大学評価・学位研究　第８号　平成20年12月（研究ノート・資料）
［独立行政法人大学評価・学位授与機構］

論文生産性および研究協力ネットワークに関する指標の相関分析

An analysis of the correlation among research productivity and collaboration network indices

芳鐘　冬樹
YOSHIKANE Fuyuki

Research on Academic Degrees and University Evaluation, No. 8 (December, 2008) [the essay/material]
National Institution for Academic Degrees and University Evaluation

1. Introduction ··· 45

2. The importance in the collaboration network ··· 46

3. Data ··· 48

4. Methodology ··· 49
4.1 Indices ··· 49
4.2 Analysis of correlation ··· 50

5. Results and discussion ··· 50
5.1 Correlation between researchers' properties in the same period ··· 51
5.2 Correlation between researchers' properties and their subsequent productivity ··· 52
5.3 Correlation between researchers' properties and their collaborators' productivity ··· 52

6. Conclusions ··· 53

ABSTRACT ··· 56

An analysis of the correlation among research productivity and collaboration network indices

YOSHIKANE Fuyuki *

Abstract

This article investigates which properties of researchers, including their importance in collaborative networks, have strong relationships with their own productivity or that of their collaborators in the domain of computer science. The results show that some indices representing researchers' properties have high correlation coefficients with productivity of themselves or their collaborators in the same period but, as for subsequent productivity, there is no index with a similarly high correlation coefficient. However, the importance as the leader, reflecting global structures of co-authorship networks, shows a relatively high correlation with collaborators' subsequent productivity, when we focus on the relationships between object researchers and "their cooperators", that is, the collaborators who have published co-authored papers with them "as the secondary author".

Keywords

Research collaboration; Co-authorship network; Research productivity

1. Introduction[1]

In this article, we investigate and clarify which properties of a researcher, including the importance in collaboration networks, have strong relationships with his own productivity or that of his collaborators. We deal with not only the synchronic correlation but also the diachronic one; that is, the relationships with their subsequent productivity. Our analysis, which clarifies the correlation between the researchers' properties and productivity, serves following purposes; to identify factors affecting productivity, and to grasp the characteristics of research domains in terms of knowledge production. Such information is also useful in more practical applications; for example, when a researcher searches his partners in research collaboration, and when one judges the substitutability between performance indices in research evaluation.

Collaboration networks, which represent cooperative relationships in research activities, are not observed only in authorship credits of co-authored papers. Some of them are observed in the acknowledgments of papers and not in authorship credits. However, we assume that authorship credits cover all collaborators that "substantially and technically" contribute to their co-authored papers, while acknowledgments are addressed only to subsidiary supporters (Cronin

* Assistant Professor, Department of Research for University Evaluation, National Institution for Academic Degrees and University Evaluation

1 This work was partially supported by Grant-in-Aid for Scientific Research (C) 20500225 (2008) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and we would like to show our gratitude to the support. An earlier version of this study, where some network indices are excluded, will be presented at *the 10th International Conference on Science and Technology Indicators*.

et al, 2003). There is a related problem concerning co-authorship; that is, honorific authorship, by which co-authorship credits are sometimes regarded to be irresponsible (Cason, 1992). Although many studies have pointed out the problem of honorary co-authors that have no substantial contribution to the work, ethical guidelines regarding authorship issues have been developed in recent years in many domains (e.g., ICMJE, 1997). These guidelines state that authorship credits should be determined by substantial and technical contributions to the work. Some questionnaire surveys illustrate that the majority of researchers have reached a consensus following these guidelines (Hoen et al., 1998; Bartle et al., 2000). Assuming that co-authorship credits appropriately represent substantial contributions, we regard the co-authorship network as the substantial collaboration network and measure the importance in co-authorship networks for each researcher.

In recent years, research collaboration has become more and more active. In many fields, especially in natural sciences, not only the ratio of co-authored papers but also the number of co-authors per paper is increasing (e.g., Adams et al., 2005). A large number of studies have analyzed the correlation of properties between authors and their co-authors. Some focused on the productivity of papers and others focused on affiliation, age, and sex, etc. (e.g., Kretschmer, 1994; 1997; Kretschmer and Gupta, 1998; Kundra and Kretschmer, 1999; Bahr and Zemon, 2000; Bozeman and Corley, 2004). Also, Yasuda (2004) reported that there is a positive correlation between the productivity or citation ratio of a researcher and the importance in the research collaboration network. These studies mainly examined the synchronic correlation between researchers' properties (both are of the same period). On the other hand, the diachronic correlation of properties, that is, the correlation between their subsequent and precedent activity, has not yet been sufficiently studied using quantitative methods. One of a few exceptions is Yoshikane et al. (2007), who analyzed the correlation between the researcher's productivity subsequent to a collaboration and the collaborator's precedent activity, targeting only newcomers and their senior collaborators. In this article, we extend their methodology and analyze the correlation between researchers' properties and each of the following four aspects of productivity: (1) productivity of themselves in the same period, (2) subsequent productivity of themselves, (3) their collaborators' productivity in the same period, and (4) their collaborators' subsequent productivity. We selected computer science as the object domain because, in this domain, research collaboration is very active and one should take into account collaboration (co-authorship) networks in the analysis of researchers' performance.

