Co-author Information and Authors’ Affiliation Information in Scientific Literature Using Centralities
-The Researchers who Act as Mediators between Organizations-

Yuji MIZUKAMI*†, Keisuke HONDA‡, Shigenori SUZUKI§, Junji NAKANO‡, Akira OTABE*†

*Nihon University, 1-2-1 Izumicho, Narashino, Chiba 275-8575, Japan
‡Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
§Research Institute of Info-Communication Medicine, 5-17-11 Honcho, Koganei, Tokyo 184-0004, Japan

Abstract

The Institute of Statistical Mathematics aims to further develop statistics by promoting the exchange of data and resources between domestic and foreign researchers. This has been examined from a statistical viewpoint, which suggests using a new index whereby the progress and effectiveness of the joint use can be objectively assessed. We analyzed co-author information using betweenness centrality in articles written by researchers at the Institute of Statistical Mathematics in 2013, obtained from the Web of Science. The analysis results enable identification of the researched people that are in the center of their research field.

Key words: Network analysis, Joint work analysis, Betweenness centrality, Organization

1. Introduction

Exchange of data and resources at different research fields is one of the effective methods for the promoting innovations at research fields. Based on this fact, the Institute of Statistical Mathematics promotes the exchange of data and resources between domestic and foreign researchers. We have been examined a lot about this topic and we may suggest a new index for measuring its progress and effectiveness.

Our analytical approach consisted of two stages: in stage 1, we applied betweenness centrality to identify people that are in the center of each research field (network sector), and in stage 2, we proposed four new formulas that are based on betweenness centrality to separately analyze researcher networks within and outside their organizations.

We utilized co-author information from Thomson Reuters’ Web of Science and collected information by researchers belonging to the Institute of Statistical Mathematics between 2009 and 2013. The data from 2009 to 2013 are utilized for preliminary analysis and that from 2013 onward is utilized for the main analysis in this paper. In total, the data consisted of 153 papers from 502 researchers.

The results from stage 1 enable identification of the people that are in the center of each research field, whereas those from stage 2 enable identification of the features of cosmopolite researchers. We expect that the high number of cosmopolite researchers could exert a positive influence to promote the joint use of resources and assets.

2. Literature review and background

2.1 Classification of research evaluation

The first category for evaluating research analyzes the number of published works in the literature. The second category analyzes the number of citations in an article and between articles [1]. This classification system is shown in Table 1. The purpose of analyzing the number of articles first is to find the productivity index of the article and the scale index of research activities. For this, we mainly use a simple tabulation technique. In contrast, by analyzing citation statistics second, we can identify the consumption index of the article, and the quality index of research activities. The consumption index is indicative of the degree of utilization. For this, the citation analysis technique or joint work analysis is used [1]. While these are minor cases, we still include the analysis of acknowledgments, co-word, and attendant classification for analyzing the citation statistics [2]. Acknowledgment analysis targets acknowledgements, whereas co-word analysis targets plural words. Finally, attendant classification analysis targets the field classification co-appearance phenomenon for analysis [2]. When analyzing the number of articles, we distinguish betweenness and summarize the following article analytics: field, year, country of origin, and affiliation.

*Corresponding author: mizukami.yuuji@nihon-u.ac.jp

Received: October 22, 2015
Accepted: July 12, 2016
Furthermore, the correlation analyses with other economic statistics indices have been carried out [1]. For example, Vergidis et al. [3] conducted an international comparison of article productivity in microbiology from 1995 through 2003 and discovered that the productivity is highest in Western Europe, followed by North America. Furthermore, productivity in Asia, Central and South America, and Eastern Europe saw the highest growth rates during this period. Conversely, citation analyses summarize the number of citations per article based on academic journal distinction and country. Furthermore, IF has been used to gauge academic journal distinction. According to academic journals, IF uses the Science Citation Index database of Thomson Reuters. Moreover, the IF is calculated based on the total number of articles published within the last two years and the number of times that those articles were cited in the selected year [1]. After Vergidis et al. [3] performed the international comparison of the average IF in microbiology from 1995 through 2003, the results showed that journals in North America had the highest IF of 3.4, followed by Western Europe with 2.8, and the average IF in other locations was 2.4.

