity of departments in terms of disciplinary coverage. A commercial software (Excel Statistics for Windows 2012 by SSRI Co. Ltd., Tokyo) for multivariate analysis by these methods was used for this purpose. As described in Gautam (2015), “Hierarchical clustering utilized agglomeration with Ward’s criterion to build a cluster hierarchy by proceeding bottom-up, starting from the smallest clusters available and merging those nearest to each other at each step”. A 7-clusters solution was obtained to establish the clusters for departments and disciplines, separately. Correspondence analysis involved visually displaying both row (disciplines, $i=1$ to I) and column (departments, $j=1$ to J) categories of a contingency table $P = (p_{ij})$, $i \in I$, $j \in J$, whereby the distances between the presenting points reflect the patterns of co-occurrences in P (Mirkin, 2011). Theory underlying the correspondence analysis is detailed by Clausen (1998).

Stacked bar plots of department-discipline relationships.

A two-dimensional representation of the department-discipline relationship, in which the data on percent disciplinary shares stacked along the vertical axis against the departments arranged along the horizontal axis has been used for visualization of the results of analyses for each of the disciplinary classification schemes as shown in Gautam (2015, 2016). In such representation, the departments and disciplines are arranged using the results of clustering and correspondence analyses.

Science mapping.

Considering the departmental and disciplinary clusters revealed from the multivariate analysis, the publications data were regrouped into three major disciplinary classes. This process sometimes involved breakdown of interdisciplinary departments such as SCI into several disciplinary subsets (e.g., the earth science subset from SCI department (Gautam, 2016). The datasets for these classes were subjected to science mapping using Vosviewer – a Leiden University Software for visualizing scientific landscapes (van Eck & Waltman, 2010).

Results and interpretation

The results of analyses using the KAKEN-L3 disciplinary schemes will be presented first in the following sequence: departmental and disciplinary clusters, department-discipline relationship in correspondence plot, two-dimensional stacked bars plot showing the disciplinary shares for each department. Then, custom dataset creation and its comparative analysis for disciplinary contributions using four different schemes will be described. Finally, the results of science mapping to extract the most important relationships in terms of keywords representing the research themes, journals used for publications of research results and also the knowledge base used in conducting the research and interpretation of the results will be given.

Departmental and disciplinary clusters.

Figure 1 shows a tree diagram with seven clusters of departments, which are identified by abbreviations (with major fields of study in brackets) below:

Cluster I: AGR (agriculture, forestry); FISH (fisheries); FLSC (field-based biosphere-related sciences)

Cluster II: EES (earth, environmental and ecological sciences); ILTS (studies of low-temperature regions); MUSE (museum-based geological, biological, archaeological etc. specimens)

Cluster III: SCI (natural sciences: physics, chemistry, biology, geoscience, mathematics); CRIS (creative transdisciplinary sciences)

Cluster IV: ENGG (engineering, including applied physics, applied chemistry, etc.); GCSE (integrated chemical sciences & chemical engineering); RIES (electronic science); CRC (catalytic chemistry)

Cluster V: GIST (information science & technology, computer science); RCIQE (quantum electronics)

Cluster VI: CIIS (application of isotopes); HLTS (health sciences, nursing); IGM (genetic medicine); MEDH (medical fields including those related to hospital); LFSC (life sciences); PHARM (pharmacology & pharmacy); DENT (dentistry)

Cluster VII: RCZC (zoonotic infections); VETM (veterinary medicine)

At a higher hierarchical level, these seven clusters may be combined to 3 broader Groups, as shown in Fig. 1, engaged in fields as follows:

Clusters I, II & III: Basic physical, environmental, biological, agriculture and fisheries sciences. It includes the geosciences.

Clusters IV & V: Applied physical sciences and engineering. It includes the information science and technology.

Clusters VI & VII: Medical, health and life sciences

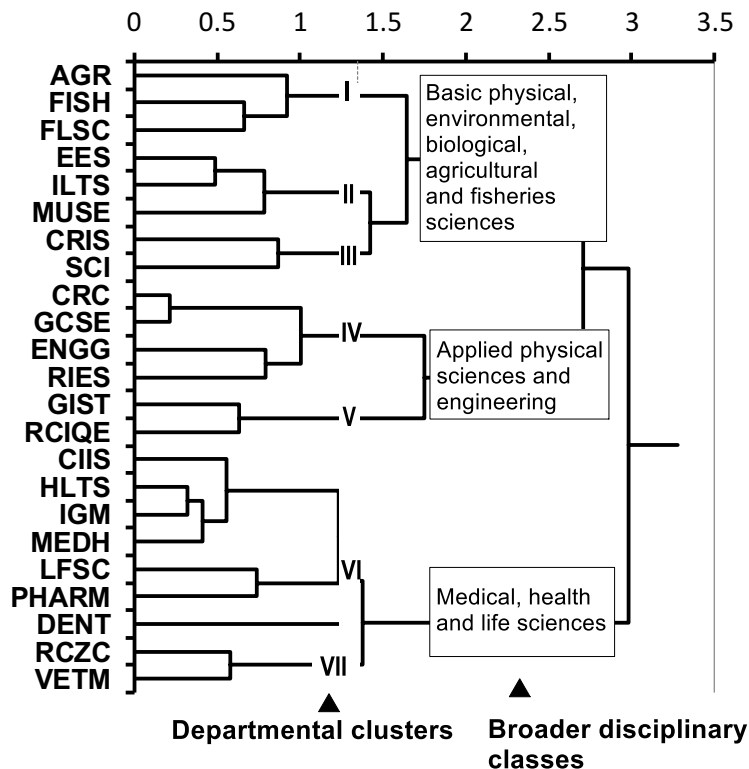


Figure 1. Tree diagram showing the results of hierarchical clustering of departments using agglomerative method with Ward’s criterion. The 7-clusters (I-VII) effectively group departments conducting research in overlapping KAKEN-L3 disciplines, although several departments such as CRIS, ENGG, MUSE, SCI, may not be uniquely assigned into the same cluster while using different classification schemes because of the largely interdisciplinary nature of their research output.

A tree diagram constructed from data for 7 clusters (numbered from 1 to 7) related to disciplines as variables is shown in Fig. 2. The clusters are further grouped into 3 classes in a way similar to that used in the case of departments, but the names assigned to each group differ slightly. One of the striking differences between the department-based and discipline-based clusters is the division of the physical sciences into basic (related to SCI) and applied (related to ENGG). Because of the large size and multidisciplinary nature (with basic/theoretical branches of physics, chemistry, biology, mathematics, and geology) nature of SCI, in contrast to MEDH and ENGG that are also large but have narrower disciplinary orientations, it is impossible to achieve accurate and unique assignment of this department to any cluster. A better approach would be to subdivide SCI into several parts and treat them as separate departments for clustering and subsequent analysis as shown by Gautam (2015).

