



Analyzing Promising Researchers Using Network Centralities of Co-authorship Networks from Academic Literature

Masanori Fujita¹ · Hiroto Inoue² · Takao Terano³

Received: 3 April 2020 / Accepted: 16 July 2020 / Published online: 19 September 2020
© The Author(s) 2020

Abstract

We propose a new way of using the betweenness centrality measure with co-author networks from an academic literature database to evaluate young researchers. It is difficult to discover and evaluate promising young researchers with indexes based on the number of cited papers, such as the *h*-index to which published papers introduce a lag and whose impact only becomes apparent after they have been cited by other papers. We validated the effectiveness of the measure as an index for evaluating young researchers. Our investigation of 1.92 million publications in the biological sciences shows that Research Fellows with the Japan Society for the Promotion of Science (JSPS) have higher rankings and progress more quickly than other researchers. In addition, differences between JSPS Research Fellows and other researchers were observed at earlier stages using the proposed method than with the *h*-index and with centralities from literature published in the past 4 years. We expect that the proposed use of the betweenness centrality measure can be applied effectively to extract promising young researchers.

Keywords Academic literature database · Co-author network · Betweenness centrality · Young promising researcher

✉ Masanori Fujita
fujita.m.ai@m.titech.ac.jp

Hiroto Inoue
h-inoue@mail.dendai.ac.jp

Takao Terano
terano@cuc.ac.jp

¹ Tokyo Institute of Technology, Tokyo, Japan

² Tokyo Denki University, Tokyo, Japan

³ Chiba University of Commerce, Ichikawa, Japan

Introduction

Talented researchers are indispensable to innovation and development in science and technology. However, it is difficult to discover and evaluate promising young researchers such as doctoral and postdoctoral students. Indexes based on the number of cited papers, such as the h -index, can be used to evaluate researchers and the impact of publications, yet already-published papers introduce a lag to such indexes, and their impact only becomes apparent after they have been cited by other papers. As such, these indexes are unsuitable for evaluating young researchers and students who have only conducted research for short periods and have yet to accumulate achievements. For this reason, peer review by experts is generally adopted when evaluating young researchers. However, peer reviews lack breadth and quantitiveness.

Therefore, to find and evaluate promising young researchers, especially those involved in group research, we focused on the betweenness centrality of co-authorship networks composed of academic literature databases and the characteristics of its time series variation. We propose a novel use of this index to supplement the lagging nature of citation indexes [1, 2]. The betweenness centrality of co-authorship networks is not an index that directly measures the impact of research results, like an index of citations. Rather, the betweenness centrality quantitatively shows the degree of contribution to the betweenness of a research community. It indicates whether a researcher is in a key position to connect research groups or researchers in a particular field. In addition, the betweenness centrality of co-authorship networks is an index that can be measured simultaneously with the publication of a paper. Therefore, an index that uses the betweenness centrality of co-authorship networks is more immediate than an index of citations, and it is quantitative, unlike peer review. Further, it allows for a broad comparison of researchers recorded in a database.

The purpose of this study is to determine whether the betweenness centrality of co-authorship networks comprising academic literature databases is useful for evaluating promising young researchers. To this end, we clarify the betweenness centrality of Research Fellows with the Japan Society for the Promotion of Science (JSPS), who were evaluated as promising according to the existing JSPS peer review process, and according to the characteristics of time series variation with the h -index. We show that researchers are able to be evaluated earlier by using the betweenness centrality than by using the h -index.

The paper is organized as follows. First, we survey related research. Following this, we describe how we extracted 15 years of bibliometric information in the biological sciences from an academic literature database (JSTPlus) provided by Japan Science and Technology Agency (JST). The information amounted to approximately 1.92 million items of data, which we used to analyze temporal changes to the betweenness centrality of co-author networks and the h -index. Finally, we introduce a statistical model that we developed to characterize JSPS Research Fellows and comprehensively search the database for researchers with similar characteristics.

Related Research

Representative evaluation indexes for researchers include the number of articles published in academic journals with a high impact factor based on the citation network of the written documents, and citation indexes such as the h -index. The impact factor is not an index that directly evaluates the researcher or the paper itself. Rather, it represents an evaluation of the journal based on the number of citations of articles published in each journal [3]. The h -index is defined as the maximum value of h such that the given author has published h papers that have each been cited at least h times. It is an index that indicates the impact of an individual researcher's activity [4]. These citation indexes are obtained based on the number of citations of published documents, and they are quantitative, revealing the impact of research results. They enable evaluations of talented researchers who have already accumulated research experience and produced results. However, these indexes merely follow past research results. As such, it is difficult to evaluate young researchers who have a promising future, yet still have insufficient research achievements [5, 6]. A typical example of a social network other than a citation network composed of citation relationships in a document database is a co-authorship network composed of document co-authorship relationships. Here, we focus on the centrality of co-authorship networks because the measure has immediacy, whereas citation indexes such as h -index, based on a citation network, are necessarily delayed.

