# A BIBLIOMETRICS APPLICATION FOR EVALUATING CONTRIBUTION OF A RESEARCH INSTITUTE TO SCIENCE

**H. Nakamura[1], Y. Kajikawa[2], S. Suzuki[1]**
**1 Center for Aviation Innovation Research, The University of Tokyo, Tokyo, Japan**
**2 Graduate School of Innovation Management, Tokyo Institute of Technology, Japan**

**Keywords**: *Technology Landscape, Bibliometrics, Citation Analysis, Patent Analysis*

## Abstract

*It is needed for a research institute to understand how the works of the institute contribute to the growth of different scientific domains. However there are not many methods to evaluate it. This paper create a landscape of scientific domains related to an aviation research institute, taking a bibliometrics approach and focusing on papers citing works of the institute. With this approach, we try to highlight works of the institute in a certain scientific domain and to evaluate the ratio of the leadership in the domain.*

## 1 Introduction

It requires a long time and huge investment for a scientific discovery or technological invention to be matured enough to be applied in an aircraft system. Today's works at aeronautics research institutes have huge importance for the future of the aviation industry.

Therefore, research institutes, whichever it depends on public or private funding, need to make it clear their research strategy, capability and outcomes to the stakeholders. And to do that, research institutes often present following statistics to the public; what facilities the institute owns, how much and where the institute got funding, how many researchers the institute employs, how much collaboration the institute has with other institutes and industries, and how many publication and oral presentation are made from the institute. While these information are all very important, it is difficult to show the evidence of contribution of institutes to science and industries, that is the main object of a research institute.

Proceeding and academic papers are one of very tangible indicators of such contribution. Besides of information about funding, collaboration of authors of different institutes, impact to other researchers is appeared in the form of forward citation. In practice, only the outer layer of paper information is used in the evaluation of works of a research institute, such as the number of publication and sometimes the impact factor of the journal that the works are published. Using publication information is getting difficult because the number of publication is increasing exponentially and science is getting more complex to take simple statistics.

There are research groups to use information of publication effectively in information science. Computer-based approach is compatible with large scale of data and citation analysis has established itself as one of the most effective approaches in structuring academic documents. A citation can be regarded as the self-organizing dynamics of scholars' communication [1-4]. There are approaches to create a scientific landscape based on the communication. Researchers and decision-makers related to science and innovation management evaluate citation analysis approaches as quantifiable and objective approaches that can compensate and validate the experts' judgments [2-3], and can be used in administration fields. Therefore we would like to use this approach for managing an aeronautics research institute.

Adopting such a computer-based citation analysis approach, we would like to create a landscape of scientific domain related to a research institute based on the citation information of papers citing papers of the institute. We assume that such an approach can make the broad and various influence of the institute to the science visible.

## 2 Methodologies

### 2.1 Citation Analysis

This paper followed the approach used in the previous paper [2]. Citation network analysis was employed to structure scientific papers and create a scientific landscape. The analysis procedure is schematically illustrated in Fig. 1.

The data are converted into a non-weighted, non-directed network in which a paper is represented as a node and citations as links. The maximum connected components (MC) of the network are extracted. Papers not citing other papers in the MC were regarded as digressional from the mainstream of those research fields and eliminated them. Finally, the network was divided into clusters using a topological clustering method [5-6].

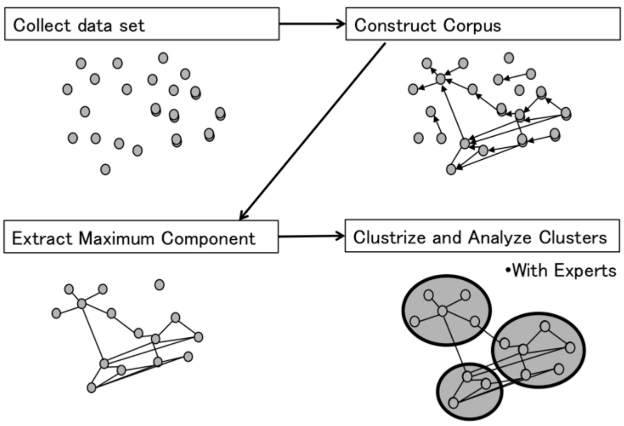



Fig. 1. Schematic diagram of the citation network analysis

After clustering the network, each cluster was characterized by an expert-based approach. To identify the characterization of clusters and to title them, we reviewed the abstracts of core papers that have frequent citations with the other papers in the cluster, frequently appearing words and the frequently appearing journals of each cluster with experts.