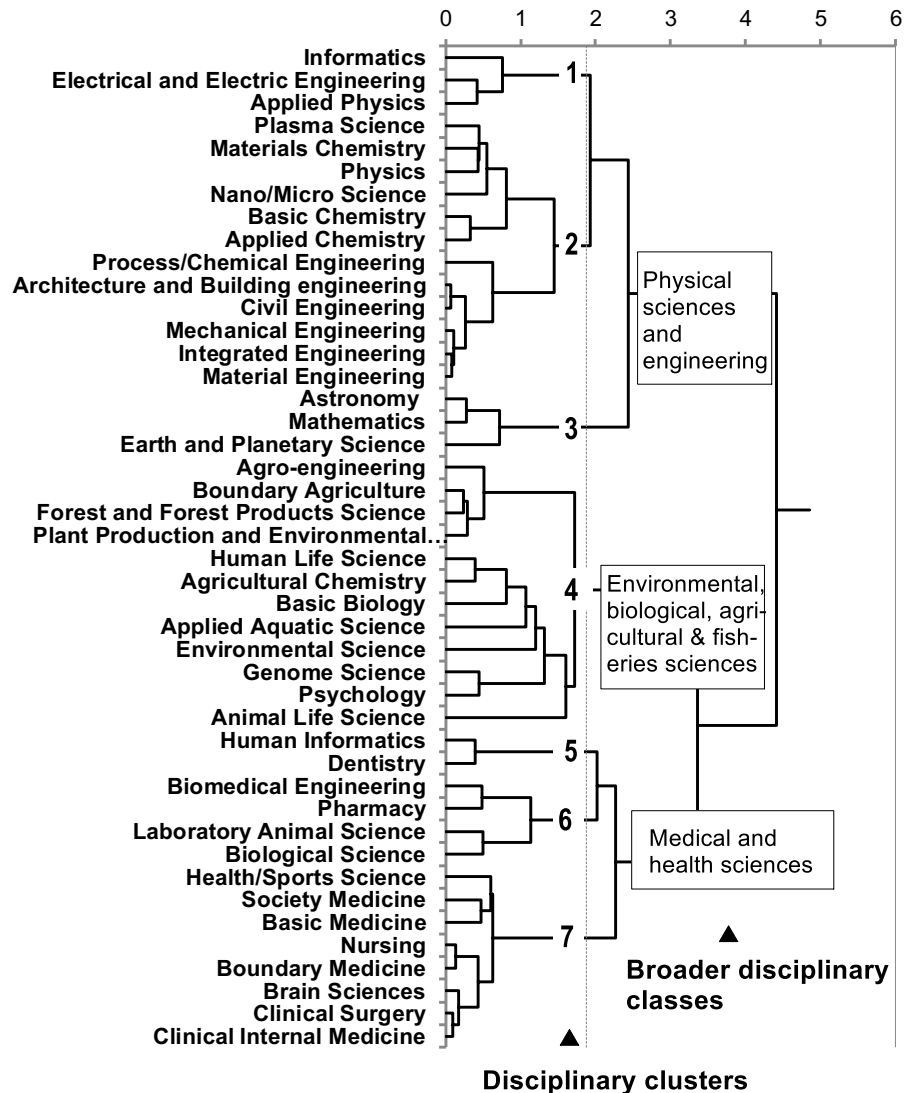


Figure 2. Tree diagram showing the results of hierarchical clustering of KAKEN-L3 disciplines using agglomerative method with Ward's criterion. The 7-clusters (1-7) effectively group closely related disciplines. Although Cluster 3 comprising Astronomy, Mathematics and Earth and Planetary Science shows greater affinity to the uppermost class, it may in practice be related also to the class in the middle.

Department-discipline relationships from correspondence analysis.

The mutual relationship of departments and disciplines can be reasonably demonstrated using a 2D representation determined by the degree of correspondence in terms of publications, where the values of two axes give the relative position of each entity (department or discipline). This is illustrated in Fig. 3, in which, for clarity, departments (Table 1) and subject categories (Table 2) are abbreviated. The size of each symbol is proportional to the logarithmic publication count (obtained by whole counting method) assigned to the department or subject category concerned.