In a related study of co-authorship networks, Melin clarified the characteristics of collaboration within research institutions, within countries, and internationally, based on the co-authorship of papers, arguing that the analysis of co-authorship relationships is useful for science and technology policy [7]. Barabasi et al. analyzed a co-authorship network of papers in two fields (viz., mathematics and cognitive science), and Newman studied three fields (viz., mathematics, biology, and physics). They clarified the differences in collaborative patterns among academic fields [8, 9].

The relationship between the centrality of co-authorship networks and the achievements of researchers has also been discussed. For example, Shinoda constructed a co-authorship network focusing on the order of authors of articles published in Transactions of the Japanese Society for Artificial Intelligence, and showed the identification and transition of central researchers from the society based on a centrality index [10]. Abbasi et al. constructed a co-authorship network of papers published in information and library sciences, showing that researchers with a high betweenness centrality achieved good results [11]. Several methods have already been proposed to search for promising researchers using a co-authorship network. These include evaluation indexes that combine parameters such as the co-authorship of papers and the names of journals [12, 13], along with classification methods using machine learning [14, 15]. However, these studies do not focus on young researchers who have only just started to publish their research results, and they are not designed to find and actively evaluate promising young researchers and students.

In this paper, JSPS Research Fellows are taken to be promising young researchers. By comparing the characteristics of the temporal changes of the betweenness

centrality of the JSPS Research Fellows of co-authorship networks with their h -index, we determined whether the betweenness centrality of co-authorship networks is a useful index for evaluating and finding promising young researchers.

Analytical Data and Method

Academic Literature Database

To construct the co-authorship networks, we used an academic literature database called JSTPlus provided by the Japan Science and Technology Agency (JST). JST-Plus is a database of academic literature from more than 50 countries, covering all fields of science and technology. It contains over 24 million papers, and approximately 700,000 new papers are added annually.

We selected the biological sciences for our analysis, designated with the JST classification code “E”. All of the information was extracted for publications with classification code “E” issued in the 15 years from 2001 to 2015. The identification numbers (IDs) of the extracted documents totaled 1,920,191. IDs are assigned to authors, institutions, and cited literature, and name aggregation processing was applied to these IDs. Table 1 shows the number of paper IDs and author IDs extracted. The “New extractions” column in Table 1 shows the number of IDs extracted for the first time that year from among the author IDs totaled each year. The publication year of the cited literature is needed to calculate the time series variation of the h -index, so the corresponding document information was extracted separately.

Table 1 Number of paper IDs and author IDs extracted from the database

Year	Paper ID	Author ID		
		Yearly total	New extractions	JSPS
2001	115, 627	342, 216	342, 216	108
2002	110, 113	325, 499	252, 934	85
2003	115, 032	329, 202	236, 067	99
2004	118, 545	342, 507	242, 739	106
2005	119, 051	350, 396	251, 490	113
2006	122, 191	363, 198	252, 092	153
2007	132, 722	392, 749	264, 226	153
2008	130, 656	385, 763	248, 683	165
2009	138, 432	407, 731	262, 952	141
2010	137, 717	421, 509	270, 000	149
2011	138, 405	438, 580	283, 824	121
2012	131, 785	428, 624	275, 723	94
2013	147, 125	494, 480	334, 015	75
2014	136, 272	464, 333	306, 998	54
2015	126, 518	457, 627	315, 287	39
Total	1, 920, 191	Total	4, 139, 246	1655

Extraction of JSPS Research Fellows

The JSPS Fellowship Program is a subsidy program for young researchers who are Ph.D. candidates or postdoctoral researchers. The JSPS Fellowship Program has been ongoing since 1985. Selection is mainly based on screenings of proposals. JSPS asks the researchers selected for the program to indicate that they have received a JSPS grant at the time of publication, for example by noting “JSPS Research Fellow” or the equivalent when stating their affiliations in papers.

We extracted authors for whom the words “Japan Society for the Promotion of Science” or “JSPS” were included in the affiliated institution column of the database as researchers with a history of obtaining a JSPS Research Fellowship. As a result of searching the affiliated institution in JSTPlus, the total number of JSPS researchers extracted from 2001 to 2015 was 1655. Because an average of approximately 100 JSPS Research Fellows were selected from each year, we considered this to be a sufficient sample size with the above-mentioned method of checking the personal history. The “JSPS” column in Table 1 shows the number of IDs for which a history of a JSPS Research Fellows as of 2015 was confirmed from among the “New extractions” author IDs for the year. The reason for the small number in the “JSPS” column in Table 1 around 2015 is that the period for JSPS Research Fellows generally follows the next few years after a new selection.