The landscape of retrieved clusters is visualized by a large graph layout (LGL) [7]. LGL is based on a spring layout algorithm where links play the role of spring connecting nodes. As a result of this layout, the group of patents citing each other is located in closer positions and only the intra-cluster links for each cluster are shown with the same color. The position of each cluster is intuitively understandable.

### 2.2 Data

This paper collected papers from the Science Citation Index (SCI) and the Social Sciences Citation Index (SSCI) compiled by the Institute for Scientific Information (ISI). The Web of Science, which is a Web-based user interface for ISI's citation databases, is used.

Firstly, we retrieve data of a certain institute by searching the institution in the organization field of the database. Secondly we exclude papers classified to "Astronomy and Astrophysics" in Web of Science Categories. This process is added because we aim to focus on a works related to aeronautics researches and one of focused institutes in this research had a big cluster of astronomy and astrophysics papers. This process is not necessary for other purpose. Thirdly, we search papers that refer papers retrieved in the previous process, using the citation report function of the database. This process is for investigating the contribution of the institute toward various science domains.

We chose Japanese Aerospace Exploration Agency (JAXA) and German Aerospace Center (DLR) to create the landscape as a case study. 5,902 JAXA papers were found in Web of Science in May 2013 after the Web of Science category filtering. 37,819 papers cited the JAXA papers above at this time. On the other hand, 7,387 DLR papers were found in the Web of Science in June 2014 after the Web of Science category filtering. 62,134 papers cited the DLR papers above at this time.

## 2.3 Indicator of Leadership

After creating a scientific landscape of papers citing papers of the focused institute, calculate the ratio of the number of papers published by the focused institute to the total number of papers in the cluster $s$, namely $I_s$. $I_s$ is translated as the rate of the involvement of the focused institute with each sub-domains. Small $I_s$ indicates the leadership position of the institute in the $s$ domain. Large $I_s$ indicates the uniqueness of the institute in the $s$ domain. Table 1 is our hypothesis about what $I_s$ can indicate.

Table 1. Our hypothesis about Indicator $I_s$

| | Scientific domains that the research institute originally targeted | Scientific domains that the research institute originally not targeted |
|---|---|---|
| Small $I_s$ | Leader | All-Rounder |
| Large $I_s$ | Pioneer | Lost Child |

## 3 Case Studies

### 3.1 Japanese Aerospace Exploration Agency (JAXA)

Fig. 2 is the landscape created by citation analysis on papers cite the JAXA papers. The MC was 40,576 and the average published year was 2004.1.

The MC were structured into 135 clusters. Top 32 clusters, that are related to aeronautics domains such as Fluid Dynamics, Structure and Materials, were investigated. Even though JAXA's paper categorized in the "Astronomy and Astrophysics" in Web of Science Categories were excluded, major scientific domains that JAXA influence seems the area of astronomy.

Cluster 3 is a group of papars, 4,716 papers in total, on Computer Fluid Dynamics (CFD). The average published year of the Cluster 3 papers is 2002.1. The involvement of JAXA to Cluster 3, $I_3$ is 14%. At Fig. 3, we took the same citation analysis approach to Cluster 3 and analyzed the sub-clusters. Top sub-clusters is plotted in a graph; the vertical axis is $I_s$ and