The IF is typically used as an index for the purchase of academic journals in libraries, but it is considered unsuitable for examining personal achievements [1]. Thus, here we recommend analyzing co-authors for assessing personal achievements.

Co-author analysis is applicable to assessing the performance of the individual. For example, it is possible to grasp the progress of collaboration between researchers at an institution where the collaboration exists [2].

### 2.2 Co-authored analysis and its index

Co-author analysis consists of two categories, the first one analyzes the strength of the ties between the authors, and the second one analyzes the author centrality in the network.

In analyzing the strength of the ties between the authors in a network, indices such as the following can be cited as typical examples: first is frequency of co-authorship (FOC) as a basic index. It is the index of frequency that represents the co-authored papers between authors. Moreover, it is an index for absolute ties strength. Second, the cooperation index is an index that corrects the FOC by averaging the number of articles of each author in order to eliminate the influence of the author on the productivity. It is also an index of the relative ties strength [4]. Third, homophily index (HI) corrects the FOC by averaging the number of articles of each group to which the author belongs in order to eliminate the influence of the group on the productivity. Moreover, it is an index of the relative ties strength [5]. Fourth, the co-authorship index corrects the FOC by the number of articles in each group in a manner similar to the HI. However, the number of articles of either group is used in this index. It is also an index of the proportion of co-authorship perspective of the group [6]. Lastly, the affinity index is the ratio of the number of co-authorship articles of a certain group and the other groups. Therefore, it is an index of the connection strength of a certain group that indicates how much stronger it is than other groups [4]. In addition, there are other indices related to the strength of author ties, including the relative frequency of co-authorship [7].

The second category analyzes author centrality within the network, which quantitatively represents whether the author plays an important role within the network. Indices including the following can be cited as typical examples. The degree of centrality is the most simple centrality index that indicates that when the author (node) connection is concentrated, the centrality is high. Moreover, the number of connections of each author refers to their degree of centrality. Closeness centrality is the index related to the distance between authors. Thus,

\[
b_i = \frac{r_i}{c_i} \quad (1)
\]

\[
r_i = \sum_{i_0=1}^{N} \sum_{i_0=1}^{i-i_0-1} \frac{g_i(i_0i_0)}{N_i_{i_0}} \quad (2)
\]

\[
c_i = \frac{(N - 1)(N - 2)}{2}, \text{if Undirected Graph}
\]

\[
c_i = \frac{(N - 1)(N - 2)}{2}, \text{if Directed Graph}
\]

Attention: Brackets are used for minor cases.

<table>
<thead>
<tr>
<th>Category</th>
<th>Purpose</th>
<th>Analytical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of the number of articles</td>
<td>Productivity index of the article</td>
<td>Simple tabulation</td>
</tr>
<tr>
<td>Quotation statistics and analysis</td>
<td>Consumption index of the article</td>
<td>Quotation analysis</td>
</tr>
<tr>
<td></td>
<td>Quality index of research activities</td>
<td>Joint work analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Acknowledgment analysis)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Co-word analysis)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Attendant classification analysis)</td>
</tr>
</tbody>
</table>

### Table 1 Classification of research evaluation

*Quotation* refers to their degree of centrality. Moreover, the number of connections of each author refers to their degree of centrality. **Closeness centrality** is the index related to the distance between authors. Thus,
there is a higher centrality of the authors as the distance between the other authors is closer [8] [9]. Furthermore, it is the author’s higher closeness that can be taken contact in a number of authors by smaller step. Finally, Betweenness centrality (BC), which was first mentioned by Shimbel [10] in 1953 and developed later by Freeman [8] [9], is an index that takes the higher centrality author as the greater path through the author. Furthermore, the higher centrality author facilitates the flow of information and resources in the network [8] [9]. The definition of BC is given in formulas 1, 2, and 3.