Betweenness Centrality of Co-authorship Networks and Its Normalization

After constructing a bipartite graph with the paper ID and author ID as vertices, we constructed co-authorship networks that extracted the author IDs exclusively. We generated the co-authorship networks as an undirected, unweighted graph, forming one for each year. Therefore, the vertex representing the author ID is missing, unless the author published at least one paper annually. Edges representing co-authorship relationships cannot be maintained unless at least one paper is co-authored every year. The separated graphs that were not connected by edges were numbered Component 1, Component 2, and so on, in descending order of the number of vertices. Fig. 1 shows Component 4 (299 authors) and Component 17 (101 authors) for 2001 as examples of co-authorship networks.

The density of the vertices in Fig. 1 represents the magnitude of the betweenness centrality in the graph. For graph g consisting of n vertices, the betweenness centrality b_i of vertex i indicates the degree to which i is located on the shortest path of the combination of two vertices (j, k) other than i . This is expressed as

$$b_i = \sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}}, \quad (1)$$

where g_{jk} represents the number of shortest paths connecting vertex j and vertex k , and $g_{jk}(i)$ represents the number of paths passing through vertex i in g_{jk} . The

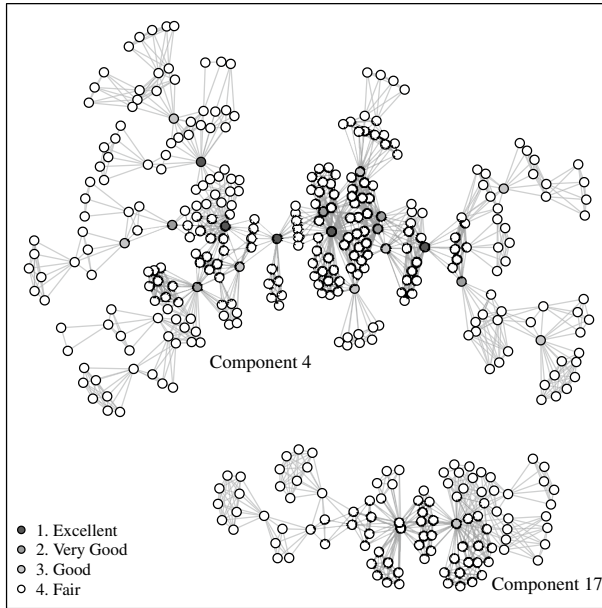


Fig. 1 Example of co-author networks of 400 authors

Table 2 Example of betweenness centrality of authors shown in Fig. 1

i	b_i	rank_1	y_i	c_i	rank_2	y_i^*
1	2.305×10^4	400	1.000	1	400	1.000
2	2.299×10^4	399	0.998	1	399	0.990
3	1.959×10^4	398	0.995	1	398	0.981
4	1.365×10^4	397	0.992	1	397	0.971
5	1.052×10^4	396	0.990	1	396	0.962
6	9.883×10^3	395	0.988	2	395	0.952
7	7.081×10^3	394	0.985	2	394	0.943
8	7.072×10^3	393	0.982	2	393	0.933
9	5.753×10^3	392	0.980	2	392	0.924
10	5.197×10^3	391	0.978	2	391	0.914
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
104	0.500×10^0	300	0.750	4	303	0.076
105	0.250×10^0	296	0.740	4	296	0.010
106	0.000×10^0	148	0.370	4	295	0.000
107	0.000×10^0	148	0.370	4	295	0.000
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
399	0.000×10^0	148	0.370	4	295	0.000
400	0.000×10^0	148	0.370	4	295	0.000

betweenness centrality of the co-author networks shown in Fig. 1 is obtained as shown in Table 2.

In Table 2, y_i , c_i , and y_i^* , together with b_i , represent values obtained by normalizing or classifying b_i with $n = 400$. There are several methods of normalizing the betweenness centrality. We opted to use normalization by ranking. Specifically, the following two methods were applied.

With the first normalization method, b_i is ranked in order of size as $n, (n - 1), \dots$, each year, and $\text{rank}_1(\cdot)$ is applied to give the mean ranking when the same ranking occurs. The normalized value y_i is then

$$y_i = \frac{\text{rank}_1(b_i)}{\max(\text{rank}_1(b))}. \tag{2}$$

In addition to being able to unify the maximum value in any year as $\max(y_i) = 1$, this method is useful for extracting a certain percentage of the upper class of people, such as those with a high betweenness centrality b_i in the top 1.0% (Excellent), top 2.5% (Very good), top 5.0% (Good), and below this (Fair). Let the class c_i of y_i be

$$c_i = \begin{cases} 1 & \text{if } 0.990 \leq y_i \leq 1.000, \\ 2 & \text{if } 0.975 \leq y_i < 0.990, \\ 3 & \text{if } 0.950 \leq y_i < 0.975, \\ 4 & \text{otherwise.} \end{cases} \tag{3}$$

First, we focus on researchers whose b_i is in the top 5% ($y_i \geq 0.95$). A cumulative logit model was applied to the time series variation of the class c_i . Let the probability $P(y = c | x)$ that the response y according to time series x is classified as class c be

$$P(y = c | x) = \begin{cases} \frac{\exp(\alpha_c + \beta_c x)}{1 + \sum_{l=1}^{m-1} \exp(\alpha_l + \beta_l x)} & \text{if } c < m, \\ \frac{1}{1 + \sum_{l=1}^{m-1} \exp(\alpha_l + \beta_l x)} & \text{if } c = m. \end{cases} \tag{4}$$

Here, α_l and β_l denote regression coefficients, and m represents the number of classes. Here, $m = 4$, as shown in Eq. (3). For time series x , newly extracted timing was employed, such that $x = 1$ from 2001 to 2015.