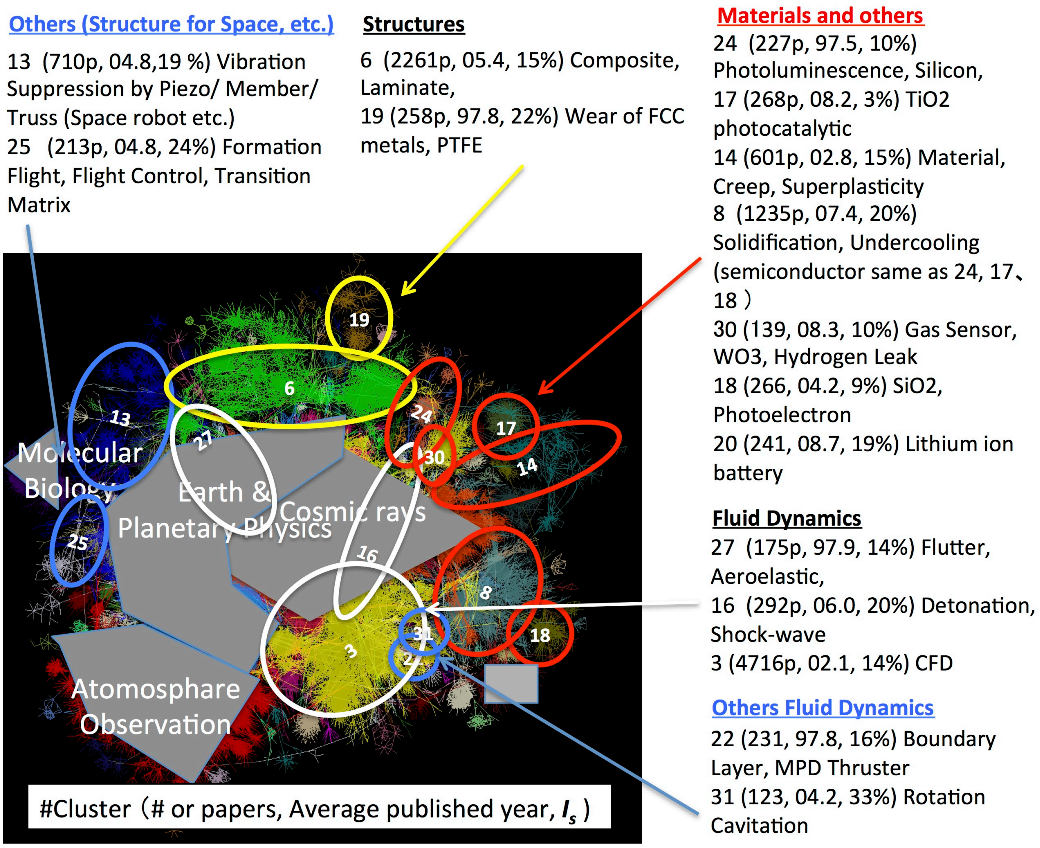



Fig. 2. The landscape of scientific domains where an institute influence (JAXA)

