Transactions on Machine Learning
and Data Mining
Vol. 6, No.1 (2013) 3-18
ISSN: 1865-6781,
ISBN: 978-3-942952-24-8

ibai Publishing
www.ibai-publishing.org

The Study of the Role Analysis Method of Key Papers in the Academic Networks

Akira Otsuki[1] and Masayoshi Kawamura[2]

[1] Corresponding author, Tokyo Institute of Technology, 2-12-1, Ookayama, Meguro_Ku, Tokyo 152-8550, Japan,
cecil2005@hotmail.co.jp
[2]The University of Tokyo, 2-11-16, Yayoi, Bunkyo_Ku, Tokyo 113-8656, Japan

ABSTRACT. In this study, we identified the problems of applying Guimera et al.'s [9] methods to a target network of articles in academic journals. Guimera et al. proposed both a clustering method and a role analysis model based on clustering. In concrete terms, they defined a Z-SCORE (Zi) and participation coefficient (Pi) as targets for metabolic networks. Although Guimera et al. methods were intended for application to metabolic networks, we believe they can be adapted to the citation networks formed by academic articles. We then proposed a new role analysis method and visualization system as a target of the academic article networks. Specifically, a unique algorithm is used to extract key articles from within clusters, after which role analysis is performed. The results are then evaluated by examining the availability of given academic articles. Finally, we performed a comparative evaluation of our method. Results showed that our method was able to show the movement of key paper innovation more clearly than Guimera et al.'s method.

Keywords: Role Analysis Method, Clustering, Citation Analysis, Academic Landscape, Database

1 Introduction

Big Data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus, according to research by McKinsey's Business Technology Office [1]. But mass aggregations of data are not of much use without context and meaning. For example, what does a given dataset mean for the present and future, and how can it be used to address current and future problems? To answer such questions, we often need to extract some of the context of the data through analysis. For example, by analyzing the data contained in related academic journals, we can extract a high-level view of the research landscape covered by those

journals, and apply more specific analyses thereafter. A number of approaches to such high-level analysis have been proposed, For example, "Centrality" [2] is within the scope of graph theory and network analysis, determine the relative importance of a vertex within the graph. "Small World" [3] is the hypothesis that anyone in the world relatively easily linked if we trace the acquaintance relationship. "Structural Holes" [4] describes a set of new measures based on ego networks, and the purpose of Structural Holes is to clarify how to compute the redundancy measures. Finally, "Clustering" [5-8] is the method of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

Guimera et al. [9] proposed both a clustering method and a role analysis model based on clustering. In concrete terms, they defined a Z-SCORE (Zi) and participation coefficient (Pi) as targets for metabolic networks. A metabolic network is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. And then expressed Pi as the X-axis and Zi as the Y-axis of a graph. This graph shows Role-specific regions in the zP parameter space. They then showed that metabolic nodes can be classified into seven different roles. When the metabolic nodes will be put on this graph, we will be able to understand each nodes role. For example, one node will do role of hub about another cluster nodes. But to apply their methods to networks of academic articles, we first identify the distinguishing features of the problem space, here restricted to the network of academic journals covering the topic of creativity support systems. Once this network is clustered, key papers in each cluster are extracted using our algorithm [10-11]. Our algorithm will extract key papers from the each research area cluster by calculate each cited paper's importance by applying the variance value to the PageRank algorithm. Variances values will investigate variances of the publication year of the cited literature. After which role analysis, visualization methods can be applied.

2 Related Studies

We will discuss research related to our study, first covering existing base systems for analyzing academic journals. We then discuss the problems of applying Guimera et al.'s work to academic journal networks.

2.1 Bibliometrics

Bibliometrics is a technique developed by Garfield [12], who had previously proposed the Science Citation Index in the 1950s as a tool to help scientists retrieve early scientific research. The Science Citation Index evolved into bibliometrics, which exposes academic papers through quantitative analyses of "what topics are hot," "which papers are cited most frequently," "what studies are related to one another," and "who qualifies as an important researcher" in a certain research area.

Bibliometrics involves three analysis techniques, as described.

The first analysis techniques is "Direct Citation"; in Fig. 1, Papers A and B are cited in Paper C, and Paper C is cited in Papers D and E. In this case, direct citation

deems that there are links between Papers A/B and Paper C and further links between Paper C and Papers D/E. As a result, there are five nodes and four links in the network. When direct citation is used, a certain paper is deemed to have links with all papers that cite the pertinent paper.

The second technique is "Co-Citation", which was proposed by Small [13]. In Fig. 1, both Paper A and Paper B are cited in Paper C. In this case, co-citation deems that there is a link between Paper A and Paper B; thus, there are two nodes and one link in the network. For pairs of papers in which co-citation was used, i.e., all papers contained in the list of cited literature of a certain paper, there is a link between the paired papers.