Although the normalization method described above is suitable for the analysis of the upper class of the betweenness centrality, it is unsuitable for analyzing the average transition of the whole. Owing to the structure of the co-authorship networks, many equivalent values such as $b_i = 0$ occur, the minimum value becomes $\min(y_i) \neq 0$, and the frequency distribution becomes discrete from this lowest level to the middle level.

Therefore, with the second normalization method, when the same rank occurs, $\text{rank}_2(\cdot)$ gives the maximum rank that is consecutive with the previous rank, and the normalized value y_i^* is given as:

Fig. 2 Transition of the betweenness centrality class *c*

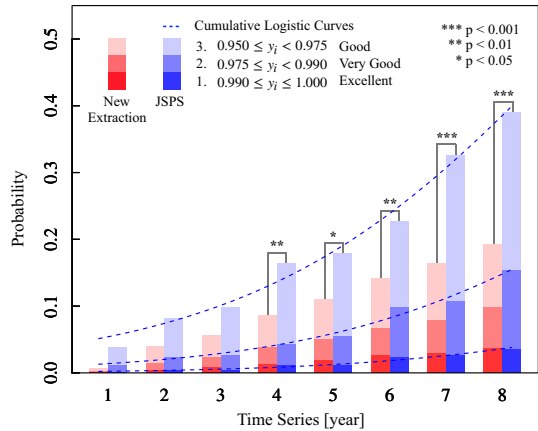


Table 3 Regression coefficients of cumulative logit model applied to class *c* of the JSPS Research Fellow group

	Estimate	Std. Error	<i>t</i> value	Pr(> <i>t</i>)
α_1	-3.540	0.1406	-25.181	< 0.0001
α_2	-4.887	0.2386	-20.482	< 0.0001
α_3	-6.515	0.4791	-13.597	< 0.0001
β_1	0.330	0.0249	13.255	< 0.0001
β_2	0.406	0.0397	10.224	< 0.0001
β_3	0.469	0.0764	6.144	< 0.0001

$$y_i^* = \frac{\text{rank}_2(b_i) - \min(\text{rank}_2(b))}{\max(\text{rank}_2(b)) - \min(\text{rank}_2(b))}. \tag{5}$$

When adopting this method, $\max(y_i^*) = 1$, $\min(y_i^*) = 0$, and the distribution is not discretized. The logistic growth curve is derived as follows:

$$\hat{y}^* = \frac{K}{1 + \exp\{-(\alpha + \beta x)\}}, \tag{6}$$

and fitted to the change of y_i^* according to time series *x*.

R 3.6.0 (data.table 1.12.2, Matrix 1.2.17, igraph 1.2.4.1, VGAM 1.1.1) was used for these analyses.

Results

Transition of the Upper Class of Betweenness Centrality

Figure 2 and Table 3 show the results of applying Eq. (4) to the transition of the upper class *c* of the betweenness centrality. The horizontal axis in Fig. 2 represents

the time series x , where its range is $1 \leq x \leq 8$. Here, the year when $x = 1$ ranges from 2004 to 2008. With the authors shown in Table 1, the data of the newly extracted 1,259,230 authors for the last 5 years were integrated, and the composition ratios of class c were tabulated. The red stacked bar chart shows the composition ratios of class c for the newly extracted researcher group. For the sake of comparison, the blue bars show the results of re-extracting data for 690 researchers with a personal history (until 2015) as a JSPS Research Fellow, and totaling the composition ratios again. Data with $x > 1$ have missing values in either group, but these are excluded from the composition ratios of the chart.

From Fig. 2, the probability is low so that the betweenness centrality falls within the top 5% of the total at the stage of $x = 1$, when newly extracted from the database for the first time. However, we confirmed the tendency for it to increase over time. In particular, the composition ratio of the upper class of JSPS Research Fellows tended to increase at an accelerating rate. As a result of applying the Kolmogorov–Smirnov (KS) test, the composition ratio showed a significant upward bias ($p < 0.05$) from the stage of $x \geq 4$, compared to the newly extracted researcher group. The cumulative logit model was applied to the transition of the composition ratio of the JSPS Research Fellow group, and a significant regression coefficient was shown ($p < 0.001$).

To summarize the above, the betweenness centrality of researchers, such as those selected as JSPS Research Fellows, increases at an accelerated rate compared to general new entrants. After approximately 4–8 years have passed since they were new entrants, it is highly probable that they fall within the top 5%. Specifically, it was shown that the growth of the normalization value y of the betweenness centrality proposed in this paper is possible to the 0.95 point.