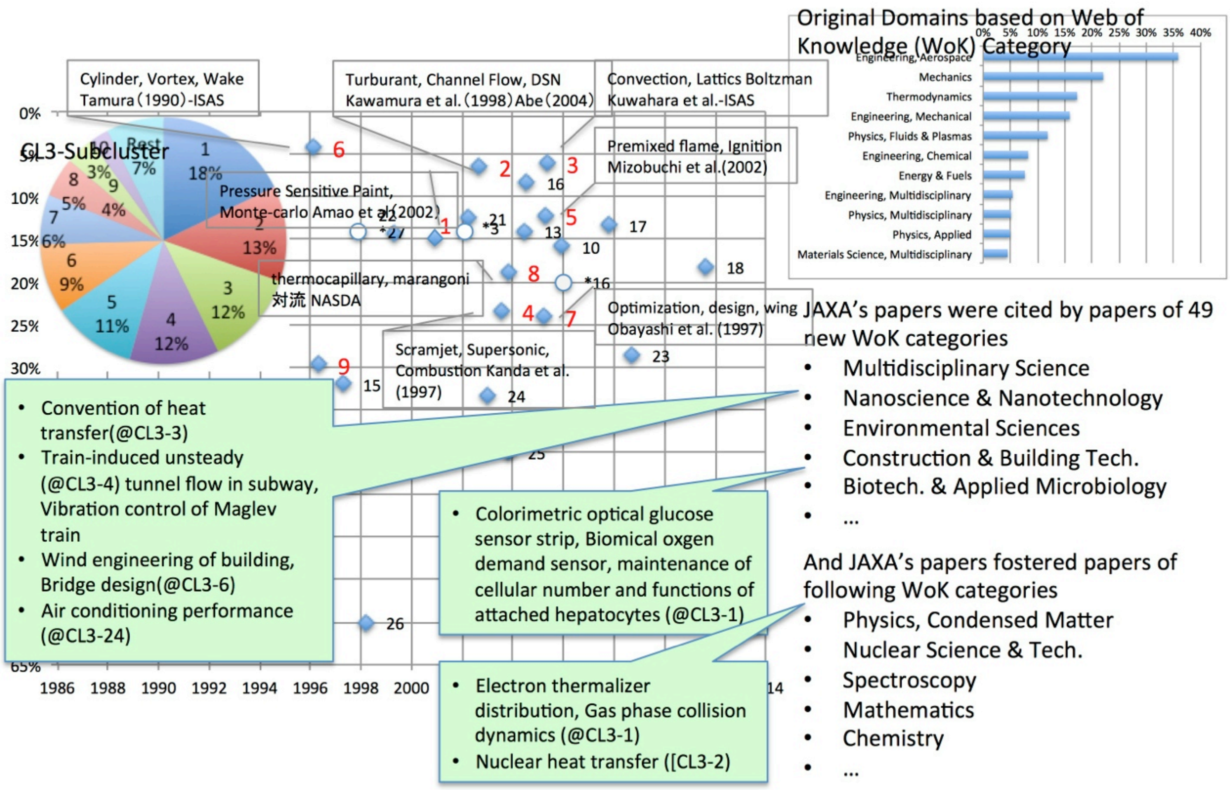



Fig. 3. A snap shot of an influence analysis based on the citation analysis approach (JAXA CL3)

the horizontal axis is the average published year of papers in each sub-cluster. The characteristics of each sub-cluster and core JAXA papers are shown in the graph.

For example, papers in the sub-cluster 3-3 discuss convection and Lattics Bolzman. $I_{3\text{-}3}$ is relatively small and the average published year of the sub-cluster 3-3 is relatively young, which can indicate that JAXA takes leadership in this area and this area is growing.

Fig.3 also shows another perspective that our approach can bring to evaluate works done by the institute. It compares the Web of Knowledge Categories of JAXA papers and the rest in Cluster 3. There are 49 Web of Knowledge Categories, which are not originally covered by JAXA's work. The CFD papers of JAXA are cited not only by other aeronautics CFD papers but also referred by papers of nanoscience and biotech domains. These analyses have possibility to contribute for better understanding of value of the researches in an institute at the wider perspective than at only the aeronautic perspective.

## 3.2 German Aerospace Center (DLR)

Fig. 4 is the landscape created by citation analysis on JAXA academic papers and papers cite the JAXA papers. The MC was 66,836 and the average published year was 2007.7.

MC were structured into 195 clusters. The landscape of DLR seems quite different from JAXA's landscape. However, the size of each clusters appeared on Fig. 2. and Fig. 4 are not coherent so that we need to conduct sub-cluster analysis if we compare two institutes. For example, CFD clusters weren't cound in top clusters until we analyzed sub-cluster of the cluster 1 in Fig. 4. Cluster 10 is a group of papars, 1,237 papers in total, on Robotics. The average published year of Cluster 10 papers is 2008.9. The involvement of DLR to the Robotic cluster, $I_3$ is 10%. Fig. 5 is the snapshot of Robotics sub-cluster analysis. The Robotics works in DLR are influencing divers scientific domain such as medical domains.