This article is organized as follows. First, in Section 2, we introduce the viewpoints from which to measure the importance in research collaboration networks for the preparation of adopting indices. In Section 3, the source data and the general conditions of the domain of computer science are described. After explaining our methodology, including indices, in Section 4, we then show and discuss the results of our experiments examining the correlations among researchers' properties in Sections 5 and 6.

2. The importance in the collaboration network

This study observes the status of a researcher in the research collaboration from two viewpoints: (1) the degree of importance in the co-authorship network as the first author and (2) that as the secondary author, who is an author of a co-authored paper but not the first author. We differentiate between these two viewpoints, based on the assumption that the first author designs the whole research as the leader and plays the special role, which is different from the roles of secondary authors. In some domains, including computer science, such a criterion is specified in guidelines (Zobel, 1999), and we also see this in the results of awareness surveys (Bridgwater et al., 1981). So, we can regard this assumption to be reasonable, at least to some extent. We measure the importance in the network not from a general viewpoint such as "power" (e.g., Batagelj and Mrvar, 2000) but from the two viewpoints, (1) and (2), separately on the basis of the idea that both the roles, as the leader in producing papers and as the cooperator collaborating with the leader, are meaningful in the network of research collaboration, and that these two roles are essentially different.

The global structure of the network as well as direct relations between authors should be considered when observing the importance in the co-authorship network because information, knowledge, or connections spread beyond direct ties. We assume the following network model, for setting the operational definitions and measures corresponding to the two viewpoints described above.

- Assuming directed graphs where the ties (links) are oriented from co-authors (secondary authors) to the first author for each paper (see Fig. 1). An author who has published a paper as the first author may possibly also be a secondary author in another paper. Some ties may be bidirectional.
- Assuming weighted graphs where the strength of co-authorship relations is taken into account.

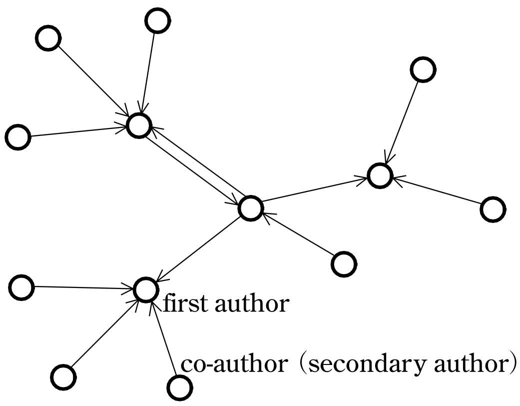

Fig. 1 An example of a co-authorship network.

Many indices have been proposed for measuring the strength of co-authorship relations between researchers, i.e., the weight of ties in graphs (e.g., Narin et al., 1991; Arunachalam et al., 1994). In this study we use, as the weight, the co-authorship frequency normalized for the number of authors on each co-authored paper, on the basis of the assumption that the strength of co-authorship relations between a pair of researchers grows in proportion to the number of times they have published together, and that the relationship between authors of a co-authored paper becomes less intimate as the number of authors increases.

We assume the above-mentioned weighted directed graphs, and calculate the importance as the leader/cooperator of each researcher in the network of research collaboration, paying attention to the number of collaborating partners, the relationship strength with each partner, and moreover the importance as the leader/cooperator of each partner. For each researcher n_i, the importance as the leader $CL^{<n_i>}$ and that as the cooperator $CF^{<n_i>}$ are obtained using the following operations, respectively (Yoshikane et al., 2006; 2007).

$$CL^{<n_i>} \leftarrow \sum_{n_j:(n_j,\, n_i)\in E} a_{i,j}\, CF^{<n_j>} \quad (1)$$

$$CF^{<n_i>} \leftarrow \sum_{n_j:(n_i,\, n_j)\in E} a_{j,i}\, CL^{<n_j>} \quad (2)$$

$a_{i,j}$ represents the value in cell (i,j) of the adjacency matrix of the network, and is equal to the relationship strength of the tie oriented from n_j to n_i, that is, the number of co-authored papers where n_i is the first author and n_j is a secondary author[2]. As mentioned earlier, we normalize the number of co-authored papers according to the number of authors on each paper. In operations (1) and (2), the mutual dependency that "a researcher who assists important leaders plays an important

[2] The value of diagonal cells $a_{i,i}$ is 0

role as the cooperator, and a researcher who organizes important cooperators plays an important role as the leader" is assumed. By repeating recursive substitution in (1) and (2), the global structure of the co-authorship network is reflected in the importance of each researcher.