3. Conceptual framework

3.1 Preliminary analysis - confirmation of prerequisites

The co-author data from Web of Science is utilized in this paper. To confirm the prerequisites, data validity is confirmed. Abbasi et al. [11] analyzed the co-author steel structure data field from 1999 to 2009 using three centralities. Consequently, as the co-authors network expands, an author can increase the flow of communication and information, thereby increasing the authors’ betweenness centrality, who can invite new authors. Thus, high betweenness centrality is better than both a high degree of centrality (authors who have many ties) and high closeness centrality (authors who have many direct ties).

In this paper, according to the preliminary analysis, we confirm that co-author networks have been expanding over the last several years.

3.2 First main analysis - attributes of high betweenness centrality authors

Abbasi et al. [11] analyzed the co-author data of the steel structure field from 1999 to 2009. Based on the results, they concluded that many high betweenness centrality authors supervised students and other researchers, and they wrote articles with those students and other researchers. Altogether, the high betweenness centrality authors belong to the elite of the organization.

Conversely, Rogers [12] proposed a model in his article called cosmopolite (which means diffusion of innovations). The cosmopolite people belong to the organization but also have contacts outside the organization. Those people serve as ties to the organization and other organizations. Thus, cosmopolite people are divided into two types: those who belong to the organization’s elite and those who belong to the organization’s lower ranks. The upper cosmopolite consciously focuses the attention outside of the organization to make a management decision. The lower cosmopolite is in the forefront of the field and is always exposed to the external environment. Therefore, they often have outside contact through their daily work.

We identify the high betweenness centrality authors and discover authors’ attributes in order to determine which author belongs to the elite of an organization, as alleged by Abbasi et al. [11], or who belongs to both the elite and lower ranks in the organization, as alleged by Rogers [12].

3.3 Final main analysis - inter-organizational betweenness centrality

According to Rogers [12], people are responsible for different roles in the organization by their human qualities and intentionality. Thus, the phenomenon can be broadly classified into local and cosmopolitan. The local people are very loyal to the organization, and as a priority, adapt to the organization’s norms. Conversely, cosmopolitan people have a high interest outside of the organization. Thus, they have the combined characteristics of both Local and Cosmopolitan. Therefore, these people can serve to tie the organization and other organizations.

The purpose of this research here is to determine the local-to-cosmopolitan ratio of the cosmopolite people (the high betweenness centrality people). To achieve this, we need to obtain the number of betweenness indices of both the inside organization (for local) and the inter-organization (for cosmopolitan). However, the number of betweenness of inside and inter-organization cannot be obtained separately by existing betweenness centrality methods. For example, in Fig. 1, the number of betweenness of node 3 is 12, and it is not possible to distinguish between inside organization and inter-organization.

In this section, we propose new indices for determining the number of both inside organizational betweenness centrality (OBC) and inter-OBC based on the existing betweenness centrality. For example, in Fig. 1, the number of inter-organizational betweenness of node 3 is 12, and the number of inside organizational betweenness of node 3 is 1, and this path is \{1, 3, 2\}. Moreover, the number of inter-OBC will be increased if the organizations of the source and destination nodes are different. On the other hand, the number of inside OBC will be increased if the organizations of the source and destination nodes are the same.

The definition of new betweenness indices are given in
Yuji MIZUKAMI, Keisuke HONDA, Shigenori SUZUKI, Junji NAKANO, Akira OTABE

\[ bd_i = \frac{rd_i}{c_i} \quad \text{(4)} \]
\[ roo_i = \sum_{i=1; i \neq i}^{N-1} \sum_{i=1; i \neq i}^{N} \frac{g_{i}(i(i;O_{i} \neq 0_{i})))}{N_{i}i_{i}} \quad \text{(8)} \]
\[ bio_i = \frac{rio_i}{c_i} \quad \text{(5)} \]
\[ rio_i = \sum_{i=1; i \neq i}^{N-1} \sum_{i=1; i \neq i}^{N} \frac{g_{i}(i(i;O_{i} \neq 0_{i})))}{N_{i}i_{i}} \quad \text{(9)} \]
\[ bii_i = \frac{rri_i}{c_i} \quad \text{(6)} \]
\[ rii_i = \sum_{i=1; i \neq i}^{N-1} \sum_{i=1; i \neq i}^{N} \frac{g_{i}(i(i;O_{i} \neq 0_{i})))}{N_{i}i_{i}} \quad \text{(10)} \]