Transition of Representative Values of Betweenness Centrality and h -Index

To construct a betweenness centrality growth model from representative values rather than the upper ranks of JSPS Research Fellows, and to compare its characteristics with the h -index, the author IDs with an h -index of 1 or more by 2015 were extracted again from the target author IDs, and the transition of the normalized betweenness centrality y^* and h -index was obtained. The transition of the h -index obtained at each time of each year was the subject of analysis, and the h -index at the newly extracted $x = 1$ was calculated by summing the number of citations in the period up to $x \leq 1$ for paper IDs published before $x \leq 1$. The h -index for $x = 2$ was calculated by summing the number of citations in the period up to $x \leq 2$ for the paper IDs published before $x \leq 2$.

There were 467,003 newly extracted researchers who met the extraction conditions, of which 463 were JSPS Research Fellows. Figure 3 shows the median value of y^* as a scatter diagram, and the results from applying the logistic curve with Eq. (6) to the transition. Figure 4 shows the mean value of the h -index as a scatter diagram and the results from applying a logistic curve similar to the transition. The blue plot represents the JSPS Research Fellow group, and the red plot represents the

Fig. 3 Transition in median of normalized betweenness centrality y^*

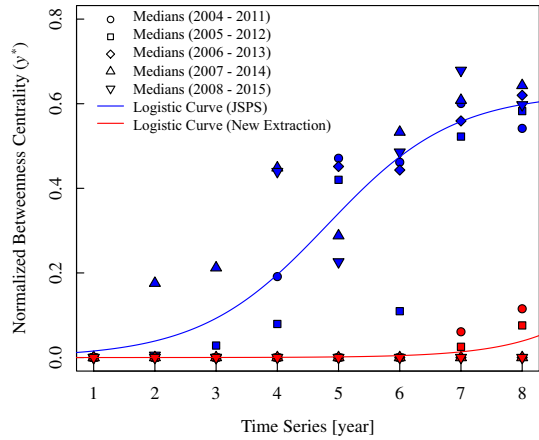


Table 4 Growth curve regression coefficients applied to the median of y^*

	Estimate	Std. Error	<i>t</i> value	Pr(> <i>t</i>)
(a) JSPS Research Fellow group (blue line)				
<i>K</i>	0.626	0.0654	9.566	< 0.0001
α	-3.674	0.7752	-4.740	< 0.0001
β	0.963	0.2484	3.878	< 0.0001
(b) Newly extracted researcher group (red line)				
<i>K</i>	-	-	-	-
α	-10.814	3.7964	-2.849	< 0.01
β	1.087	0.5526	1.967	n.s.

Fig. 4 Transition of mean value of *h*-index

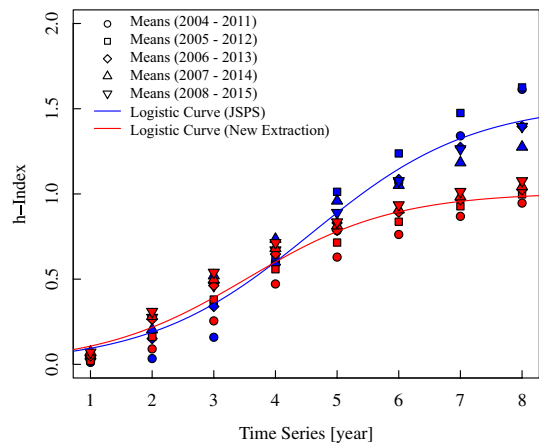


Table 5 Regression coefficients of growth curve applied to the mean of the h -index

	Estimate	Std. Error	t value	$\text{Pr}(> t)$
(a) JSPS Research Fellow group (blue line)				
K	1.534	0.0705	21.75	< 0.0001
α	-2.725	0.1892	-14.40	< 0.0001
β	0.762	0.0735	10.37	< 0.0001
(b) Newly extracted researcher group (red line)				
K	1.008	0.0347	29.10	< 0.0001
α	-2.969	0.2674	-11.11	< 0.0001
β	0.839	0.0883	9.50	< 0.0001

newly extracted researcher group. Tables 4 and 5 show the regression coefficients of the growth curves.

Figure 3 shows that the betweenness centrality y^* of the JSPS Research Fellow group grows at an accelerated rate over time and significantly fits to the logistic curve ($p < 0.001$). The regression coefficient K of this growth curve was estimated to be 0.626, as shown in Table 4, and it was shown to approach this value after 8 years. By contrast, the y^* of the newly extracted researcher group was almost 0, even after 8 years since the first academic publication, and the regression coefficient K could not be estimated, even when applying Eq. (6). Therefore, other regression coefficients were estimated at $K = 1$, but the regression coefficients β related to the rise of the curve were not significant ($p > 0.05$).