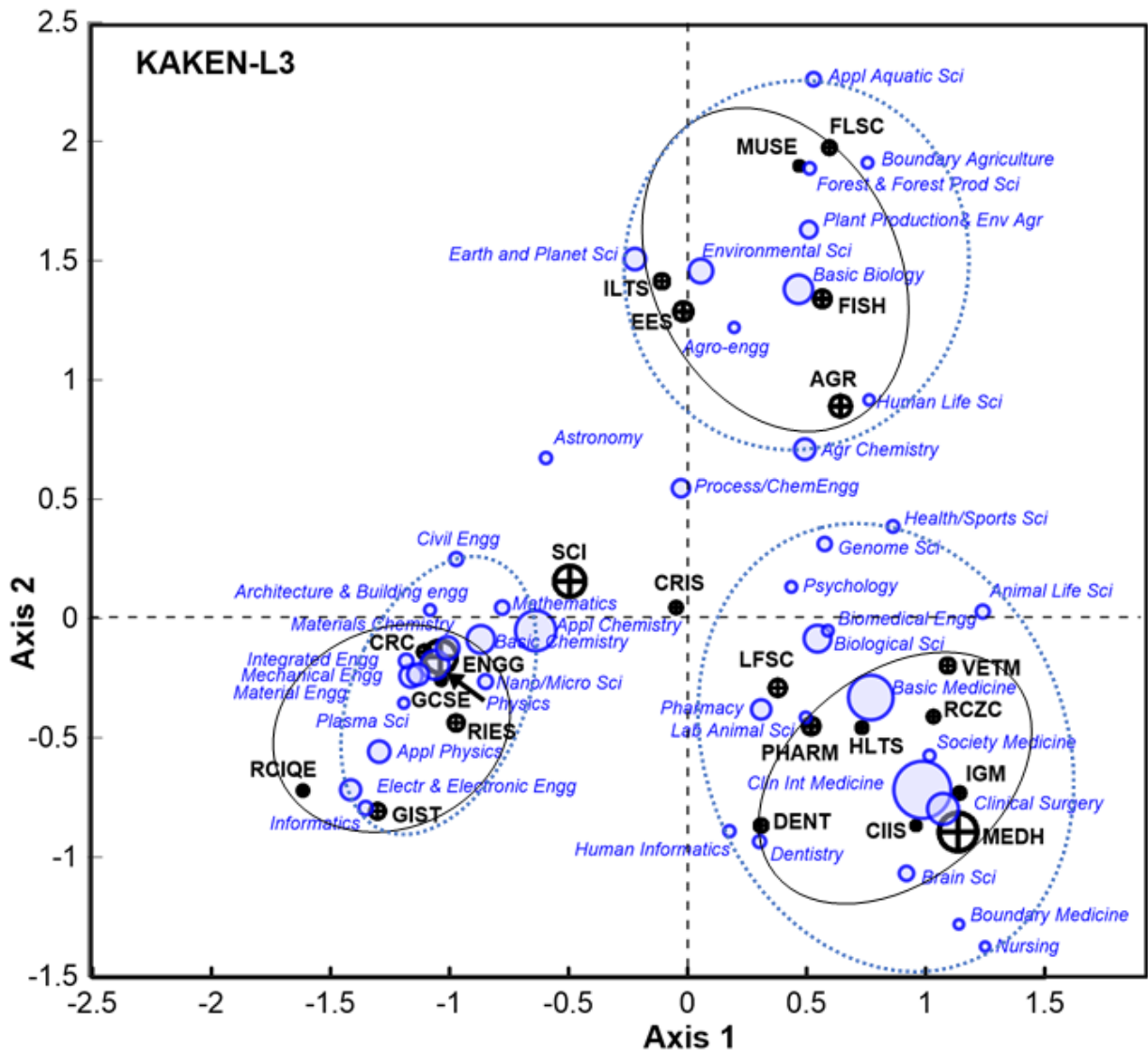


Figure 3. Visualization of department/discipline contingency data in terms of KAKEN-L3 disciplinary scheme using correspondence analysis. Crossed-out and filled circles stand for departments and subject categories. The size of the symbols is proportional to the logarithm of the publications counts ranging from 50 for CIIS to 2,662 for MEDH (for departments) and 19 for Nursing to 4,200 for Clinical Internal Medicine (for subject categories). Coordinates for departments and subject categories vary depending on the discipline-department relations in terms of 2 major axes (explaining 31.6 and 20.6% of total variance, respectively).

Visual inspection of the plot in Fig. 3 reveals that each disciplinary group (Fig. 1) recognized from the clustering of departments occupies a specific position (confined in one of the quadrants), where the departments generally plot within a relatively narrow region shown by ovals drawn by solid lines. Some exceptions are evident though, as CRIS, SCI and LFSC (to some extent) occupy positions that are farther from all other groups owing to their largely interdisciplinary and multidisciplinary nature.

Disciplines forming a specific disciplinary group (Fig. 2) also plot in a restricted region (ovals drawn by dotted lines) but somewhat wider than that for the corresponding departmental group. This is also logical because most of the departments are engaged in research in multiple disciplines and the department-discipline proximity is the indicator of how closely they are related to each other. For example, Astronomy and Process/Chemical engineering occupy positions that do not seem to be close to any particular departments. The position of Astronomy in this representation is governed by the proportion of documents produced mainly at SCI, ENGG, ILTS and CRIS, all of which contribute to this discipline.

The position of Process/Chemical engineering seems to be dictated from the proportions of papers shared with ENGG and probably two other departments (AGR and FISH), but verifying this requires examination of the author affiliations at article level. Similarly, 'Applied Physics' plots in between several departments (e.g., ENGG, GIST, RIES, RCIQE) that contribute to this discipline in varying degree. In contrast, Mathematics, Environmental Science and Dentistry plot very close to SCI, DENT and EES, respectively, implying that research in each discipline is restricted to unique department. The Earth and Planetary Sciences discipline occupies a position close to ILTS and EES implying their close relation; in fact, it is contributed by SCI and to some extent also by MUSE additionally, and the position is consistent with this.

A single correspondence diagram thus enables the recognition of mutual relations among the departments, among the disciplines and also the combined department-discipline relationships, provided that the departments and disciplines themselves are not too cross-disciplinary and complex. A comparison of the correspondence diagrams presented in Gautam (2015, 2016) with that presented in Fig. 3, covering altogether four disciplinary schemes, reveals that the use of any of the schemes provides a fairly good picture on the department-discipline relationship although the details on the mutual relationships differ.

Visualization of department-discipline relationships using stacked bars.

The department-discipline relationships in Fig. 4 are shown using vertical bars depicting the stacked per cent contributions of individual disciplines along the vertical for departments placed along the horizontal. The sequences of disciplines and departments are primarily based on the results of hierarchical clustering described above. The stacked bars enable further appreciation of the clusters of departments engaged in similar disciplines. Data for both (whole and fractional) counting modes are plotted to show that large differences may be possible such that the use of a particular counting mode requires proper justification.