formulas 4–11. For example, in Fig. 1, roo3 is 3 \{\{4, 3, 7\}, \{5, 4, 3, 7\}\}. The second one is bio3, and this value is 8 \{\{1, 3, 4\}, \{1, 3, 4, 5\}, \{1, 3, 7\}, \{2, 3, 4\}, \{2, 3, 4, 5\}\}. The third one is bii3, and this value is 1 \{1, 3, 2\}. The fourth one is rd3, and this value is 11 \{path of roo3 and rii3\}. The final one is rii3, and this value is 12 \{path of roo3, rii3, and rii3\}.

4. Research methodology

4.1 Data source

We obtained the article information from the Institute of Statistical Mathematics 2013 using the Web of Science. The data consisted of a total of 153 papers from 502 researchers for the main analysis. The data retrieval conditions are as follows: search mode is “Basic search,” search word is “Inst Stat Math,” category is “Address,” time span for preliminary analysis is “2009–2013,” and time span for main analysis is “2013–2013.” In addition, these search words are utilized by the universities and public institutions’ English name variation table (Web of Science edition) of the National Institute of Science and Technology Policy.

4.2 Data analysis

The data analysis consisted of five steps. The first step is data collection (the details of data retrieval conditions are given in section 4.1). The second step is data mining. First, the adjacent matrix table is constructed from spreadsheet software. Second, the network data is made using the iGraph function in the statistical analysis tool (R, version 3.1.2). Finally, the network diagram is made using the network visualization tool (Cytoscape, version 3.2.0). The third step is a preliminary analysis, whereby we confirm that the co-authors network has been expanding over the last several years as an assumption (details are given in section 3.1). The analyzed data incorporate article information from 2009 to 2013 (details are given in section 4.1). The fourth step is the first main analysis, whereby we determine the attribute of high betweenness centrality authors (details are given in section 3.2). The analyzed data contains article information from 2013 (details are given in section 4.1). The fifth step is the last main analysis, whereby we determine the ratio of inside OBC and inter-OBC (details are given in section 3.3). The analyzed data contain article information from 2013 (details is in section 4.1). In addition, inter-OBC is proposed in this paper.

5. Results

5.1 Preliminary analysis - confirmation of prerequisites

In this section, according to the preliminary analysis, we confirmed that the co-authors network has been expanding over the last several years.

| Table 2 Comparison of the network size (2009–2013) |
|-----------------|-----|-----|-----|-----|-----|
|                | 2009 | 2010 | 2011 | 2012 | 2013 |
| Total number of authors | 276  | 365  | 345  | 427  | 502  |
| Total number of betweenness (\(\sum\beta_i\)) | 485  | 1278 | 318  | 2272 | 14297 |
| Total number of authors, if betweenness (\(\beta_i\)) > 0 | 27   | 38   | 22   | 34   | 51   |

International Journal of Japan Association for Management Systems
The comparison of network size is shown in Table 2. There are three indices in the table. The first index “Total number of authors” is used to simply understand the network size. The second index “Total number of betweenness ($\sum r_i$)” is important to understand the expanding size of the network, based on Abbasi et al. [11] (the details are given in section 3.1). The third index “Total number of authors, if betweenness ($r_i$) > 0” is a subset of the second index, for reference. The trend of those indices had been increasing, except in 2011, when the data were influenced by reorganization in 2010 at the Institute of Statistical Mathematics. We determined that the co-authors network has been expanding in the last five years (2009–2013).