In contrast to this feature, Fig. 4 shows that the logistic curve fits well for the time series change of the h -index for both groups. It should be noted that there is no difference in growth between the two groups from new extraction to 4 years thereafter, and a difference of only approximately 0.5 arises even after 8 years. The h -index calculated from the number of academic publications and the number of citations is an integer value such as 0, 1, 2, ..., as defined. Therefore, the difference in the h -index value between the two groups of less than 1 is not sufficient to evaluate differences in these two groups of researchers.

According to these results, the h -index has a lagging property, as explained in Sect. 1. At first, no differences appeared between the general group of newly extracted researchers and the JSPS Research Fellow group. In contrast to the characteristic of a difference that slowly and steadily occurs, the betweenness centrality is an index that is able to show differences between the two groups at an early stage.

Extraction of a Researcher Group that Shows Correlation with the JSPS Research Fellow Betweenness Centrality Growth Model

We here propose using the Pearson product–moment correlation coefficient r as a simple method of widely searching the database for researchers whose betweenness centrality is increasing in the same way as the JSPS Research Fellow group. The procedure for the proposed method is as follows.

Step 1. Extract the document information for the specific academic field to be searched from the database and normalize it using Eq. (5) to calculate the betweenness centrality of the co-authorship networks. The betweenness centrality is calculated on a yearly basis and made into time series data.

Step 2. Build a JSPS Research Fellow model for the field. When constructing a median model, as shown by the blue line in Fig. 3, Eq. (6) is applied to approximately 8 years of time series data, to observe the accelerating increase to the betweenness centrality.

Step 3. Evaluate the correlation between the JSPS Research Fellow model and the n year data of the newly extracted researcher group. The threshold for the correlation coefficient r is the value at which the null hypothesis is rejected in a test of no correlation for a sample size n . That is, according to a t distribution of $n - 2$ degrees of freedom, the test statistic t_0 in the uncorrelated test is

$$t_0 = r \sqrt{\frac{n-2}{1-r^2}}. \quad (7)$$

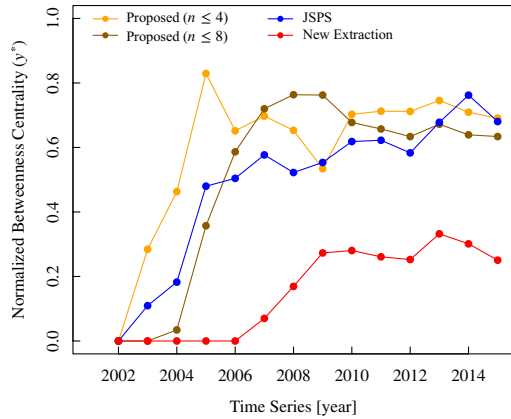
Therefore, this is solved for r , and it becomes the threshold value. For example, if $n = 4$ at a significance level of 5%, then researchers who show a correlation of $r > 0.900$ are extracted. If $n = 8$, then those with a correlation of $r > 0.621$ are extracted. This method can be applied even if part of the time series data is missing. Owing to the nature of Eq. (7), however, r cannot be evaluated unless $n > 2$.

Table 6 shows the number of author IDs correlated with the median model of the JSPS Research Fellow group constructed by the above procedure (Fig. 3 and Table 4). Figure 5 shows the betweenness centrality of researchers newly extracted in 2002 and the transition to the h -index by group. The values in parentheses in Table 6 indicate the number of newly extracted researchers who were confirmed as

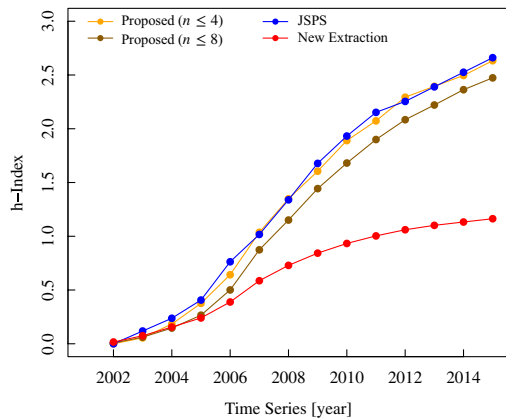
Table 6 Author ID search hits ($h \geq 1$)

Year	New extraction (JSPS)	Extraction by the proposed method (JSPS)		
		$n \leq 4$	$n \leq 6$	$n \leq 8$
2002	105, 811 (59)	109 (3)	783 (3)	1218 (11)
2003	97, 377 (78)	121 (1)	708 (10)	1144 (15)
2004	98, 480 (88)	110 (1)	696 (11)	1101 (21)
2005	104, 672 (80)	160 (3)	893 (10)	1285 (18)
2006	96, 325 (106)	147 (2)	690 (11)	1087 (24)
2007	89, 511 (98)	131 (3)	618 (12)	1025 (17)
2008	78, 015 (91)	103 (3)	562 (12)	884 (27)
2009	72, 410 (76)	88 (0)	627 (14)	–
2010	63, 086 (56)	106 (0)	480 (10)	–
2011	53, 864 (41)	98 (1)	–	–
2012	41, 485 (13)	61 (1)	–	–