The third technique is "Bibliographic Coupling" a technique proposed by Kessler [14]. In Fig. 1, both Paper D and Paper E cite Paper C. In this case, this technique deems that there is a link between Paper D and Paper E; thus, there are two nodes and one link in the network. When bibliographic coupling is used for pairs of papers that cite a certain paper, it is deemed that there is a link between the paired papers.

Note that in our own work, we select the first of these techniques, Direct Citation.

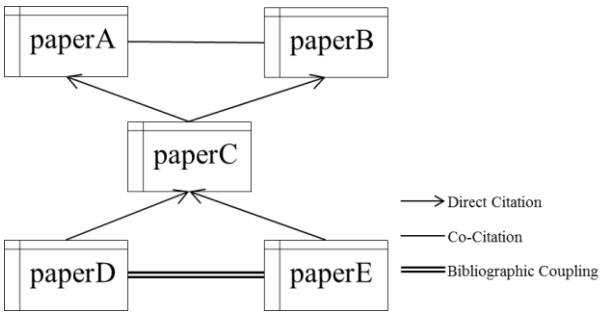


Fig. 1. Three analysis techniques of Bibliometrics

2.2 Cluster Analysis (Clustering)

Cluster analysis (Clustering) is a technique used to divide a large volume of data such as academic papers, into clusters. By clustering data according to common features, can simplify the overall structure of complex data and understand it more directly and thoroughly. We will describe about a typical example of clustering methods. In the initially proposed clustering algorithm by M. E. J. Newman [5-6], the commonly used technique focuses on central links, each of which is “cut,” one by one, starting from the most central. Later, Girvan and Newman [7] focused on the links that mediate the clusters nearest to other clusters using modularity as the evaluation function, and proposed an algorithm that cuts links in descending order of their mediating power. The CNM method (Aaron Clauset, M.E.J.Newman and Cristopher Moore method [8] is a high-speed version of the Newman method meant for application to massive networks.

2.3 Extraction of the key Paper in the Cluster

We previously proposed [10-11] a method for dynamically specifying the key papers (nodes) in each area identified through clustering. The PageRank algorithm [15] is a technique used to determine the most "important" page quantitatively by using calculations in the presence of mutual referencing relations as hyperlink structures. We did investigate variances values from the publication year of the cited literature and calculate each cited paper's importance by applying the variance value to the PageRank algorithm. The common method for obtaining the variance is expressed as follows (1) and the obtained value of variance is stored as Variance. Then calculate each cited paper's importance by applying the variance to the PageRank algorithm.

$$Variance = \frac{\sum(x - \bar{x})^2}{(n-1)} \tag{1}$$

Assuming that the sum of the scores of the citations that "flow out" from a given paper and the sum of the scores of the citations that "flow in" to that paper are equal, we treat this value as the overall pertinence score for the paper. Papers with higher scores are considered more important. By applying the variance value specifically to the "flow in" score for each paper, it is possible to identify the key papers in each area. Although scores have been assigned equally in the conventional algorithm, when there are multiple citations that "flow in," the severity reflecting the state of variance in the citation year is calculated in this study with the consideration that more citations will "flow in" to papers with higher variance values.

2.4 Guimera et al.'s Role Analysis Model

Guimera et al. [9] considered the relationship between the function and structure of cells to form the hub of networks for elucidating metabolic networks. Concretely, they defined the within-module degree, or Z-SCORE (Zi), and the participation coefficient (Pi) as follows:

$$Z_i = \frac{k_i - \overrightarrow{k_{s_i}}}{\sigma_{k_i}} \tag{2}$$

$$P_i = 1 - \sum_{s=1}^{N_M} \left[\frac{k_{is}}{k_i}\right]^2 \tag{3}$$

Zi expresses a particular node's degree of coupling in the cluster to which it belongs. Pi expresses the degree to which an edge of one node is coupled in a cluster other than its own. If all edges of a node exist within its own cluster, Pi = 1. Conversely, if the edges of nodes reside equally in all clusters, Pi = 0. Guimera et al. discerned the seven-role model in Fig.2 using statistical classification. They surmised that the role of a node is defined mainly by its within-community degree and its participation coefficient. They definition of the roles is firstly determined by the within-module degree. They classify nodes with $z \geq 2.5$ as module hubs and nodes $z < 2.5$ as non-hubs. Both hub and non-hub nodes are then more finely characterized by using the values of the participation coefficient. Simple calculations suggest that non-hub nodes can be naturally assigned into four roles:

- *Ultra-peripheral nodes* (role R1).