5.2 First main analysis - attributes of high betweenness centrality authors

In this section, we identify the high betweenness centrality authors and their attributes. According to Abbasi et al. [11], authors are only identified as being a part of an organization’s elite, but Rogers [12] states that authors can belong to either the lower ranks or the elite of an organization.

The Pareto chart of number of betweenness ($r_i$) is shown in Fig. 2. The author’s name in this paper is a pseudonym. The number of betweenness ($r_i$) of authors A is 6014 and accounts for 42.07% of the total number of betweenness ($\sum r_i$). Moreover, the percentage of betweenness ($r_i$) of the top five authors is 86.01% of the total percentage of betweenness ($\sum r_i$). Based on these results, we only analyzed the top five authors in this paper.

The network diagram of betweenness ($r_i$) of the top five authors (2013) is shown in Fig. 3. This network diagram is made by Cytoscape version 3.2.0. This cluster consists of 122 authors, but there are only five top authors. The comparison of betweenness ($r_i$) of the top five authors is shown in Table 3. There are five indices in the table. The first index, “Number of betweenness ($r_i$),” is utilized for determining rank. Other indices were changed with the number of betweenness ($r_i$). (The details of correlation of each centrality is mentioned in Uddin et al. [13].)

The properties of the top five authors are shown in Table 4. In this table, the name of the organization is a pseudonym except the Institute of Statistical Mathematics. Authors A, C, and E are professors or emeritus professors and belong to the elite of the

![Table 3 Comparison of betweenness ($r_i$) - top five authors (2013)](image)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>Number of betweenness ($r_i$)</td>
<td>6014.00</td>
<td>3310.50</td>
<td>1774.50</td>
<td>469.00</td>
<td>412.00</td>
</tr>
<tr>
<td>Closeness centrality</td>
<td>5E-6</td>
<td>5E-6</td>
<td>5E-6</td>
<td>5E-6</td>
<td>5E-6</td>
</tr>
<tr>
<td>Degree centrality</td>
<td>59</td>
<td>39</td>
<td>18</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Total number of joint articles</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Maximum number of co-authors in an article</td>
<td>16</td>
<td>17</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

![Fig. 2 Pareto chart of number of Betweenness ($r_i$) (2013)](image)

![Fig. 3 Network diagram of Betweenness ($r_i$) - top five authors (2013) (made by Cytoscape version 3.2.0)](image)
organization. Conversely, the second type of author belongs to the lower ranks in the organization. Authors B and D are assistant professors and work for multiple departments. They belong to the lower ranks of the organization.

In summary, we identified the top five high betweenness centrality authors, and we concluded that there are two types of authors belonging to either the lower ranks or the elite of the organization, as mentioned by Rogers [12].

5.3 Final main analysis - inter-organizational betweenness centrality

The cosmopolite people, as mentioned by Rogers [12], have the combined characteristics of both local and cosmopolitan. Therefore, those people can serve to tie two different organizations, as explained in detail in section 3.3.

In this section, we clarify the local and cosmopolitan ratio (localized level) of the cosmopolite people (the high betweenness centrality people). For this, we need to calculate the number of betweenness indices of both the inside organization (for local) and inter-organization (for cosmopolitan).

The calculated results of the top five authors are shown in Table 5, with details of each index given in section 3.3. There are two ratio indices in this table. The first ratio is the \( \frac{b_{ii} + b_{ii}}{b_{1}} \) index, and it is utilized to understand the efficiency of the OBC. The second ratio is the \( b_{1} \) index, and it is utilized to understand the localized level (details of the local and cosmopolitan frameworks are given in section 3.3).

In terms of efficiency (defined as \( \frac{b_{ii} + b_{ii}}{b_{1}} \)), the number of observations is just five, but the average efficiency is 17.527 \%, the standard deviation is 4.236, the minimum efficiency is 10.024 \% for author E, and the maximum efficiency is 20.387 \% for author C. Based on this result, we may conclude that less than \( 30.235 \% \) of the betweenness centrality (\( b_{1} \)) are the internal organizational links or the gateway organizational links (group A). More than 69.765 \% of the betweenness centrality (\( b_{1} \)) are external organizational links (without via any inside authors except the author (i)).

