Deciphering the Department-Discipline Relationships within a University through Bibliometric Analysis of Publications Aided with Multivariate Techniques

Pitambar Gautam
Creative Research Institution (CRIS)
Hokkaido University
Sapporo, Japan
pgautam@cris.hokudai.ac.jp

***Abstract*— This study explores a practical approach to decipher the department-discipline relationships between the organizational research units dedicated to natural science, technology, engineering & medical (STEM) fields and 22 disciplinary categories used in Essential Science Indicators database (ESI 22 fields), for a Japanese national university as seen in a set of peer-reviewed journal publications (articles & review) indexed in the Web of Science (WoS) Core Collection database for a 5-years period. The procedure involved several steps such as (i) identification of publications of each organizational research unit through disambiguation of the affiliation data; (ii) assigning each publication to the corresponding ESI field based on journal title; (iii) aggregating bibliometric information of all publications per research units and disciplines, and (iv) performing multivariate analysis, e.g., clustering and correspondence analysis, to extract proximity relationships and internal structures that enable regrouping the obtained data and visualizing them using two-dimensional plots and bar diagrams. This approach may be easily adapted for analysis using other available disciplinary (subject areas or categories) schemes. Moreover, such analysis can be further extended to a lower hierarchical level, such as research divisions or research teams comprising a complex multidisciplinary department. The proposed affiliation-based analysis is useful for initial understanding the disciplinary contribution of the university departments to overall research output, e.g., for analysis of ranking based on performance for past 5-6 years tracing past history. It can be easily adapted to the bottom-up research performance analysis (based on current researchers) required for research administration or research strategy formulation based on the research output of the immediate past.**

Keywords—bibliometry; scientometry; multivariate statistics, correspondence analysis; clustering; cross-disciplinarity; ESI fields

I. Introduction

Significant differences exist in the regional and national organizational structures of universities around the globe in terms of the disciplinary coverage. Notwithstanding this, traditional bibliometric/scientometric analyses and university rankings rely largely on indicators aggregated at the level of disciplines dependent on specific databases that index the bibliographic data (including citations) on research publications appearing in peer-reviewed journals [1]. The most popular disciplinary classifications are: (i) 251 subject categories and 22 fields offered by WoS and ESI databases of Thomson Reuters [2]; and (ii) 27 major and 300+ minor subject areas or categories by Scopus database of Elsevier [3]. Although aggregation into several other classification schemes (e.g., OECD: Frascati Fields of Science, UK: RAE Units; Australia: ERA 2012 FOR Levels) used in different parts of the world becomes possible if one has access to more sophisticated analytic tools, this seems to be of limited interest in Japan owing to the peculiar research organizational structure of the Japanese universities, which adhere to the KAKENHI (Grant-in-Aid for Scientific Research) classifications (Categories, Areas, Disciplines and Research Fields) [4]. Assigning research publications to the KAKENHI scheme and mapping the correspondence between the WoS and ESI subject categories, is a topic of much interest for research groups and database vendors in Japan, but it is likely to take some time before this can be materialized into practice. One popular approach adapted by the National Institute of Science & Technology Policy (NISTEP) is to aggregate down publications into STEM fields corresponding to 18 ESI fields to 8-tier subject portfolios and aggressively use for various purposes, e.g., for benchmarking the Japanese Universities for their research and development capacity [5]. This scheme together with ESI 22 fields and information on research competencies in the form of science maps have been increasingly used for placing each Japanese research-intensive university in the national and international context and to judge their relative competitiveness [6].

University research administrators (URAs) require bibliometric indicators and maps derived from publications data for immediate use to aid formulating/planning the research strategy within their university. The objective is provide data to the university executives as well as to the heads of “departments” - organizational units, such as graduate school, research institute or center, each of which has varying level of autonomy and executes research activities according the programs specific to it. In this context, it is vital to be able to clearly decipher the relationships between the departments and disciplines (e.g., the share of each unit to the total contribution

in certain portfolio or subject class), both at the level of broad subject areas as well as much smaller categories.