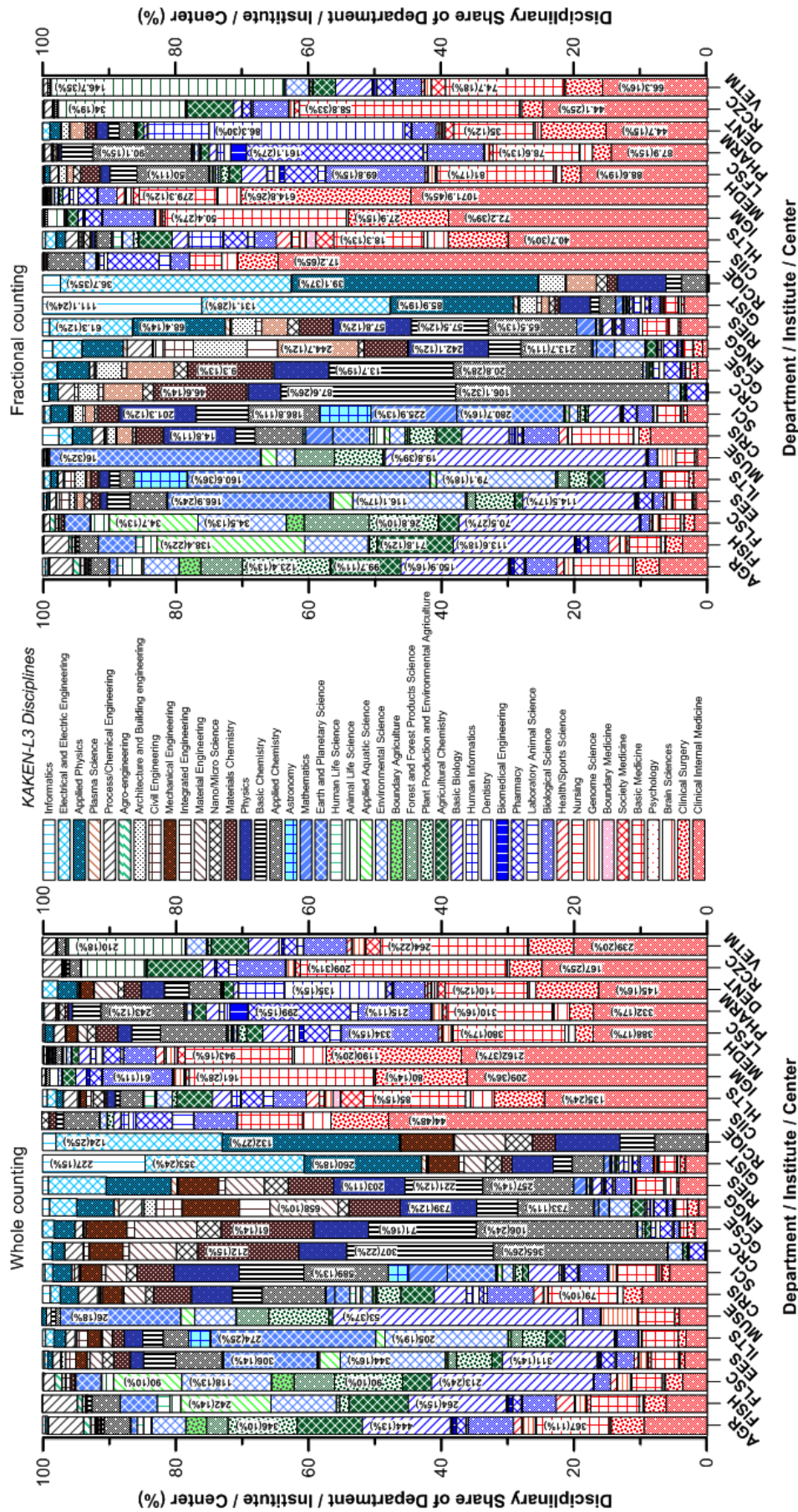


Figure 4. Visualization of department-discipline (KAKEN-L3) relationships as stacked bars. For each department along the horizontal, each segment in the vertical bar represents the % disciplinary share (labels only for 10% or more). Departments and disciplines are placed using the results of cluster analysis. Due to possible multiple disciplinary assignments of each paper, the numbers obtained by whole counting are highly exaggerated. In contrast, fractional counting yields smaller numbers and thus may underestimate the disciplinary contribution. Due to the large variability of the contributions, only the most significant fields of engagement of departments may be distinct.

"Medical, Health and Life Sciences" Publications Disciplinary Assignments by 4 Different Schemes

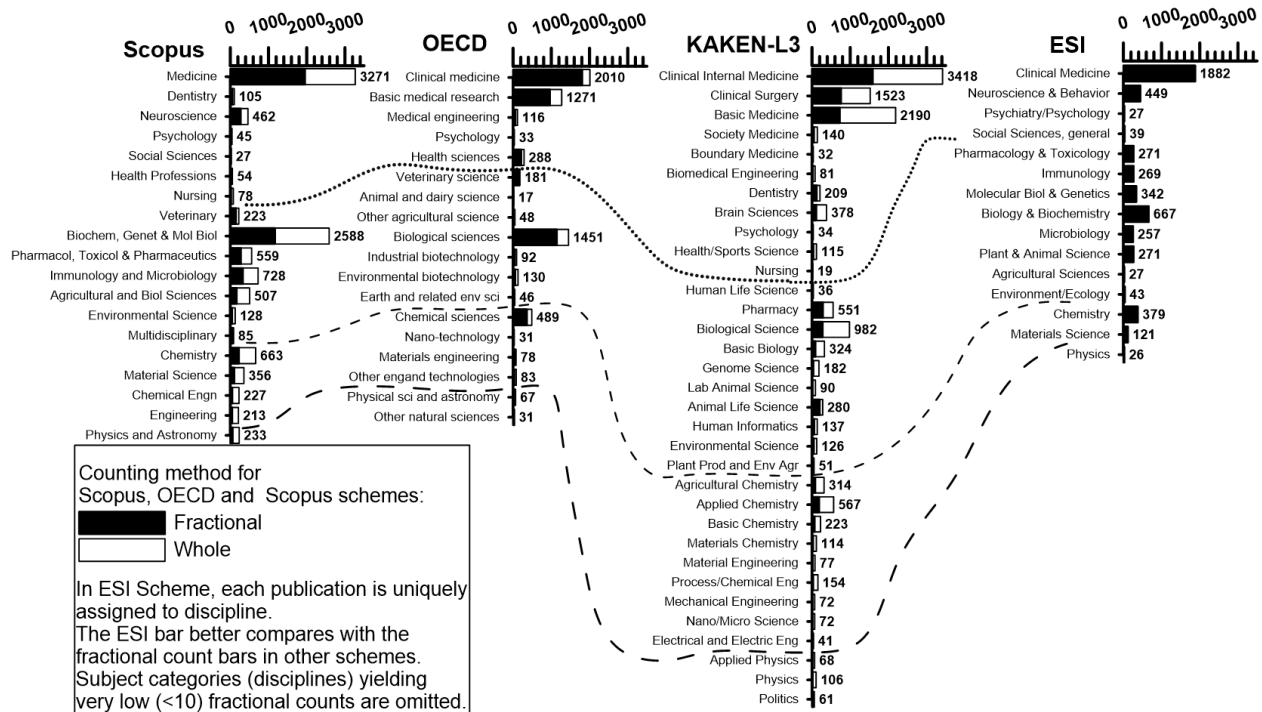


Figure 5. Comparison of the disciplinary assignments of a subset of publications by 4 classification schemes. The numbers of publications for the full length of each horizontal bar in all schemes are whole counts at discipline level. Note that the publications assignments may largely differ among different schemes even for the categories with exactly the same name. The fractional counts plotted as the filled portions of horizontal bars may be appropriate for very rough comparison in terms of the production volume.

Construction of custom publication subsets and their analysis by science mapping.