In terms of the localized level (\( b_{ii} \)) results, the number of observations is just five, but the average localization level is 2.954 \%, standard deviation is 2.765, minimum localized level is 0.000 \% for authors D and E, and the maximum localized level is 5.917 \% for author C. Based on this result, we may conclude that less than 11.204 \% (the average value + three standard deviations) of group A are local links, and more than 88.796 \% of group A are cosmopolite links.

To summarize, 69.765 \% of the betweenness centrality (\( b_{1} \)) are external OBC (\( b_{ii} \)) (without any inside authors, except the author (i)). 26.847 \% (30.235 \% * 88.796 \%) of the betweenness centrality (\( b_{1} \)) are external OBC (\( b_{ii} \)). Cosmopolitan links (the links between an author belonging to the outside organization and another author belonging to the organization). 3.388 \% (30.235 \% * 11.204 \%) of the betweenness centrality (\( b_{1} \)) are internal OBC (\( b_{ii} \)). Local links (the links between two authors both belonging to the organization).

6. Discussion and conclusion

The Institute of Statistical Mathematics aims to promote the joint use of resources and assets by outside researchers. We expect that the high number of cosmopolite researchers could positively contribute to promoting joint use of resources and assets. As alleged by Roger [12], the cosmopolite people have both the local and cosmopolitan characteristics. Therefore, these people should serve as a tie between the organization and outside the organization.

Our analytical activity in this paper consists of two stages. In stage 1, we used betweenness centrality to identify people in the center of each research field (network sector); the analysis results enable the identification of such people. In stage 2, we proposed four new analysis formulas based on betweenness centrality to individually analyze the researcher networks within and outside the organizations. The analysis results enable the identification of the features of cosmopolite researchers. In addition, we utilized the article information from the Institute of Statistical Mathematics.
2013 using Thomson Reuter’s Web of Science. The main data used for the analysis consisted of 153 papers from 502 researchers.

Our results in stage 1 indicate that we can identify top five high betweenness centrality authors, and we realized that there are two types of authors, which belong to either the elite or lower ranks of an organization, as mentioned by Rogers [12]. Based on this result, we proposed that the organization may increase hiring of researchers (i.e., visiting professors) for contributing to the elite of the organization, as well as to increase hiring of young researchers (i.e., assistant professor) for contributing to the lower ranks of the organization.

An interesting point regarding young researchers was highlighted during the analysis. The reason they have many contacts with outside organizations is that they are at the forefront of the field, and are always exposed to the external environment. Therefore, they often have contact outside their daily work, according to Rogers [12]. However, when they are later promoted, this assumption no longer has any merit, which means their contact outside of the organization decreases. There are no previous studies of this phenomenon. Future studies in this field are needed.

Our results from stage two indicate three key findings. 69.765 [%] of the betweenness centrality (b_i) is not related to any of the inside authors except the author (i). 26.847 [%] (30.235 [%] * 88.796 [%]) of the betweenness centrality (b_i) are linked to Cosmopolitan, and the links are between authors belonging to the outside and within the organization. 3.388 [%] (30.235 [%] * 11.204 [%]) of the betweenness centrality (b_i) comes from Local links, which are between two authors, both belonging to the organization. There are more outside links compared to inside links.

An interesting question was raised following this result. Had author (i) invited outside authors to his/her organization? Or, had author (i) been invited by outside authors to other organizations? Our purpose is to invite outside authors to the organization of author (i). However, we do not have any previous studies regarding this phenomenon, and we realize that future studies in this field are needed.

A limitation of this paper is that we only analyzed a single case from one organization. Therefore, our contribution mainly provides a hypothesis. In order to generalize our findings, more case studies are needed in other years and/or other scientific domains.

**Acknowledgements**

The authors would particularly like to thank Yosuke Mizutan for his invaluable assistance in obtaining the data used for this study. This study was funded by the Institute of Statistical Mathematics joint research program (27-Kyoken-1027).

**References**