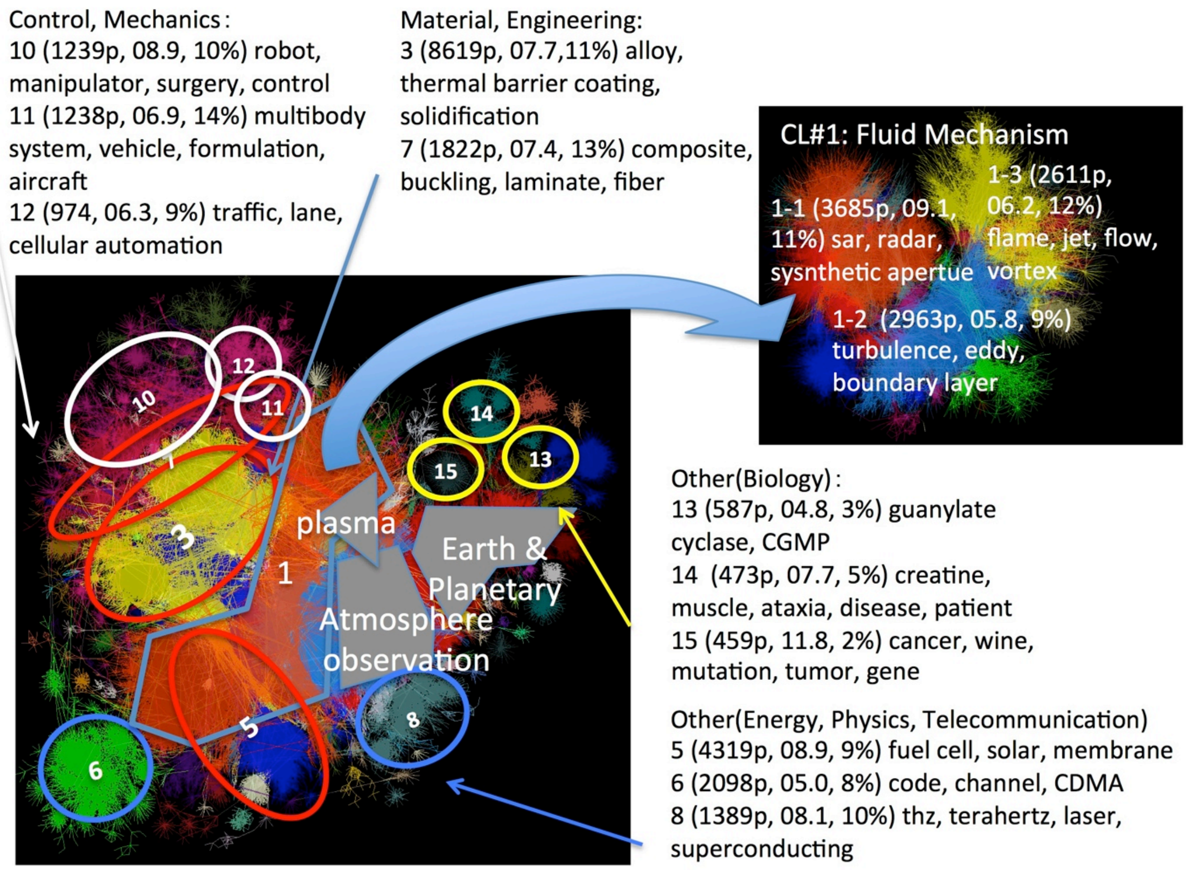



Fig. 4. The landscape of scientific domains where an institute influence (DLR)