Three main subsets of publications were considered to test the efficiency of construction of training datasets and to illustrate the methods for science mapping on them using the methodology described in Gautam (2016). The subsets considered are as follows: (i) ‘earth & environmental sciences, agriculture and fisheries’ formed by publications belonging to AGR, FISH, FLSC, MUSE, EES, ILTS, the ‘biology and geoscience’-related partial dataset from SCI, and additional publications from LFSC dealing with ‘biology’; (ii) ‘medical, health and life sciences’ formed by publications belonging to MEDH, HLTS, DENT, IGM, RCZC, VETM, LFSC, PHARM and the partial subset from SCI related to life science; and, (iii) ‘physical sciences, engineering and computer science’ formed by publications belonging to ENGG, GCSE, GIST, RCIQE and partial subset from SCI excluding publications related to geology, life sciences and biology. Partial disciplinary subsets for cross-disciplinary departments were discriminated using a procedure, outlined by the author earlier (Gautam, 2016; Fig. 4), dealing

with an analysis of the ‘earth & environmental sciences, agriculture and fisheries’ subset with 3,725 papers.

Comparative analysis of disciplinary profiles by multiple schemes.

The ‘medical, health and life sciences’ dataset comprised 5,390 publications. The cognitive structure in terms of the subject categories has been established for 5,090 documents, which are assigned to all four classifications (InCites-based (ESI22, OECD Frascati 38, KAKEN-L3 66) and Scopus 27 subject areas schemes) listed in Table 2. Figure 5 shows the results of disciplinary analysis that helps to understand the broad picture regarding the similarities and differences evident from the four schemes. The numbers for each field (times the actual count) for the university obtained by whole counting method are: 5,090 for ESI (x 1 time); 6,520 (x 1.3 times) for OECD; 10,623 (x 2.1 times) for Scopus; and 12,899 (x 2.5 times) for KAKEN-L3 for Scopus. The numbers for individual subject categories for OECD, Scopus and KAKEN-L3 schemes represented by the horizontal bars, therefore, indicate the partial affinity of the documents. It should be noted, however, that the definitions even for seemingly identical categories differ by schemes and therefore direct comparisons are impossible. The length of the filled bar is proportional to the fractional count that may be used to compare the output in seemingly similar categories across the schemes, but each individual number could be higher or lower than the possible realistic value.

For a crude interpretation of the disciplinary profiles within the ‘medical, health and life sciences’ subset, three curved and dotted or dashed lines that broadly differentiate the medical & health sciences, life sciences, chemical sciences & engineering, and physics disciplines are drawn in Fig. 3. Medical disciplines (including dentistry) may be reasonably differentiated from the health sciences disciplines in the three schemes in the left side, but not in the ESI scheme in which the ‘Clinical Medicine’ includes both. It may be practical to calculate bibliometric indicators using one or more of these schemes that best describe the particular dataset analyzed.

Science mapping: Co-words map.

Figure 6 shows the co-words map constructed from keywords occurring in the ‘medical, health and life sciences’ subset. The network is formed by co-occurring keywords related to at least 15 publications, and each cluster has at least 10 nodes. The 5-clusters solution identifies often cross-disciplinary research themes, actively pursued at the concerned university during the 2009-2013 period. The largest cluster plotting in the right side is mainly related to cancer research as well as its care using radiotherapy and chemotherapy but not limited to these topics as shown by the presence of prolific terms such as bone-marrow transplantation and epidemiology. A smaller cluster in the lower part, with terms like scaffolds, hydrogel and collagen, reflects research in regenerative medicine. The other three clusters (clockwise) deal with infections, gene delivery, and protein structure analysis, respectively. Detailed examination of a particular cluster can be made by fully utilizing the visualization capabilities offered by the VOSviewer software (Van Eck & Waltman, 2010). In VOSviewer, one can zoom into particular location of the map to get an enlarged visual representation of the links and strengths among the nodes, and get details

about a particular node (cluster number, number of links, total link strength, number of occurrences) by simply clicking on it (Van Eck & Waltman, 2010). Interpretation of the clusters for titles, topics, etc., may be best done by consulting experts in the relevant fields and after collection of additional information related to the research activities. The science mapping exercise is interactive. It allows the selection of several parameters that best suit the data and the extraction data by refining clustering and visualization that better match the reality.

Medical, Health and Life Sciences Co-Words Map

5 clusters (min occurrence: 15)