This study explores a practical approach to decipher the department-discipline relationships as seen from ESI 22 fields (hereafter, disciplines) for a university using a set of publications (articles & review), contributed by organizational research units (faculty, graduate school, departments, research institutes and centers, referred hereafter simply as "department") related to the STEM fields, indexed in WoS database and spanning for a 5-years period.

II. Methods of analysis

A. Extraction of Publication Data

Publications data for a Japanese national university were harvested from the WoS Core Collection (SCI-E, SSCI and A&HCI) using the advanced search and organization-enhanced options (accessed: February 23, 2015. The search was restricted to articles & reviews published during 2009-2013.

B. Data cleaning, and Assigning Publications to Departments and Disciplines

The retrieved bibliographic data were subjected to cleaning that involved identifying publications that didn't belong to the concerned university and excluding them from the final list. After that, each publication was assigned to the relevant department (e.g., AGR for Faculty or Graduate School of Agriculture) by identifying all possible affiliation variants using a semi-automatic worksheet-based matching approach. Likewise, each publication was assigned to unique discipline (e.g., Agricultural Sciences comprising ESI 22 fields) primarily based on journal title following Thomson Reuters' master journal list [7] . The whole-count method was used implying that if a publication was contributed by several departments, it was assigned fully to each of them. As the publication counts recorded in WoS for the non-STEM departments (e.g., economics, law, literature, etc.) of the university are considered to represent a small fraction of their expected total publication output (mostly in Japanese), this analysis is restricted to departments engaged in STEM fields. Furthermore, to ensure the representative size, departments that produced less than 10 papers a year were also excluded.

C. Multivariate Statistical Analysis

To establish the similarity of departments in terms of disciplinary coverage, the percentage share of each discipline, as a size-independent measure, has been used as raw data for multivariate analysis using hierarchical clustering and correspondence analysis using a commercial software (*Excel Statistics for Windows 2012* by SSRI Co. Ltd., Tokyo).

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-cluster solution was obtained to determine the departmental clusters. Correspondence analysis involved visually displaying both row (disciplines) and column (departments) categories of a contingency table P = (pij), i ϵ I, j ϵ J, in such a way that distances between the presenting points reflect the patterns of co-occurrences in P. Theory underlying these analytical methods is found in standard texts, such as [8].

D. Stacked Bar Plots of Department-Discipline Relationships

Data on percent disciplinary shares of departments arranged considering the outcome of clustering and correspondence analyses were used to generate the stacked bar plots. For some departments lacking in predominance of a single or a few disciplines and therefore showing multidisciplinarity [9], further analysis was conducted at the level of smaller research units discriminated by further scrutiny of the affiliation data.

III. Results And Interprettion

A. Tree diagram based on hierarchical Clustering

Fig. 1 shows the tree diagram showing seven clusters of departments, which are identified by abbreviations (with major field of study in brackets) below:

1. AGR (agriculture); FISH (fisheries); FLSC (field-based sciences); VETM (veterinary medicine)
2. RCZC (zoonotic infections)
3. EES (environmental sciences); MUSE (museum); ILTS (low-temperature studies)
4. CIIS (isotope science); MEDH (medicine); HLTS (health sciences); DENT (dentistry); IGM (genetic medicine)
5. CRC (catalytic chemistry); GCSE (chemical sciences & engineering)
6. LFSC (life sciences); PHARM (pharmacology & pharmacy)
7. CRIS (creative transdisciplinary sciences); RIES (electronic science); SCI (natural sciences); ENGG (engineering); GIST (information science & technology); RCIQE (quantum electronics)

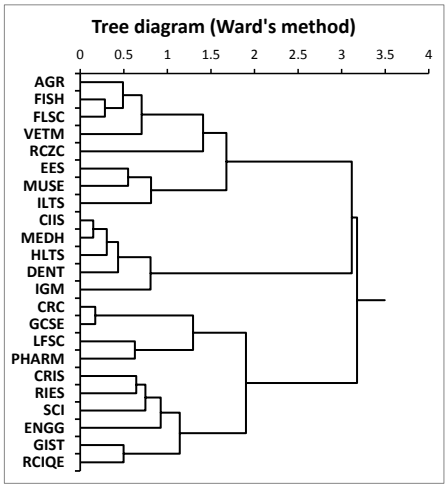

Fig.1. Tree diagram showing the results of hierarchical clustering of departments using agglomertive method with Ward's criterion. For explanation, refer to the text IIIA above.