The common idea that ties with more important nodes contribute to the importance more than those with less important nodes is shared among the centrality measure of Bonacich (1987, 2007), the PageRank algorithm (Brin and Page, 1998), the HITS (Hyperlink-Induced Topic Search) algorithm (Kleinberg, 1998), and so on. Of these, the HITS algorithm is most similar to *CL* and *CF* in that both of these assume two different roles considering the direction of relationships. Authorities and hubs in the HITS algorithm correspond to leaders and cooperators, respectively. In addition to the direction of relationships, the strength (weight) of relationships is taken into account in the calculation of *CL* and *CF*. After assigning 1 to $CL^{<n_i>}$ and $CF^{<n_i>}$ as the initial value, we calculate them by recursively repeating substitution and normalization of vectors in a manner analogous to the HITS algorithm, which is shown below.

Iterate (G, k)
  G: a collection of g researchers
  k: a natural number
  Let z denote the vector $(1, 1, 1, ..., 1) \in R^g$.
  Set $CL(0) := z$.
  Set $CF(0) := z$.
  For $i = 1, 2, ..., k$
    Apply the (1) operation to $(CL(i\text{-}1), CF(i\text{-}1))$, obtaining new *CL*-weights $CL^{new}(i)$.
    Apply the (2) operation to $(CL^{new}(i), CF(i\text{-}1))$, obtaining new *CF*-weights $CF^{new}(i)$.
    Normalize $CL^{new}(i)$, obtaining $CL(i)$.
    Normalize $CF^{new}(i)$, obtaining $CF(i)$.
  End
  Return $(CL(k), CF(k))$ (Kleinberg, 1998; Yoshikane et al., 2006).

In this study, substitution and normalization are repeated 10 times ($k = 10$).

3. Data

Our investigation targeted the domain of computer science, and used the SCI (Science Citation Index) CD/DVD-ROM version, provided by Thomson Scientific, for measuring the productivity of researchers in this domain and for observing their co-authorship networks. We regard the journals classified in the category of "computer science, theory & methods" in the list of source publications arranged by subject category, which is given on the Thomson Scientific web site: http://www.thomsonscientific.com/cgi-bin/jrnlst/jlsubcatg.cgi?PC=K, as the core journals in computer science. This category includes 37 journals (e.g., *Journal of Algorithms*).

The object researchers are those who have published in those journals at least one co-authored paper between 1996 and 2000. The reason why only the authors of co-authored papers are targeted is that the purpose of this study is to clarify the relationships of researchers' properties not only with the productivity of themselves but also with that of their co-authors. Thus, we focus on the researchers whose research styles are collaborative. In order to grasp the object researchers' and their co-authors' properties (the productivity of papers and the importance in the network) during a given period (1996 – 2000) and during the subsequent period (2001 – 2005), we extracted from SCI the bibliographic data of papers[3] published in the 37 journals over the ten-year-period (1996 – 2005).

It is necessary to identify authors' names, that is, to distinguish the same name for physically different persons and to integrate different names for the same person, when collecting data from bibliographic databases. However, descriptions in SCI are not always sufficient to strictly identify

[3] We extracted only original papers whose "document type" fields are "article" in the SCI database.

Table 1 Basic quantities of the data

	NP	*TA*	*DA*	DA_c	DA_{cf}	A_{av}	A_{sd}	P_{av}
'96 – '00	10181	23162	14483	13059	5518	2.28	1.29	1.60
'01 – '05	12772	31273	19792			2.45	1.46	1.58

NP: the number of papers;
TA: the total number of author tokens;
DA: the number of different authors;
DA_c: the number of authors who have published at least one co-authored paper;
DA_{cf}: the number of authors who have published at least one co-authored paper as the first author;
A_{av}: the average number of authors per paper (= *TA/NP*);
A_{sd}: the standard deviation of the number of authors;
P_{av}: the average number of papers per author (= *TA/DA*).

authors' names. In SCI, authors do not necessarily correspond one-to-one with affiliations in credits. Besides, in some cases an author has two or more affiliations, and in others there is no description of affiliations. For very frequently appearing authors, we manually checked their affiliations for identification, recognizing the above limitations. As for the others, misidentification was considered to not seriously affect the investigation of macroscopic tendencies in the targeted domain. Furthermore, we integrated variants of names, which are distinguished only by the existence of middle names or by differences in the use of upper and lower case letters.