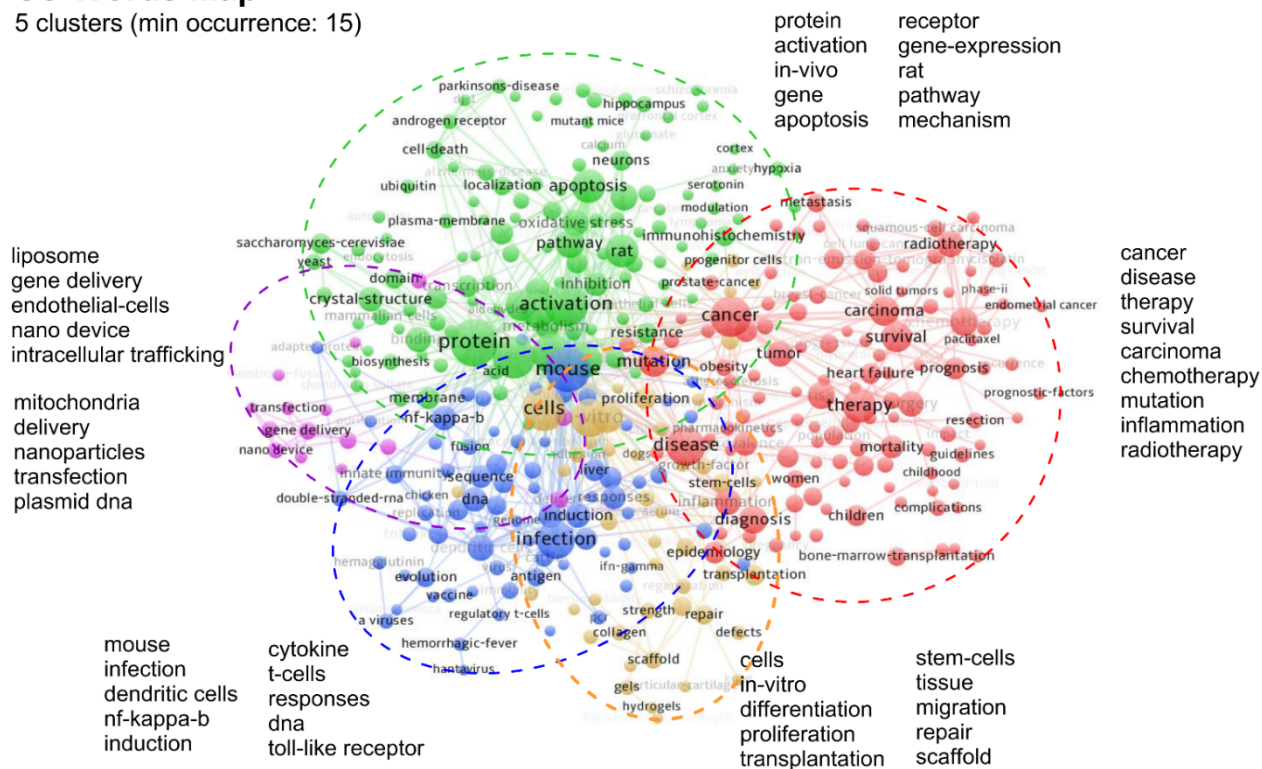


Figure 6. Co-words map constructed from the frequency of co-occurrence of all types of keywords in the 5,390 publications comprising the ‘medical, health and life sciences’ area of the university. For each of the five clusters comprising nodes confined within the ovals, delineated on the basis of keywords co-occurring at least 15 times, 10 most frequent keywords characterizing the research carried out during 2009-2013 are given. Map generated by VOSviewer version 1.6.5 (Van Eck & Waltman, 2010)

Science mapping: Bibliographic coupling map.

A bibliographic coupling map showing the network of journals (sources) in which the university’s publications in the ‘medical, health and life sciences’ area appeared is shown in Fig. 7. As in the co-word map, the closer the nodes (journals) the stronger their mutual relationship. It

is an overlay map in which the nodes (journals) are colored according to the average normalized citation score. The size of nodes, each representing a journal with a minimum of 10 publications, is proportional to the number of publications. Among the 10 most prolific journals, ‘Plos One’ with 163 publications tops the list. Consideration of the citation impact score, however, suggests that ‘Journal of Cell Biology’ tops the list. If the volume and impact are considered together, ‘Proceedings of the National Academy of Sciences of the United States of America’ and ‘Journal of Biological Chemistry’ may be considered to be two very important journals for the university’s community engaged in medical, health and life sciences. If the data are to be used at the level of smaller entities (narrow subject areas, projects, etc.), creation of smaller thematic sets followed by mapping is recommended.

Medical, Health and Life Sciences Bibliographic Coupling (Sources) - Citation Overlay Map

Min occurrence of source: 10

Top 10 journals

(by number of publications)

1. plos one
2. biochem bioph res co
3. j biol chem
4. j vet med sci
5. jpn j vet res
6. biol pharm bull
7. p natl acad sci usa
8. j neurosci
9. circ j
10. neurol med-chir

Top 10 journals

(by average normalized citations score)

1. j cell biol
2. p natl acad sci usa
3. blood
4. biomaterials
5. arthritis rheum-us
6. cancer res
7. j neurosci
8. macromolecules
9. j control release
10. int j radiat oncol

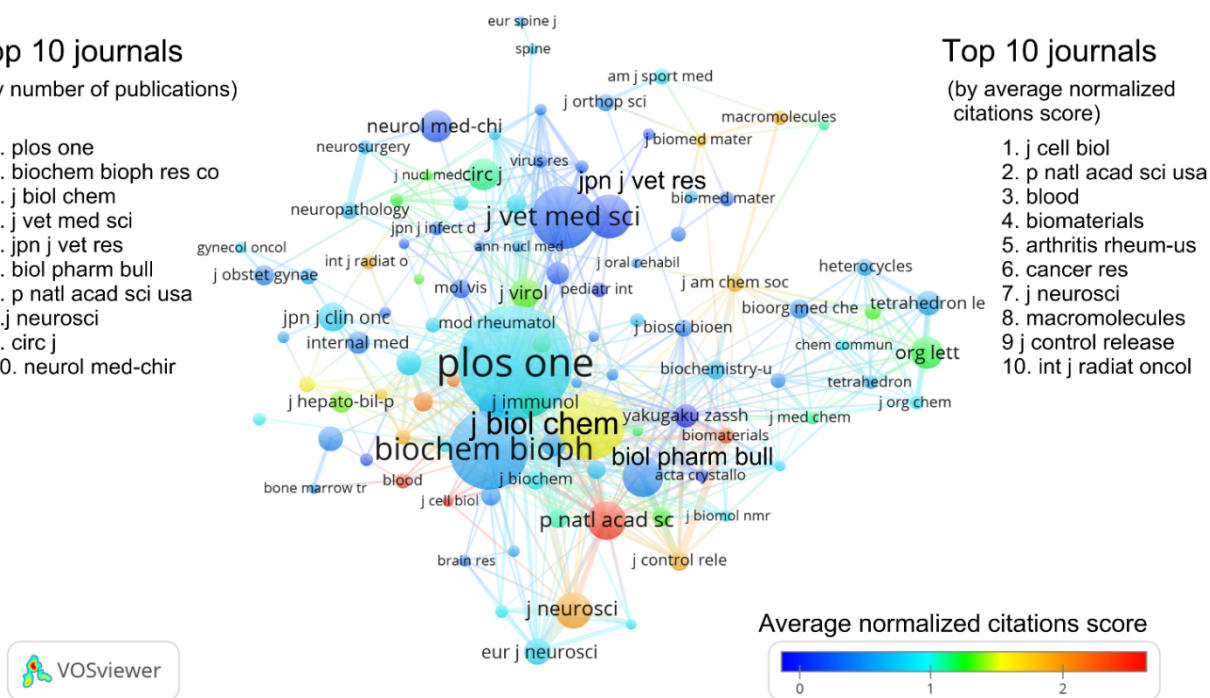


Figure 7. Bibliographic coupling map showing the network of journals (sources) in which the 2009-2013 publications appeared. The size of the nodes is proportional to the number of documents published in the journal. Unlike the network map showing clusters of broadly similar journals with the same color, the nodes (journals) in this overlay map are colored using the normalized citation score. Although the most prolific journal used by researchers in this field is ‘plos one’, it’s citation score is 1 or about the average. In contrast, ‘p natl acad sci’ ranked within top 10 by the number of publications has very high citation score occupying the second position after ‘blood’.

Science mapping: Co-citation map.