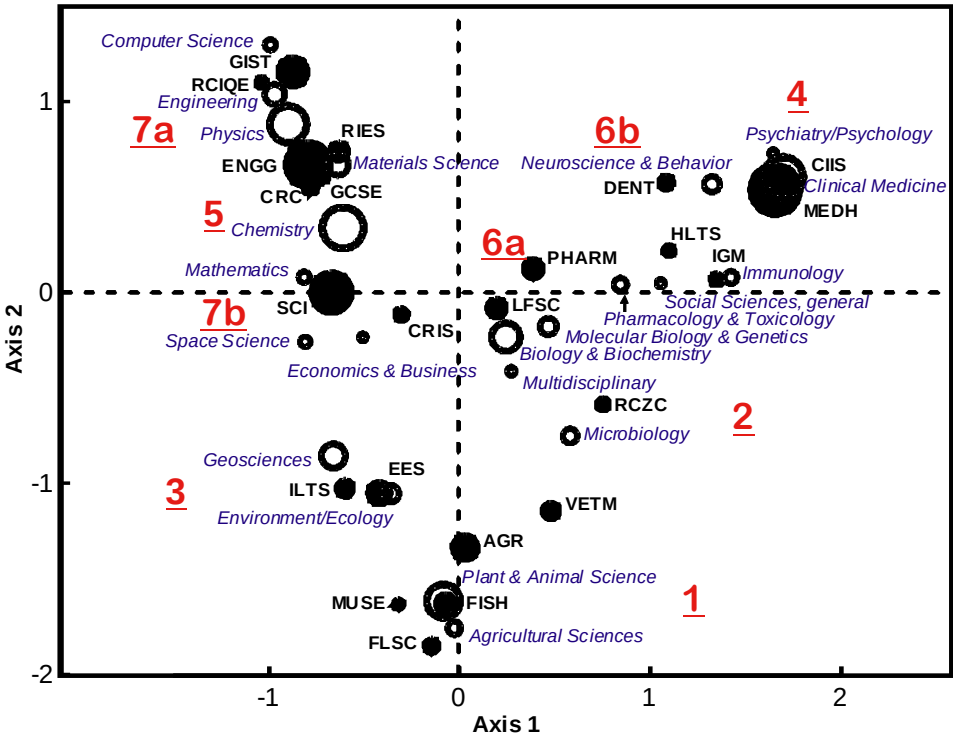

Fig. 2. Visualization of department/discipline contingency data using correspondence analysis. Solid circles stand for departments and open circles for discipline categories. The size of each symbol is roughly proportional to the total number of papers. Underlined numbers almost coincide with those identified as departmental clusters

B. Department-Discipline Relationships from Correspondence Analysis

Fig. 2. shows the 2D representation of the mutual relationship of departments and disciplines 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). The size of each symbol reflects the publication volume, ranging from about 100 (for CIIS and MUSE) to 2800 (MEDH). Underlined numbers that mostly match with the cluster numbers inferred from the tree diagram based on departments (Fig. 1) mark the areas corresponding to refined clusters representing the departments and disciplines. Cluster 7 here comprises two distinct sub clusters: 7a - comprised by Computer Science, GIST, RCIQE, Engineering, Physics, ENGG, RIES and Materials Science; and 7b - comprised by (i) SCI surrounded by Chemistry, mathematics and Space Science, and (ii) CRIS, which may be considered to be surrounded by SCI and clusters 5, 6a, 3, 5 revealing true multidisciplinary nature of this department dealing with natural sciences. Cluster 2 is defined by RCZC, Multidisciplinary and Microbiology.

C. Visualization of Department-Discipline Relationships using Stacked Bars

Stacked bar diagram in Fig. 3 attempts to better visualize the department-discipline relationships and readily reveal the clusters. Four major departmental clusters marked by predominant (>20%) disciplinary contributions of Plant & Animal Science, Clinical Medicine, Chemistry, and Physics, respectively, can be identified from left to right. Other less prominent clusters marked by the closely grouped identical patterns are indicated at the top. Closer inspection of Fig. 3 reveals that three departments (SCI, ENGG and CRIS) are highly multidisciplinary while the maximum contribution of a single discipline is only 20-23%. Deciphering the details of such departments need further analysis as will be demonstrated below in the example of SCI.