  If a node has all its links within its module ($P \approx 0$).

- *Peripheral nodes* (role R2).

  If a node has at least 60% its links within the module, then for $k < 4$ it follows that $P < 0.625$

- *Non-hub connectors* (role R3).

  If a node with $k < 4$ has half of its links (or at least two links, whichever is larger) within the module, then it follows that $P < 0.8$. Thus, a plausible region for non-hub connectors is $0.62 < P < 0.8$.

- *Non-hub kinless nodes* (role R4).

  If a node has fewer than 35% of its links within the module, it implies that $P > 0.8$. They surmise that such nodes cannot be clearly assigned to a single module. They thus classify them as kinless nodes. They will demonstrate later that non-hub kinless nodes are found in most network growth models, but not in real-world networks.

Similarly, hubs can be naturally assigned into three different roles:

- *Provincial hubs* (role R5).

  If a node with a large degree, $k \gg 1$, has at least 5/6 of its links within the module, then it follows that $P = 1 - (5/6)^2 - (k/6)(1/k^2) = 0.31 - 1/(6k) \approx 0.30$.

- *Connector hubs* (role R6).

  If a node with a large degree has at least half of its links within the module, then it follows that $P = 1 - 1/4 - (k/2)(1/k^2) = 0.75 - 1/(2k)$. Since $k \gg 1$, $P < 0.75$ for such nodes.

- *Kinless hubs* (role R7).

  If a hub has fewer than half its links within the module, i.e., $P > 0.75$, then they surmise that it may not be clearly associated with a single module. They then classify it as a kinless hub.

In total, they seven roles correspond to seven regions of the zP parameter space (Fig.2). They confirmed that this seven-role model conforms to reality through experiments on real metabolic networks for E. coli.

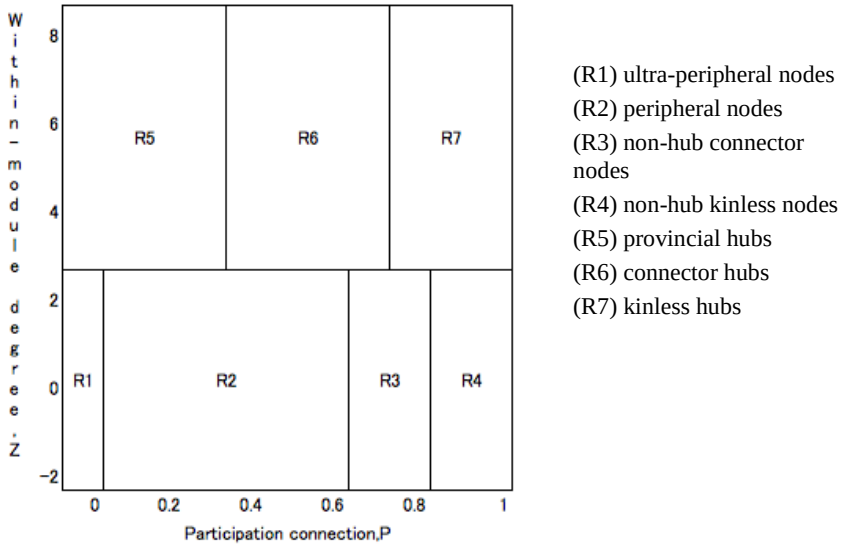


Fig. 2. Guimera et al.'s role analysis model (Guimera, R 2005)

Because Guimera et al.'s methods target metabolic networks, their methods use Z-SCORE as a within-module degree. Our purpose is to examine how distant is the within-module degree from the average in the same cluster. In the case of citation networks for journal articles, it is meant that compared with the difference between the average of the internal cluster. Therefore, although there are many citations, the Z-SCORE will be 0 if the within-module degree is the same as the average for the cluster. In this study, we define a function {Internal links/Total links (Total degree)} as shown in section 3.3 as a substitute for Z-SCORE.

Our Study focuses to the Loss Rate plost(R) that was defined by Shibata [16]. It seen in (4), and then calculated the average of plost(R) in each classification. In this result, the value for the plost(R) of (R5) is larger than the value for plost(R) of (R3). That is to say, Shibata considers the nodes that belong to (R3) to be more important to the structure than the nodes that belong to (R5). He also indicates that the large degree hub is important conventionally, but that the global role of nodes is also important. Our study focuses on the nodes that play a larger role in the global system. We consider role analysis methods as the target of the research area of "Creativity Support Systems."

For a pair of species, A and B, we define the loss rate as the probability,

$$plost(R) = p(RA = 0 \mid RB = R) \tag{4}$$

That a metabolite is not present in one of the species (RA= 0) given that it plays role R in the other species (RB = R).