A co-citation map showing the 5-clusters network of journals (as sources of publications in references), each of which was cited >200 times internally, is presented in Fig. 8. The cluster types (disciplinary orientations of journals covered) and sizes provide information on the knowledge base used by the researchers authoring the medical, health and life sciences publications. The four most frequently cited journals (Journal of Biological Chemistry, Proceedings of the National Academy of Sciences of the United States of America, Nature, Science) occupying the center of the diagram serve as the knowledge sources of most publications. The list of the remaining 16 frequently cited journals listed besides the diagram includes the most important journals belonging to these clusters. It is noteworthy that journals related to chemistry (including biochemistry) forming the cluster in the right side represent an important knowledge base for the medical, health and life sciences dataset considered.

Medical, Health and Life Sciences Co-Citation (Sources in References) Map

Min citations: 200

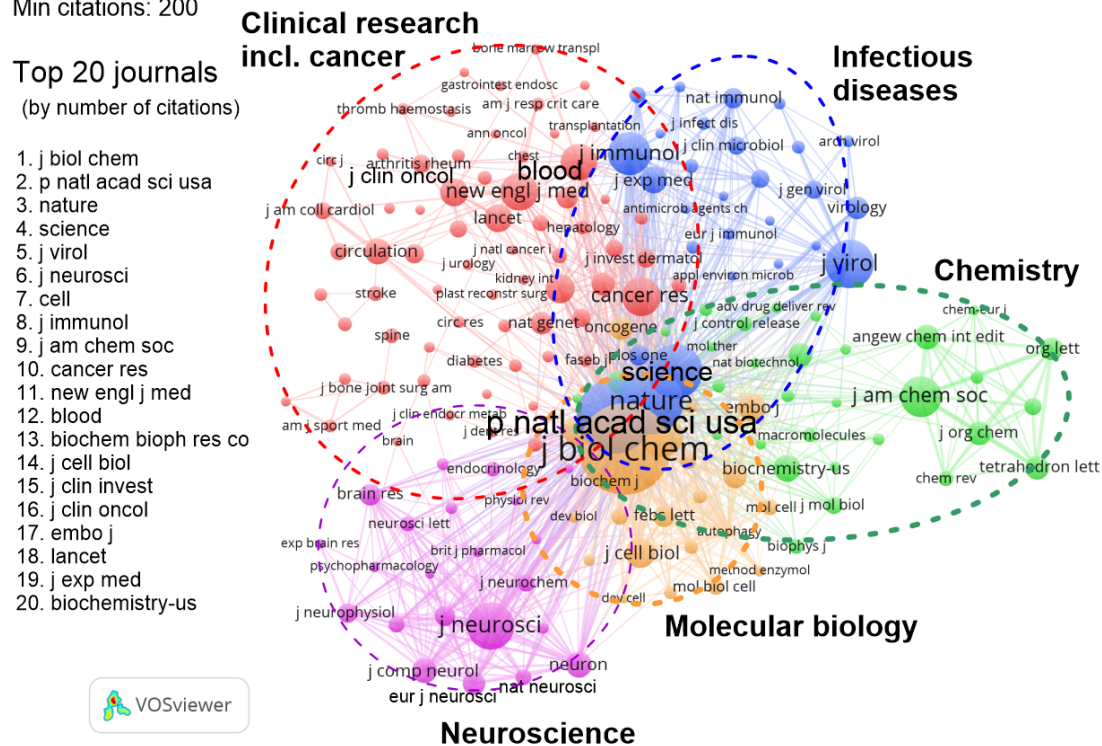


Figure 8. Map showing the co-citation relationship of sources (journals) as the units of analysis. The 5-clusters solution groups journals, cited more than 200 times each, based on their disciplinary orientations (i.e., diverse clinical research including cancer, infectious diseases, chemistry, molecular biology, neuroscience). Four journals that are most cited and possess multidisciplinary character occupy the central and near neutral (connections to journal from almost all clusters) positions despite their greater affinity to a specific cluster. Other journals

with highest impact in specific category (e.g., j neurosci; j am chem soc) can be easily identified in the diagram by their size.

Concluding Remarks

This study introduced a practical approach for comparative disciplinary analysis and science mapping of the scientific publications at department level of a university using a WoS-based dataset of publications in STEMM fields covering a 5-yrs period. It emphasized the importance of having a clear understanding of the publications-based relationships among departments and disciplines through effective visualizations by using the commonly available multivariate techniques and simple diagrams.

Because of the differences in the organizational structures, such relationships differ from university to university at the level of different entities, the application of several disciplinary classification schemes yields information leading to a more comprehensive picture on the research landscape. The step related to the construction of subsets (detailed in section 3.4 above), each of which most effectively reveals the hidden message on research entities in question, requires special care. The science mapping techniques, applied to a custom dataset, illustrated here using commonly available software to visualize the relationships based on keywords and journals in publication titles and references can be extended to the authors, projects and so on using the specific publication subsets which are based on data subjected to disambiguation in terms of relevant research entities (author, affiliating unit, etc.). The information derived using the approach presented above can be combined with other bibliometric indicators, that incorporate the citation data.

It is assumed that the users of the results of the bibliometric/scientometric analysis are either the research entities themselves, the research administrators or the executives. In any case, it is mandatory to verify the correctness of the derived information through input of experts and discussion with the users themselves before considering them while formulating new or re-aligning the existing research strategies.

Limitations of this study.

The publication records considered in this study were extracted from WoS and only those possible to assign to all four disciplinary schemes (ESI, SCOPUS, OECD Frascati, and KAKEN-L3) could be used. Their disciplinary analysis limited to schemes with relatively small number (22-65) of subject categories. In order to explore the subjects at much finer scale, however, utilizing the WoS 250+ and Scopus 300+ ASJC schemes available in WoS and Scopus, respectively, is recommended. Use of the multiple disciplinary schemes always results in better understanding the cognitive structure, possibly inadequately addressed through individual journal-based subject categories scheme, as “articles within a journal do not necessarily share a similar topic or disciplinary perspective” (Rafols, 2014).

Acknowledgements

This study utilized the analytical tools and data made available through the Office for Enhancing Institutional Capacity executing the MEXT funded program at Hokkaido University. However, the author is solely responsible for the approach and opinions related to bibliometric analysis presented in this paper.