Table 1 shows the basic quantities of the data. Of 14,483 authors who have published at least one paper between 1996 and 2000 (=*DA* ('96 – '00)), 13,059 are the object researchers in this study, that is, those who have published at least one "co-authored" paper. More than 90 percent of all the authors have published co-authored papers. We can reconfirm that computer science is a domain whose research style is collaborative. Of the 13,059 object researchers, 5,518 have published at least one co-authored paper as the first author.

4. Methodology

4.1 Indices

We divided the data into two periods (1996 – 2000 and 2001 – 2005), and investigated the relationships between (1) researchers' productivity and importance in the collaboration network in the first period and (2) the productivity, in the same period (i.e., the first period) or in the subsequent period (i.e., the second period), of the researchers themselves or of their collaborators who have published co-authored papers with them during the first period.

As researchers' productivity, its quantitative aspect, i.e., the number of produced papers, was considered in this study. For counting the number of papers, we adopted the normal count, the adjusted count, and the straight count. In the adjusted count the number of papers is normalized for the number of authors of each co-authored paper, in the straight count only papers published as the first author are counted, and in the normal count every author is counted fully whether or not there is co-authorship (Lindsey, 1980). *NOR* (normal count), *ADJ* (adjusted count), and *STR* (straight count) represent the number of papers measured by each method, respectively. *ADJ* was used to examine the productivity considering the proportion of contribution in research collaboration, while *STR* was used to examine the productivity as the leader of research.

To measure the importance in the network, we developed weighted directed graphs where the ties are oriented from secondary authors to the first author for each paper and where the strength of co-authorship relations is taken into account. As basic indices considering only direct ties, we used indegree *DI* and outdegree *DO*, that is, the number of researchers who have published co-

Table 2 Indices for measuring researchers' properties

			Previous period		Latter period (subsequent to their collaboration)	
			Object researchers	Their collaborators	Object researchers	Their collaborators
Productivity	Normal count		NOR_{r1}	NOR_{c1}	NOR_{r2}	NOR_{c2}
	Adjusted count		ADJ_{r1}	ADJ_{c1}	ADJ_{r2}	ADJ_{c2}
	Straight count		STR_{r1}	STR_{c1}	STR_{r2}	STR_{c2}
Importance in the collaboration network	Indegree		DI_{r1}	DI_{c1}	DI_{r2}	DI_{c2}
	Outdegree		DO_{r1}	DO_{c1}	DO_{r2}	DO_{c2}
	Importance considering global structures	(leader)	CL_{r1}	CL_{c1}	CL_{r2}	CL_{c2}
		(cooperator)	CF_{r1}	CF_{c1}	CF_{r2}	CF_{c2}

authored papers in which the object researcher is the first author and the number of researchers who have published co-authored papers as the first author with the object researcher, respectively. In addition, *CL* and *CF*, which are discussed in Section 2, were adopted as indices of the importance in the global structure including indirect ties. *DI* and *CL* correspond to the importance as the leader (first author), while *DO* and *CF* correspond to that as the cooperator (secondary author). We also used *CL'* and *CF'*, the variants of *CL* and *CF*, where the co-authorship frequency are not normalized for the number of authors in the calculation of the weight of ties.

Table 2 shows the list of the indices used in this study. Subscripts attached to the indices signify the first ("1") or second ("2") period, and object researchers ("*r*") or their collaborators ("*c*"). Thus, for example, NOR_{r1} represents the productivity in the first period of an object researcher measured by the normal count.

4.2 Analysis of correlation

We examined which property of a researcher has a strong correlation with productivity of himself/his collaborators in the same/subsequent period. To put it concretely, we calculated the correlation coefficients between the indices of the object researchers' properties [NOR_{r1}, ADJ_{r1}, STR_{r1}, DI_{r1}, DO_{r1}, CL_{r1}, CF_{r1}, CL'_{r1}, CF'_{r1}] and each of the following two families of indices regarding productivity.

(1) Indices regarding the productivity of the researchers themselves in the same/subsequent period [$NOR_{r1/2}$, $ADJ_{r1/2}$, $STR_{r1/2}$]

(2) Indices regarding their collaborators' productivity in the same/subsequent period [$NOR_{c1/2}$, $ADJ_{c1/2}$, $STR_{c1/2}$]

As these indices are expected to follow not the normal distribution but the power law distribution, we used Spearman's rank-order correlation coefficient r_s.