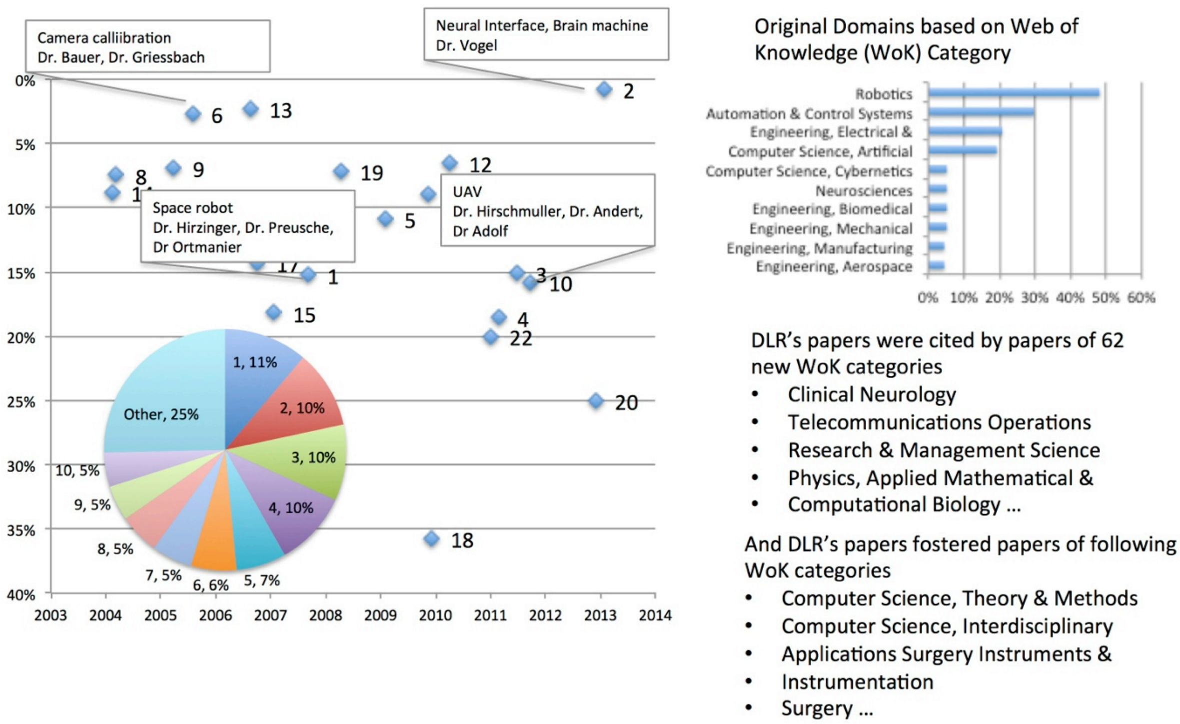


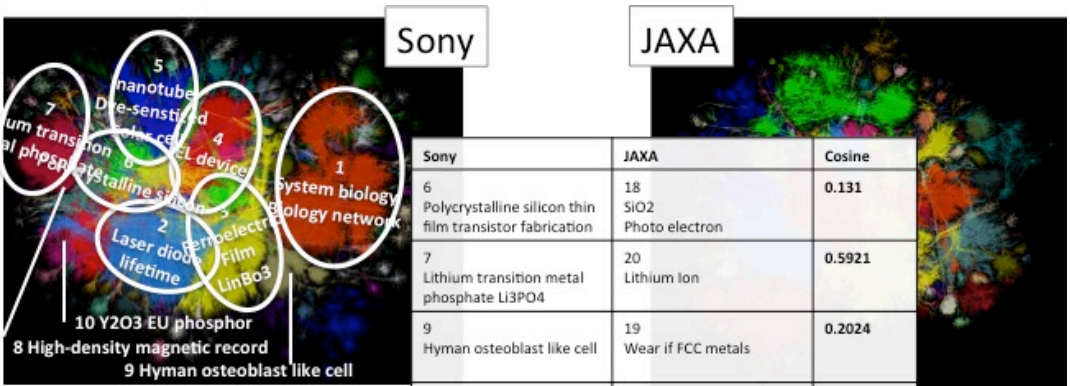

Fig. 5. A snap shot of an influence analysis based on the citation analysis approach (DLR CL10)

### 3.3 Similarity Between Two Institutes

This section is additional and we adopted similarity measurements to discuss potential collaboration between two different institutes. We tested two cases for comparing JAXA and Sony. One case (A1) is to create a landscape from academic papers published by JAXA and Sony Institute together, using our citation analysis. Then we regard clusters which include papers of both institute as common research interest domains. Another case (A2) is to create landscape of JAXA and Sony separately, following the approach described in the section 2. Then we measure the similarity of frequent words appeared in each cluster and highlight clusters with high similarity. More detail explanation of two approaches and cosine


| | Num | Sony | JAXA | | S*J |
|---|---|---|---|---|---|
| 1 | 12332 | 11370 | 962 | 92% | 0.27 |
| 2 | 8916 | 54 | 8862 | 1% | 0.08 |
| 3 | 8103 | 27 | 8076 | 0% | 0.06 |
| 4 | 7771 | 5550 | 2221 | 71% | 0.45 |
| 5 | 7297 | 7262 | 35 | 100% | 0.07 |
| 6 | 4530 | 51 | 4479 | 1% | 0.11 |
| 7 | 4001 | 24 | 3977 | 1% | 0.08 |
| 8 | 2904 | 2661 | 243 | 92% | 0.28 |
| 9 | 2634 | 2450 | 184 | 93% | 0.25 |
| 10 | 1442 | 45 | 1397 | 3% | 0.17 |
| 11 | 1314 | 1299 | 15 | 99% | 0.11 |
| 12 | 1270 | 159 | 1111 | 13% | 0.33 |
| 13 | 1140 | 1 | 1139 | 0% | 0.03 |
| 14 | 905 | 209 | 696 | 23% | 0.42 |
| 15 | 740 | 0 | 740 | 0% | 0.00 |
| 16 | 648 | 98 | 550 | 15% | 0.36 |
| 17 | 620 | 491 | 129 | 79% | 0.41 |
| 18 | 533 | 102 | 431 | 19% | 0.39 |
| 19 | 482 | 473 | 9 | 98% | 0.14 |
| 20 | 463 | 208 | 255 | 45% | 0.50 |