References

- Bartol, T., Budimir, G., Juznic, P., & Stopar, K. (2016). Mapping and classification of agriculture in Web of Science: other subject categories and research fields may benefit. *Scientometrics*, *109*(2), 979-996. doi:10.1007/s11192-016-2071-6
- Börner, K., & Polley, D. E. (2014). Replicable Science of Science Studies. In: Ding Y., Rousseau R., Wolfram D. (eds) *Measuring Scholarly Impact*, Springer, Cham, 321-341. doi:10.1007/978-3-319-10377-8_14
- Clarivate Analytics (2015a). <https://clarivate.com/products/> (access to Web of Science and InCites under licensed subscription only)
- Clarivate Analytics (2015b). Master Journal List. <http://ip-science.thomsonreuters.com/mjl/>
- Clausen, S-E. (1998). *Applied Correspondence Analysis: An Introduction*, SAGE Publications.
- De Groot, S. L., & Raszewski, R. (2012). Coverage of Google Scholar, Scopus, and Web of Science: A case study of the h-index in nursing. *Nursing Outlook*, *60*(6), 391-400. doi:10.1016/j.outlook.2012.04.007
- de Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, *98*(2), 1547-1565. doi:10.1007/s11192-013-1089-2
- Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, *83*(1), 243-258. doi:10.1007/s11192-009-0021-2
- Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online Information Review*, *32*(1), 8-21. doi:10.1108/14684520810865958
- Gautam, P., Kodama, K., & Enomoto, K. (2014). Joint bibliometric analysis of patents and scholarly publications from cross-disciplinary projects: implications for development of evaluative metrics. *Journal of Contemporary Eastern Asia*, *13*(1), 19-37.
- Gautam, P. (2015). Deciphering the Department-Discipline Relationships within a University through Bibliometric Analysis of Publications Aided with Multivariate Techniques. 2015 IIAI 4th International Congress on Advanced Applied Informatics, Okayama, Japan, 468-471. doi:10.1109/IIAI-AAI.2015.212
- Gautam, P. (2016). Comparative Analysis of Scientific Publications of Research Entities Using Multiple Disciplinary Classifications. 2016 IIAI 5th International Congress on Advanced Applied Informatics, Kumamoto, Japan, 523-528. doi:10.1109/IIAI-AAI.2016.117
- Gautam, P. (2017). Scientific Publications and World University Rankings: Focus on Bibliometric Indicators at Institution (Hokkaido University) and Department (Dental Medicine) Levels. *Hokkaido Journal of Dental Science*, *38*, 2-15.
- Glänzel, W. (2003). *A Course on Theory and Application of Bibliometric Indicators*, Course Handouts. <https://www.researchgate.net/publication/242406991> (accessed on 2016/8/18).
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787-804. doi:10.1007/s11192-015-1798-9
- Jalali, S. M. J., & Park, H. W. (2018). State of the art in business analytics: themes and collaborations. *Quality & Quantity*, *52*(2), 627-633. doi:10.1007/s11135-017-0522-7

- JSPS (2015). *Application Procedures for Grants-in-Aid for Scientific Research (KAKENHI)*. http://www.jsps.go.jp/j-grantsinaid/22_startup_support/data/27/h27_kensta_koubo_e.pdf.
- Lee, Y.-G. (2013). Multidisciplinary team research as an innovation engine in knowledge-based transition economies and implication for Asian countries. *Journal of Contemporary Eastern Asia*, 12(1), 49-63.
- Leydesdorff, L. (2013). An evaluation of impacts in "Nanoscience & nanotechnology": steps towards standards for citation analysis. *Scientometrics*, 94(1), 35-55. doi:10.1007/s11192-012-0750-5
- Lopez-Illescas, C., de Moya-Anegon, F., & Moed, H. F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. *Journal of Informetrics*, 2(4), 304-316. doi:10.1016/j.joi.2008.08.001
- Meho, L. I., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711-1726. doi:10.1002/asi.20874
- Minasny, B., Hartemink, A. E., McBratney, A., & Jang, H. J. (2013). Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. *Peerj*, 1, 16. doi:10.7717/peerj.183
- Mingers, J., & Lipitakis, E. (2010). Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics*, 85(2), 613-625. doi:10.1007/s11192-010-0270-0
- Mirkin, B. (2011). *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*, Springer.
- Moed, H. F. (2010). *Citation Analysis in Research Evaluation*, Dordrecht: Springer.
- Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, 110(2), 967-990. doi:10.1007/s11192-016-2212-y
- Moed, H. F., Markusova, V., & Akoev, M. (2018). Trends in Russian research output indexed in Scopus and Web of Science. *Scientometrics*, 116(2), 1153-1180. doi:10.1007/s11192-018-2769-8
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228. doi:10.1007/s11192-015-1765-5
- Nederhof, A. J., Meijer, R. F., Moed, H. F., & Vanraan, A. F. J. (1993). RESEARCH PERFORMANCE INDICATORS FOR UNIVERSITY DEPARTMENTS - A STUDY OF AN AGRICULTURAL UNIVERSITY. *Scientometrics*, 27(2), 157-178. doi:10.1007/bf02016548
- Prathap, G. (2013). Benchmarking research performance of the IITs using Web of Science and Scopus bibliometric databases. *Current Science*, 105(8), 1134-1138.
- Prins, A. A. M., Costas, R., van Leeuwen, T. N., & Wouters, P. F. (2016). Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. *Research Evaluation*, 25(3), 264-270. doi:10.1093/reseval/rvv049
- Rafols, I. (2014). Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence. In: Ding Y., Rousseau R., Wolfram D. (eds) *Measuring Scholarly Impact*, Springer, Cham, 169-190. doi:10.1007/978-3-319-10377-8_8
- Santa, S., & Herrero-Solana, V. (2010). Scientific production in Latin America and the

- Caribbean: an approach using the data from Scopus, 1996–2007. *Revista Interamericana de Bibliotecología*, 33(2), 379-400.
- Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430-435. doi:10.1016/j.joi.2018.03.006
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. doi:10.1007/s11192-009-0146-3
- van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In: Ding Y., Rousseau R., Wolfram D. (eds) *Measuring Scholarly Impact*, Springer, Cham, 285-320. doi:10.1007/978-3-319-10377-8_13
- Vieira, E. S., & Gomes, J. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, 81(2), 587-600. doi:10.1007/s11192-009-2178-0
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., . . . Borner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14-26. doi:10.1016/j.joi.2010.06.004
- Williams, R., & Bornmann, L. (2014). The Substantive and Practical Significance of Citation Impact Differences Between Institutions: Guidelines for the Analysis of Percentiles Using Effect Sizes and Confidence Intervals. In: Ding Y., Rousseau R., Wolfram D. (eds) *Measuring Scholarly Impact*, Springer, Cham, 259-281. doi:10.1007/978-3-319-10377-8_12<https://doi.org/>
- Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429-472. doi:10.1177/1094428114562629