Fig. 5 Comparison of the group extracted by the proposed method from the newly extracted researcher group in 2002 (yellow, dark yellow), JSPS Research Fellow group (blue), and newly extracted researchers with an h -index of 1 or more by 2015 (red)



(a) Median of normalized betweenness centrality y^*



(b) Mean value of h -index

JSPS Research Fellows by 2015. The target of extraction was author IDs with an h -index of 1 or more by 2015, as with the model described in Sect. 4.2. Data from 2004 to 2008 were used to build the model, and separator lines are included in the table. In addition, the newly extracted results for 2001 included many IDs registered in the database from before that. These were omitted from the table.

According to Table 6, when the search condition is $n \leq 4$, that is, when searching for researchers who published their first paper within 4 years, approximately 100 researchers were extracted per year, and when the search condition was $n \leq 8$, approximately 1000 researchers were extracted as promising young researchers. In addition, from Fig. 5a, the betweenness centrality of the group revealed by the proposed method transitions in the same way as the betweenness centrality of the JSPS Research Fellow group for that year. Regarding the h -index, we confirmed that the group revealed by the proposed method and the JSPS Research Fellow group transitioned in the same way, as shown in Fig. 5b.

The number of JSPS Research Fellows included in the group revealed by the proposed method was smaller than the population, as shown in parentheses in Table 6. This is because the proposed method uses only the statistical property of the betweenness centrality, and not an estimator trained by machine learning using multiple variables for the purpose of predicting future JSPS Research Fellows with high accuracy. Those who were extracted by the proposed method were found to be Research Fellows with RIKEN and the National Institute of Advanced Industrial Science and Technology—leading Japanese research institutions—and postdoctoral researchers from research institutions in other countries. The proposed method is thus considered to be effective as a means of widely searching for researchers who have not been accepted as a JSPS Research Fellow, yet have experience becoming a fellow through a similar peer review process, or people who may be evaluated by such a peer review process over the next few years.

Discussion

Evaluation Timing

The results of this study suggest that by using the proposed method, researchers with the same centrality transition characteristics as JSPS Research Fellows can be extracted earlier than they would be by using the h -index, provided that approximately 4 years of co-authorship information can be obtained. This is because citation-based indexes such as the h -index are lagging and require time for a striking difference to arise between excellent researchers and general researchers from citations by other researchers. By contrast, the centrality of co-authorship networks is an immediate index that is calculated concurrently with the publication of research papers. Furthermore, the betweenness centrality indicates whether a researcher is in a key position to connect research groups or researchers, and it shows the degree of contribution to the betweenness in a research community which is so important for achievements in collaborative research. Accordingly, as results of this study, we found that researchers extracted using the proposed method with the betweenness centrality are indeed promising, and that their h -index tends to grow in the same way as that of JSPS Research Fellows.

The JSPS Fellow system aims to train and secure excellent young researchers. A Research Fellow (DC) refers to a student who is enrolled in a doctoral program, and a Research Fellow (PD) refers to one who has recently obtained a doctoral degree (within five years). At present, the selection and evaluation of these young researchers is conducted by a peer review of the prospective researcher's "qualities as a researcher", "research plans", "ability to execute research plans", etc.

We considered these "young researchers" to be prospective JSPS Research Fellows: doctoral students or those who recently acquired their Ph.D. The proposed method is able to evaluate and extract young researchers who published their first paper within 4 years. For example, researchers who published their first paper in the second year of their doctoral program and spent three more years finishing the program correspond to these young researchers. At that stage, young researchers

are rarely evaluated highly by citation-based indices such as the *h*-index, and peer reviews are often used instead. However, it is generally difficult to guarantee a quantitative evaluation and one that offers wide coverage of the evaluation targets with a peer review process. The effort required of evaluators is moreover burdensome. By contrast, the proposed method is quantitative and offers broad coverage by using the betweenness centrality in co-authorship networks constructed from academic literature. As such, the burden on evaluators can be reduced.

Therefore, we expect that the proposed method can provide useful information to companies that are considering hiring researchers who are still doctoral candidates for future research and development, and to academic promotion organizations seeking promising young researchers.

Target Field for Evaluation

Research styles vary by field. For example, researchers tend to work alone in fields that are more theoretical, such as mathematics, and they tend to collaborate in fields that are more experimental, such as biology and physics. In interdisciplinary fields such as biotechnology and bioinformatics, it is considered effective to conduct collaborative research across multiple fields. In addition, JSPS Research Fellows must change their affiliated research institute, to create new research opportunities. These factors can influence the centrality of co-authorship networks. Furthermore, in the field of medicine, for example, research and development is conducted over a long period of time, and continual, protracted collaboration among researchers is often necessary. When taking into account these research styles, the centrality of the co-authorship networks—that is, collaborative networks—can be considered an index suitable for group research in general, and for interdisciplinary and multidisciplinary research in particular.