Fig. 6. A1 Approach to Highlight Potential Common Research Interests Between Institutes




| Sony | JAXA | Cosine |
|---|---|---|
| 6 Polycrystalline silicon thin film transistor fabrication | 18 SiO2 Photo electron | 0.131 |
| 7 Lithium transition metal phosphate Li3PO4 | 20 Lithium Ion | 0.5921 |
| 9 Hyman osteoblast like cell | 19 Wear if FCC metals | 0.2024 |
| 15 Organic contamination on silicon wafers | 18 SiO2 Photo electron | 0.0946 |
| 17 Gate etching plasma damage process | 2 Magnetic Reconnection | 0.1893 |
| 17 Gate etching plasma damage process | 9 Ion-acoustics shock in a plasma | 0.2451 |
| 17 Gate etching plasma damage process | 18 SiO2 Photo electron | 0.1305 |

Fig. 7. A2 Approach to Highlight Potential Common Research Interests Between Institutes

similarity measurement and is cound in [8] and available online.

Fig. 6 shows the results of A1 and Fig. 7 shows the results of A2. The table is the potential areas that two institutes have common interests. These analyses can be use for investigating new collaboration possibility between organizations of different domains.

## 4 Concluding remarks

The authors thought it should be useful if institutes can evaluate how much the works by the institute contribute to the scientific domain, demonstrated a citation analysis and investigated landscapes of scientific fields JAXA and DLR influence. Here their works on Astronautics and Astrophysics were excluded in the analysis because originally we wanted to focus on the influence of their aeronautics works. However we think the process is not sophisticated enough. On the other hand, the results showed the wide range of scientific domains that two institutes influence. This approach also made it possible to highlight works of which employees are core in the retrieved scientific domains.

An indicator $I_s$, the ratio of works of a certain institute in a scientific domain is designed in this research in order to evaluate whether the institute takes leadership or display the uniqueness. We plotted retrieved scientific clusters by $I_s$ and the average of published year of papers found in each cluster. We couldn't set a threshold os small and large $I_s$ in this paper because the citation behavior depends on scientific domain and also because experts' knowledge is required to evaluate this approach. Unfortunately, this research lacks evaluation of the results. Interviews with people in demonstrated institutes are necessary and we should ask not only whether the results are similar to their intuitions, whether such perspective is necessary and useful in their institutes, but also what should be a challenge in order to adopt this approach in practices within institutes. We would like to open these questions to the readers and hope to receive responses during the oral presentation. In any case, discussion with practitioners will be appeared in our future research.

The circumstance of research institutes are changing and to be competitive, they must appeal their works to public. However, as a long time is required for a technology to be matured for the aviation system, aeronautic research institutes don't have tangible evidences to show their contribution to science and industry. We found some useful tools developed in a different domain, that is, information science domain, and studied how we can adopt such an approach in practical use, wishing to contribute toward fostering aeronautics research.

## References


[1] Cronin B. (2001) Bibliometrics and beyond: some thoughts on web-based citation analysis, *Journal of Information Science*, 27, 1, p.1-7, 2001

[2] Kajikawa Y et al.. Structure of knowledge in the science and technology roadmaps. *Technological Forecasting & Social Change*, Vol. 75, pp 1-11, 2008.

[3] Nerur S.P., Rasheed A.A., Natarajan V. The intellectual structure of the strategic management field: An author co-citation analysis, *Strategic Management Journal*, 29, p.319-336, 2008.

[4] Nakamura H, Kajikawa Y, Suzuki S. Science and Technology Map Analysis of Aerospace Engineering. *Proc Conference The 28th International Congress of Aeronautical Science*, Brisbane, Australia, 2012.

[5] Newman, MEJ. Fast algorithm for detecting community structure in networks. *Phys. Rev.*, 69, 066133, 2004.

[6] Newman, MEJ, Girvan, M. Finding and evaluating community structure in networks, *Phys. Rev.*, 69, 026133, 2004.

[7] Adai, AT, SV Date, S Wieland, and EM Marcotte. LGL: creating a map of protein function with an algorithm for visualizing very large biological networks. *Journal of molecular biology,* 340(1), 179-190, 2004.

[8] Nakamura H, Kajikawa Y, Suzuki S. Science and Technology Map Analysis of Aerospace Engineering. *Proc Conference The 28th International Congress of Aeronautical Science*, Brisbane, Australia, 2012.


## Contact Author Email Address

techhn [at] mail.ecc.u-tokyo.ac.jp

## Copyright Statement