(Shibata2009)

3 Method for Role Analysis of Key Nodes in the Cluster

3.1 Clustering of Journal Papers

We perform clustering using "Girvan and Newman method" [7] on journal networks. "Girvan and Newman method" is the Cluster analysis method we described at Section 2.2 above. Concretely, we define a Modularity Q, and then calculate the cluster structure such that Q is the maximum. NM is the number of modules, L is number of links, lS is number of links of between nodes in the modules, and dS is a node's coefficient in the modules.

$$Q = \sum_{s=1}^{N_M} [\frac{l_S}{L} - \left(\frac{d_S}{2L}\right)^2] \tag{5}$$

3.2 Extraction of Key Paper in the Cluster

In this section, we extract the key papers from each cluster using the methods presented in section 2.3.

3.3 Calculation of Participation Coefficient and Within-module Degree

We calculate the degree of coupling about another cluster, (Pi) as follows:

$$P_i = 1 - \sum_{s=1}^{Nc} \left(\frac{k_{is}}{k_i} \right)^2 \quad (6)$$

$\boldsymbol{K_i}$ indicates all links of node i
$\boldsymbol{K_{is}}$ indicates the number of links about cluster from node I
$\boldsymbol{N_c}$ is the number of all clusters

Nc indicates all clusters. *Ki* indicates all links of node *i*. *Kis* indicates the number of links about cluster from node *i*. One at the beginning is expected to increase as the variance value increases. If there is no participation coefficient, *Pi* will be 0. Conversely, when the participation coefficient is large, *Pi* will increase. We then calculate the degree of coupling about cluster to which the node belongs, (*Ii*) as follows:

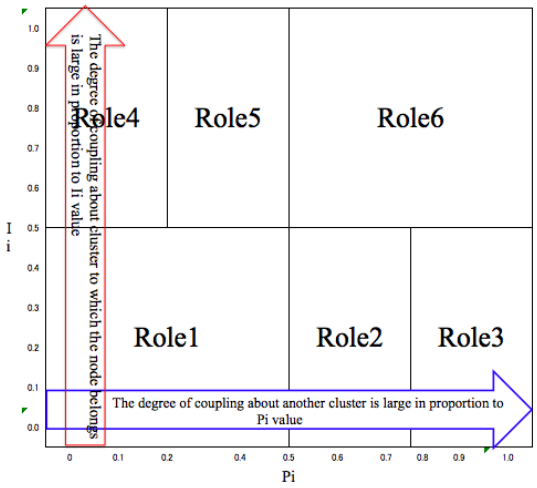
$$I_i = \frac{k_{ic}}{k_i} \quad (7)$$

$\boldsymbol{K_i}$ indicates all links of node i
$\boldsymbol{K_{is}}$ indicates the number of links in the cluster to which node i belongs

Ki indicates all links of node *i*. *Kic* indicates the number of links in the cluster to which node *i* belongs. In other words, the within-module degree is calculated using the ratio of the number of links in the cluster to which the node belongs to the total number of links. The reason for using this function instead of the *Z*-SCORE is that the purpose is to examine how distant the within-module degree is from the average for the same cluster. However, in the case of analysis citation of journal networks, it is meant that compared with the difference between the averages of the internal cluster. Therefore, although there are many citations, the *Z*-SCORE is 0 if the within-module degree of the node is the same as the average of cluster inside degrees.

3.4 Key Node Role Analysis

Method of Role Analysis of Key Nodes Finally, we calculate the Pi and Ii of all nodes per year and plot them on a role analysis chart (Fig. 3). We perform role analysis using this chart. The roles are classified into six categories. We will expound about six categories. At first, we will do classify into four equal parts categories. Then, will add two categories (Role3 and Role4). Role3 will be created by evenly divide the lower right in the four equal parts categories. And Role4 will be created by evenly divide the upper left in the four equal parts categories. This purpose is that we will investigate in detail the nodes of very strong role as the hub for other cluster or own cluster. In other words, Role3 will indicate both the most strong role as the hub for other cluster and the most feeble role as the hub for own cluster. And Role4 will indicate both the most strong role as the hub for own cluster and the most feeble role as the hub for other cluster.



Role6 The Role as a hub in own cluster or for another cluster is strong

Role5 The role of the hub in own cluster＞The role of the hub for another cluster

Role4 The role of the hub in own cluster is very strong

Role3 The role of the hub for another cluster is very strong

Role2 The role of the hub in own cluster＜The role of the hub for another cluster

Role1 The Role as a hub in own cluster or for another cluster is weak

Fig. 3. Role analysis method of key nodes

$\boldsymbol{P_i}$ is the degree of coupling about another cluster
$\boldsymbol{I_i}$ is the degree of coupling about cluster to which the node belongs

Implementation of System The above methods were implemented in Java, and the graphics and statistic calculations, in R and iGraph. We call the implemented system "OTSUKI2012".