This paper focused on biology, where group research is common. We did not analyze the effectiveness of the proposed index in other research fields. In recent years, however, collaborative research such as open innovation, industry–academy cooperation, and interdisciplinary research has become more important. Thus, it is desirable to conduct similar analyses on the relationship between the centralities of co-authorship networks in other research fields and affiliated institutions.

Conclusion

To verify whether the betweenness centrality of co-authorship networks composed of academic literature databases is a useful index for quantitative and widely covered evaluations of promising young researchers, we analyzed the betweenness centrality of JSPS Research Fellows who were previously evaluated as promising according to a JSPS peer review process and who showed characteristics of time series variation with the *h*-index.

The betweenness centrality of the group assessed by JSPS as promising young researchers was compared to the centrality of a general group of researchers. We

clarified that the former has a high probability of being positioned at the top as a result of their first academic publication, and that a growth curve applies that increases at an accelerated rate in the first few years. In addition, whereas citation-based indexes such as the h -index indicate the impact of research results in a lagging manner, the betweenness centrality of co-authorship networks is an index that can be measured simultaneously with the publication of papers. Therefore, when the above two groups of researchers were evaluated with the betweenness centrality, differences between the two groups could be observed at earlier stages than evaluations with the h -index. The betweenness centrality of co-authorship networks was thus shown to be an index that can evaluate researchers ahead of the h -index. Furthermore, we demonstrated that if co-authorship information is obtained over approximately 4 years, we can extract from the database promising researchers whose h -index grows in the same way as a JSPS Research Fellow.

This paper focused on biology, where collaborative research is common. Given the increasing importance of collaboration in research activities such as open innovation and industry–academy cooperation, the betweenness centrality of co-authorship networks may be suitable for organizational research and interdisciplinary research such as bioengineering, bioinformatics or astrophysics. Consequently, it will be beneficial to conduct similar analyses in such research fields. Future developments include clarifying the causal relationship between changes to affiliated institutions and changes to the betweenness centrality of the co-authorship networks.

Acknowledgements In this study, we used the science and technology literature database provided by JST. In using the data, we received cooperation from G-Search Limited. We would like to thank all those involved from JST and G-Search Limited.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Fujita, M., Inoue, H., Terano, T.: Searching promising researchers through network centrality measures of co-author networks of technical papers. In: The 4th IEEE International COMPSAC Workshop on Social Services through Human and Artificial Agent Models (SSERV 2017) (2017)
2. Fujita, M., Inoue, H., Terano, T.: Evaluating researchers through betweenness centrality measures of co-author networks from academic literature database: finding gatekeeper researchers in organizational research. In: IEEE BigData 2018 The Third Workshop on Application of Big Data for Computational Social Science, Seattle (2018)
3. Garfield, E.: Citation analysis as a tool in journal evaluation. *Science* **178**(4060), 471–479 (1972)
4. Hirsch, J.E.: An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci.* **102**(46), 16569–16572 (2005)

5. Schreiber, M.: How relevant is the predictive power of the h-index? A case study of the time-dependent Hirsch index. *J. Informetrics* **7**(2), 325–329 (2013)
6. Zare, R.N.: Assessing academic researchers. *Angewandte Chemie Int. Ed.* **51**, 7338–7339 (2012)
7. Melin, G., Persson, O.: Studying research collaboration using co-authorships. *Cientometrics* **36**(3), 363–377 (1996)
8. Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. *Physica A* **311**(3–4), 590–614 (2002)
9. Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci.* **101**(Suppl. 1), 5200–5205 (2004)
10. Shinoda, K.: A family tree of artificial intelligence research in Japan [in Japanese]. *J. Jpn. Soc. Artificial Intell.* **26**(6), 584–589 (2011)
11. Abbasi, A., Chung, K.S.K., Hossain, L.: Egocentric analysis of co-authorship network structure, position and performance. *Inform. Process. Manag.* **48**, 671–679 (2012)
12. Li, X.-L., Foo, C.S., Tew, K.L., Ng, S.-K.: Searching for Rising Stars in Bibliography Networks. *LNCS* **5463**, 288–292 (2009)
13. Zhang, J., Xia, F., Wang, W., Bai, X., Yu, S., Bekele, T. M., Peng, Z.: CocaRank: A collaboration caliber-based method for finding academic rising stars. In: *Proceedings of the 25th International Conference Companion on WWW*, 395–400 (2016)
14. Daud, A., Ahmad, M., Malik, M.S.I., Che, D.: Using machine learning techniques for rising star prediction in co-author network. *Scientometrics* **102**, 1687–1711 (2015)
15. Panagopoulos, G., Tsatsaronis, G., Varlamis, I.: Detecting rising stars in dynamic collaborative networks. *J. Informetrics* **11**(1), 198–222 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.