D. Detailing the Intra-Departmental Relationship between Research Groups and Disciplines

Despite the notable changes in the organizational structure of SCI during the last decade (e.g., formation of a separate life science department; reorganization of sections dealing with earth sciences and developmental biology into a new natural history sciences department, similar but varying changes in sections dealing with physics, planetary sciences and chemistry), the affiliation records in publications do not always reflect these changes as some authors still adhere to the affiliations used earlier even after completely moving to newly formed departments or sections. So, affiliation-based analysis has certain limitations. Considering the past organizational structure, all publications from SCI were distributed to several disciplinary groups, some of which existed independently earlier but not at present. The outcome shown in Fig. 4 reveals additional details useful for understanding the disciplinary contributions. It also points to the most multidisciplinary nature of BIOSCIENCE Group is obvious. Similar analysis could be extended to other multidisciplinary departments to establish their true nature.

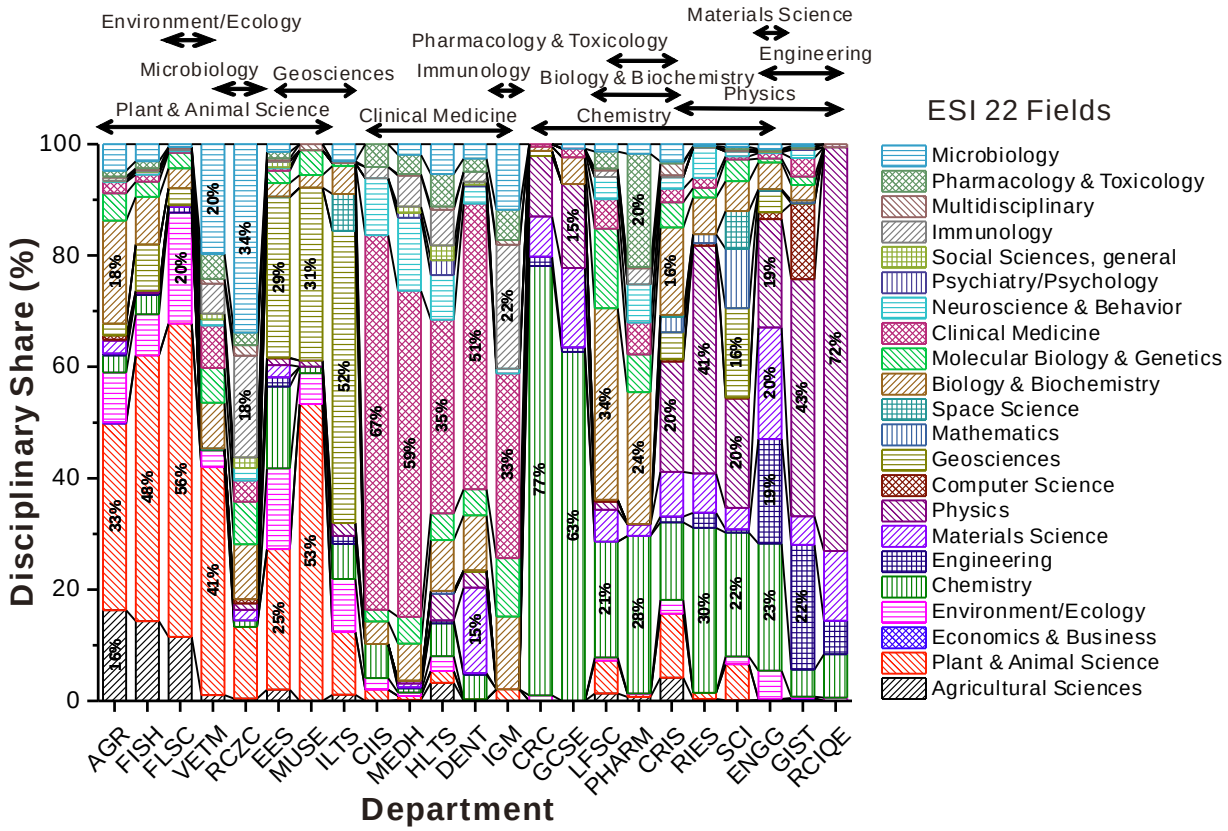

Fig. 3. Visualization of department-discipline relationships as stacked bars. For each department placed along the horizontal axis, the percent share of disciplines is shown by vertical bars. Prominent clusters and major contributory disciplines are further shown by arrows at the upper part. The most bvious contributors are: (i) Plant & Animal Science, in the left; (ii) Clinical Medicine, in the middle, (iii) Chemistry, in the lower right half; and, (iv) Physics, middle to top right.