Future Prediction Support Using OTSUKI2012 Shibata [16] suggested that the journal papers that trigger radical innovation exhibit the following two steps of development:

1. It becomes a "local hub" in the early stages of clustering.
2. It becomes a "global hub" in proportion to the growth of the cluster.

In this study, "global hub nodes" are nodes that have been growing toward the Role3 nodes in proportion from year to year, because the X-axis shows the degree of growth of the cluster. If we analyse the technology journal networks, the nodes that have been growing toward the Role3 in proportion from year to year show a possibility that the cluster will be fused. This means that the fusion of the clusters refers to the fusion of technology because this study is targeting the technology networks. The fusion of technology refers to the possibilities of innovation, but OTSUKI2012 does not suggest possibilities of innovation but only assert possibilities of innovation. Therefore, further studies of journal papers are needed before OTSUKI2012 can make accurate predictions of innovation. For current purposes, we consider a field of research showing rapid growth in publications to be one in which innovation is occurring.

4 Evaluation Experiment

4.1 Outline of Evaluation Experiment

In this section, we confirm the superiority of OTSUKI2012 to Guimera et al.'s role analysis model in Fig. 2 (GUIMERA2005) through experimentation. Our body of target papers includes papers using the search term "Knowledge Based System," totaling 7,527 papers. We show the results of clustering these papers (the academic landscape) in Fig. 4 and Fig. 5 is a graph comparing the change in the number of citations per year for the top six target papers. Among these six papers, that of Aamodt1994 shows a rapid increase in the number of citations per year from 2002. Fig. 6 is a line graph comparing the degree of cluster growth about these six papers over the last 10 years (to 2011 from 2002). The growth rate of the cluster to which Aamodt1994 belongs was larger than that of the clusters to which the other five papers belong.

Note that a prominent expert in the field stated that the theory presented in Aamodt1994 constituted "Case-based reasoning," after which the theory presented in the Aamodt1994 paper spread worldwide during the 1990s. It then became the foundation of a new field of research that spread rapidly during the 2000s into applied research and the development of systems-based decision support systems, etc.

Given this growth cycle, we can see that the cluster of Aamodt1994's affiliations provides a strong indicator of innovation. Based on this, we next compare OTSUKI2012 and GUIMERA2005 by extracting from them the key papers of the most recent five years (2000 to 2004) in which the innovation occurred.

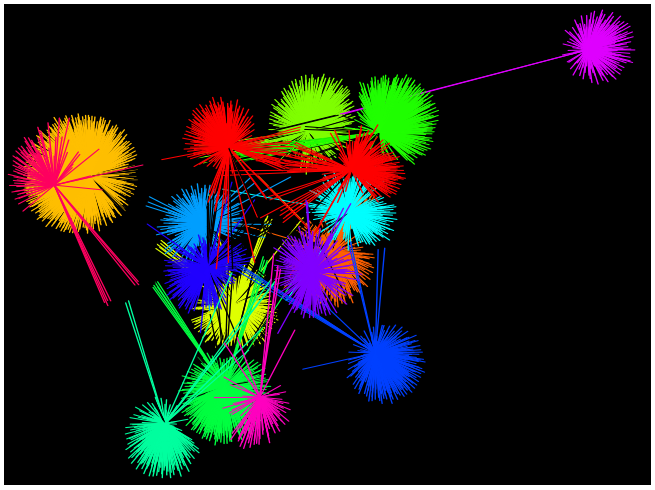


Fig. 4. Result of clustering a target article using OTSUKI2012 (the search term is "Knowledge Based System")

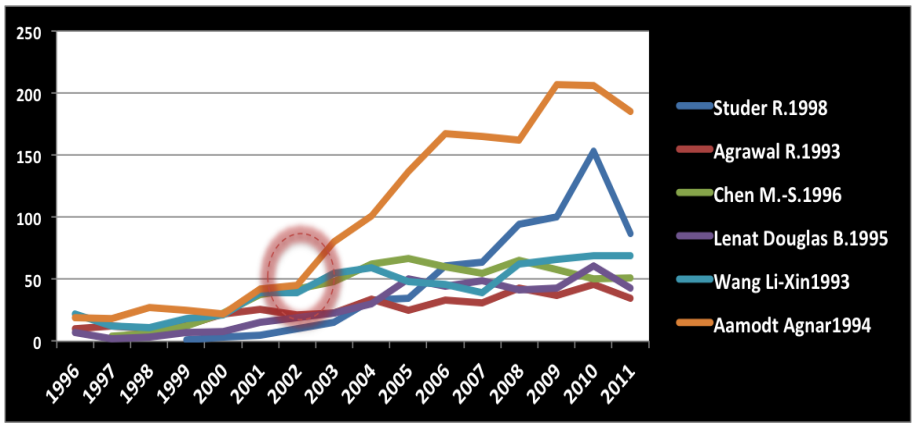


Fig. 5. Transitional phase comparison of the number of citations in the top six papers (the search term is "Knowledge Based System")

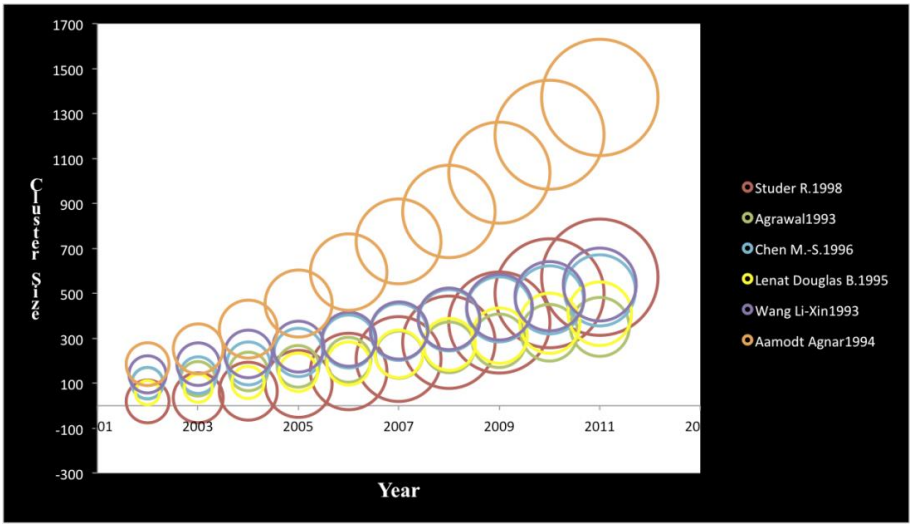


Fig. 6. Comparison of cluster size growth rates

We next calculate the degree of coupling about another cluster (Pi) and the degree of coupling about the cluster to which I belong (Li) for each year. We will plot these values in an x-y plane (Fig. 7 - Fig.9) and note for which of the two implementations, OTSUKI2012 or GUIMERA2005, is Pi growing linearly. The reasoning behind this is as follows: when we analyze technological citation networks, the nodes that have been growing toward the Role3 nodes in proportion from year to year show a possibility that the cluster will be fused. The fusion of clusters indicates a fusion of technology, which in turn indicates possible innovation. Conversely, papers for which there is oscillation between roles make it difficult to assess the potential for innovation.

4.2 Result of the Evaluation Experiment

We analyzed the last five years of innovation resulting from the Aamodt1994 paper using both OTSUKI2012 and GUIMERA2005. The results are shown in Fig. 7. Note that when using OTSUKI2012, the key paper grew linearly toward Role3 from year to year. This clearly indicates that the degree of coupling about another cluster is growing. By contrast, when using GUIMERA2005, the key paper showed oscillation between roles. This oscillation makes analysis more difficult, and we must also consider the possibility that the accuracy of the GUIMERA2005 method is decreasing.

We might suggest that these results depend on the difference in the number of areas of classification (GUIMERA2005 with seven, and OTSUKI2012 with four), but the points of comparison are equal. By comparing only at common points, we see that the growth rate of Pi for Role3 using OTSUKI2012 shows coupling about another cluster, and for R7 using GUIMERA2005 shows likewise. By doing this, we effectively eliminate the influence of the difference in the number of regions by converting rates to a shared scale.