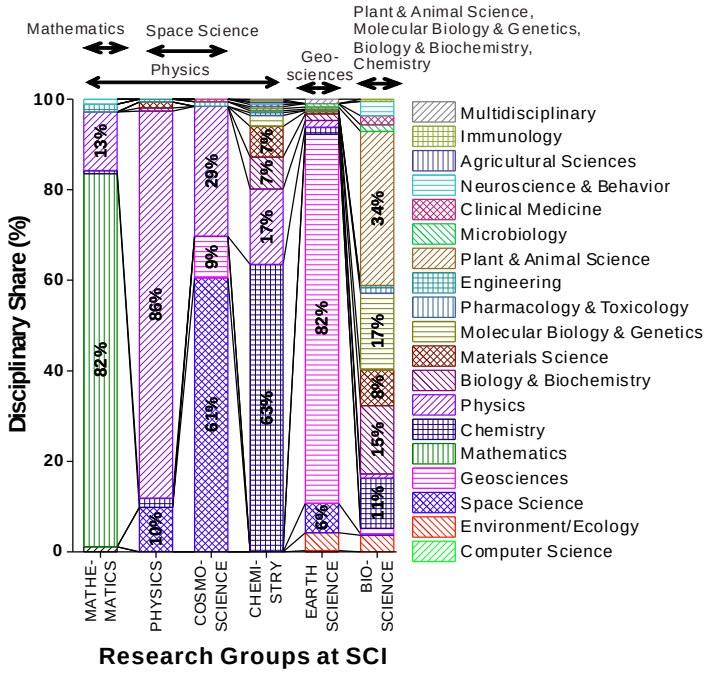

Fig. 4. Visualization of the disciplinary contribution of publications at the level of research groups within ta single department that is engaged in research in multiple disciplines

REFERENCES

[1] E. S. Vieira and J. A. A. F. Gomes, "A comparison of Scopus and Web of Science for a typical university," Scientometrics, 81(2), pp. 587-600, 2009.

[2] Thomson Reuters, Web of Science – IP & Science. http://wokinfo.com/

[3] Elsevier, Scopus. http://www.elsevier.com/online-tools/scopus

[4] Japan Society for the Promotion of Science, Application Procedures for Grants-in-Aid for Scientific Research –KAKENHI-. 2015. http://www.jsps.go.jp/j-grantsinaid/22_startup_support/data/27/h27_kensta_koubo_e.pdf .

[5] A. Saka and T. Kuwahara, "Benchmarking Research & Development Capacity of Japanese Universities 2011," NISTEP/MEXT Report, Japan, 2012.

[6] A. Saka and M. Igami, "Science Map 2010&2012 -Study on Hot Research Area (2005 - 2010 and 2007 - 2012 By Bibliometric Method," NISTEP REPORT No. 159, 2014.

[7] Master Journal List – IP & Science - Thomson Reuters, http://ip-science.thomsonreuters.com/mjl/

[8] B. Mirkin, Core Concepts in Data Analysis: Summarization, Correlation and Visualization. Springer, 2011.

[9] C. S. Wagner, J. D. Roessner, K. Bobb, J. T. Klein, K. W.. Boyack, J. Keyton, I. Rafols, and K. Börner, "Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature," Journal of Informetrics, 165, pp. 14-26, (2011).