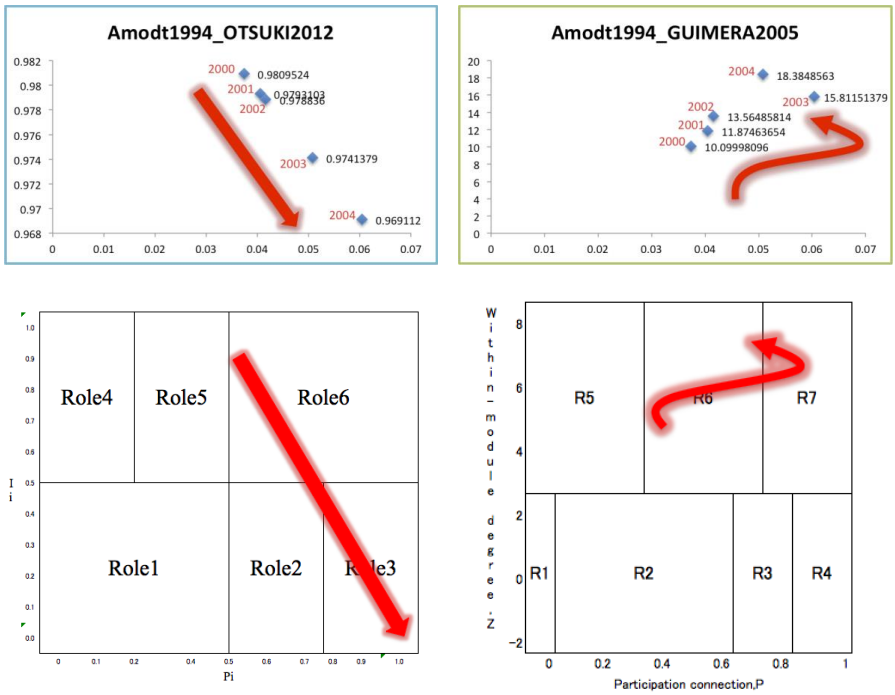


Fig. 7. Comparison of OTSUKI2012 and GUIMERA2005 (for paper Aamodt1994)

Next, we confirm that the result of this evaluation experiment is not exclusive to the query "Knowledge Based System" by performing a parallel experiment using another query, "Data Mining," which produces a set of 10,037 articles. Clustering these results yields the academic landscape shown in Fig. 8.

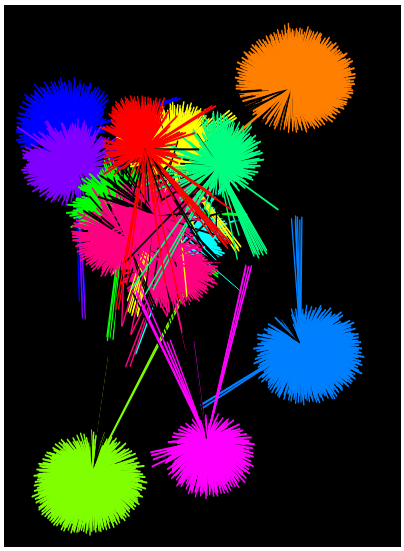


Fig. 8. Result of clustering journal articles queried with "Data Mining" using OTSUKI2012

We selected the key paper (Zimmermann2004) from this set for evaluation, based on the number of citations, vicissitudes of cluster size, and expert comments. As before, we analyzed the last five years of innovation resulting from the Zimmermann2004 paper using both OTSUKI2012 and GUIMERA2005. The results are shown in Fig. 9. Again, note that when using OTSUKI2012, the key paper grew linearly toward Role3 from year to year, indicating that the degree of coupling about another cluster is growing. In contrast, when using GUIMERA2005, the paper's growth oscillated between R6 and R7. These parallel results indicate that OTSUKI2012 more clearly showed the growth of innovation for the key paper than did GUIMERA2005 did.

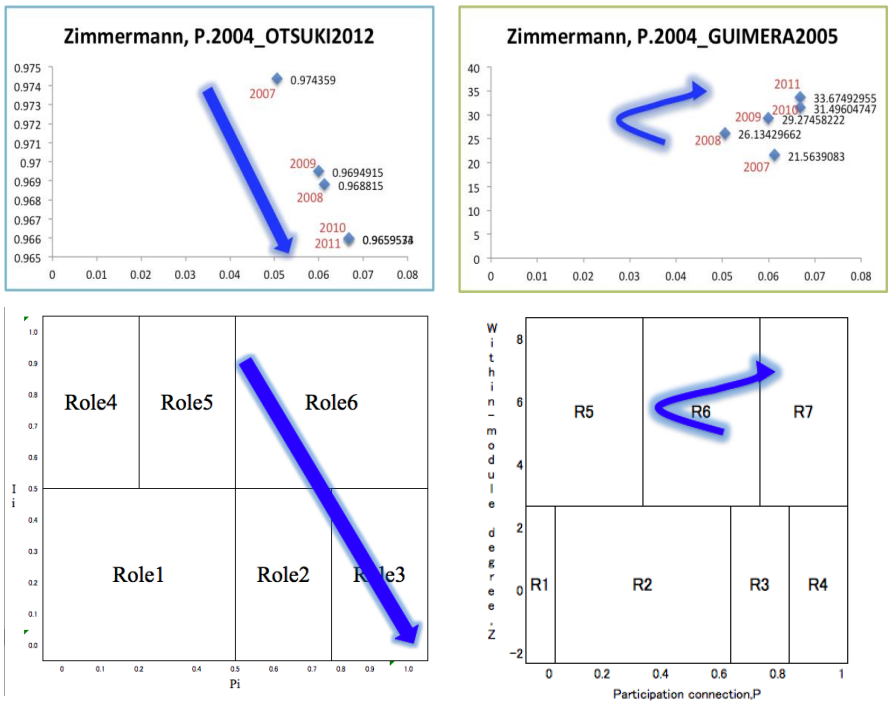


Fig. 9. Comparison of OTSUKI2012 and GUIMERA2005 (for paper Zimmermann2004)

4.3 Discussion

From these experiments, we note that in the case of Aamodt1994, the value of *Pi* (degree of coupling about another cluster) decreased after 2003 in spite of previous increases. In the case of Zimmermann2004 as well, the value of *Pi* decreased after 2010 in spite of previously increases.

The GUIMERA2005 method calculated the degree coefficient in the cluster using *Z*-SCORE, so the degree of coupling will be the same as the average in the cluster, i.e., 0. Furthermore, if it is less than the average degree is negative.

We think the reason for the oscillation between R6 and R7 is an extreme decrease of the *Z*-SCORE for the key paper, which is caused by the extreme decrease of the inner or outer join degree. OTSUKI2012 can solve this problem by using the above function (6). This means that OTSUKI2012 will be able to predict support of innovation than GUIMERA2005 as a target of the journal paper network.

5 Conclusions and Future Work

In this study, we identified the problems of applying Guimera et al.'s methods to a target network of articles in academic journals. We then proposed a new role analysis method and visualization system as a target of the academic Article Networks. Final-

ly, we performed a comparative evaluation of our method and Guimera et al.'s method. This evaluation used the key papers that appeared to have triggered large amounts of innovation. Results showed that our method was able to show the movement of key paper innovation more clearly than Guimera et al.'s method. We hope to extend our study to include other analysis methods in future works.

Acknowledgment

This study was supported by Editage Inspired Researcher Grant. We thank "Cactus Communications Pvt. Ltd." and "Leave a Nest Co., Ltd." for invaluable assistance.

References

1. McKinsey Global Institute. : Big data: The next frontier for innovation, competition, and productivity, The Report of McKinsey Global Institute, (2011). http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
2. Newman, M.E.J.: Networks: An Introduction. Oxford, UK: Oxford University Press, (2010).
3. Stanley. M.: The Small World Problem, Psychology Today, May. pp 60 - 67 (1967).
4. Burt, R. S.: Structural Holes: The Social Structure of Competition. Harvard University Press, paperback edition, (1995).
5. Newman, M. E. J.: Fast algorithm for detecting community structure in networks, Phys, (2004).
6. Newman, M. E. J.: A measure of betweenness centrality based on random walks, Social Networks, Vol. 27, No.1, pp. 39-54. Rev. E, Vol. 69,(2005).
7. .Newman, M.E.J and Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113, (2004).
8. Clauset, A., Newman, M.E.J, Moore, C.: Finding community structure in very large networks. Phys. Rev. E 70, 066111, (2004).
9. Guimera, R, Amaral, LAN: Cartography of complex networks: modules and universal roles, J. Stat. Mech.-Theory Exp., art. No. P02001, (2005).
10. Otsuki, A., Kawakami, A.: Academic Landscape using Network Analysis Considering the analysis of variance of the number of years as a weighted publication, The 73rd National Convention of Information Processing Society of Japan (IPSJ), pp. 655-657, (2011).
11. Otsuki, A. , Kawakami, A.: Academic Landscape based on network analysis considering analysis of variation in the years of lucubration publishing, New Research on Knowledge Management Models and Methods, InTec Open Science, Chapter 17, pp.371-378, (2012).
12. Garfield, E.: Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas". Science (AAAS) 122 (3159): pp.108–111, (1955).
13. Small, H.: Co-citation in the scientific literature: a new measure on the relationship between two documents, Journal of the American Society for Information Science, Vol. 24, pp.28-31, (1973).
14. Kessler, M.: Bibliographic coupling between scientific papers, American Documentation Volume 14, Issue 1, pp.10–25, (1963).

15. Page, Lawrence, Brin, Sergey, Motwani, Rajeev, Winograd, Terry: The PageRank Citation Ranking: Bringing Order to the Web, (1998).
16. Shibata, Naoki: Study on the methodology of early detection of radical innovation, a doctoral dissertation, School of Engineering the University of Tokyo (2009).