As for the correlation with the collaborators' productivity [$NOR_{c1/2}$, $ADJ_{c1/2}$, $STR_{c1/2}$], we calculated the correlation coefficients with each of the mean, minimum, and maximum values among all collaborators, considering the case that a researcher has more than one collaborator. In addition, we recalculated the correlation coefficients, taking into account only the collaborators who have published co-authored papers where an object researcher is the first author, to infer the effects of the leader of research (first author) on their cooperators (secondary authors).

5. Results and discussion

The results of calculating the correlation coefficients between the family of indices to measure the properties of object researchers, who have published at least one paper in the previous

Table 3 Correlation coefficients between indices (N=13,059)

		NOR_{r1}	ADJ_{r1}	STR_{r1}	DI_{r1}	DO_{r1}	CL_{r1}	CF_{r1}	CL'_{r1}	CF'_{r1}
NOR_{r1}			0.769*	0.412*	0.254*	0.412*	0.289*	0.380*	0.268*	0.296*
ADJ_{r1}		0.769*		0.476*	0.265*	0.166*	0.382*	0.226*	0.302*	−0.075*
STR_{r1}		0.412*	0.476*		0.905*	−0.545*	0.911*	−0.448*	0.900*	−0.494*
NOR_{r2}		0.419*	0.370*	0.208*	0.144*	0.182*	0.169*	0.200*	0.158*	0.132*
ADJ_{r2}		0.416*	0.376*	0.214*	0.144*	0.171*	0.171*	0.191*	0.158*	0.117*
STR_{r2}		0.296*	0.286*	0.298*	0.237*	−0.009	0.261*	0.035*	0.246*	−0.014
NOR_{c1}	(mean)	0.378*	0.261*	0.168*	0.121*	0.147*	0.246*	0.390*	0.228*	0.323*
NOR_{c1}	(min)	0.040*	0.215*	0.117*	0.037*	−0.147*	0.183*	0.065*	0.142*	−0.063*
NOR_{c1}	(max)	0.440*	0.224*	0.151*	0.123*	0.237*	0.220*	0.446*	0.215*	0.427*
ADJ_{c1}	(mean)	0.343*	0.503*	0.254*	0.153*	0.030*	0.324*	0.327*	0.252*	0.070*
ADJ_{c1}	(min)	−0.027*	0.465*	0.199*	0.075*	−0.275*	0.238*	−0.049*	0.146*	−0.383*
ADJ_{c1}	(max)	0.448*	0.397*	0.211*	0.145*	0.179*	0.277*	0.431*	0.234*	0.284*
STR_{c1}	(mean)	0.270*	0.304*	−0.294*	−0.360*	0.540*	−0.280*	0.687*	−0.299*	0.452*
STR_{c1}	(min)	−0.087*	0.245*	−0.201*	−0.265*	0.103*	−0.199*	0.232*	−0.223*	−0.062*
STR_{c1}	(max)	0.386*	0.180*	−0.261*	−0.302*	0.618*	−0.248*	0.726*	−0.251*	0.667*
NOR_{c2}	(mean)	0.288*	0.218*	0.133*	0.099*	0.126*	0.179*	0.281*	0.168*	0.231*
NOR_{c2}	(min)	−0.065*	0.157*	0.087*	0.021	−0.169*	0.136*	−0.013	0.095*	−0.132*
NOR_{c2}	(max)	0.374*	0.193*	0.123*	0.107*	0.223*	0.161*	0.342*	0.163*	0.343*
ADJ_{c2}	(mean)	0.287*	0.225*	0.132*	0.095*	0.126*	0.174*	0.277*	0.161*	0.220*
ADJ_{c2}	(min)	−0.066*	0.161*	0.087*	0.020	−0.170*	0.135*	−0.015	0.093*	−0.137*
ADJ_{c2}	(max)	0.373*	0.202*	0.121*	0.101*	0.223*	0.156*	0.340*	0.155*	0.334*
STR_{c2}	(mean)	0.288*	0.190*	0.023*	−0.007	0.242*	0.030*	0.350*	0.025*	0.303*
STR_{c2}	(min)	−0.065*	0.134*	−0.010	−0.060*	−0.063*	0.006	0.046*	−0.023*	−0.075*
STR_{c2}	(max)	0.354*	0.184*	0.031*	0.010	0.303*	0.034*	0.389*	0.036*	0.379*

* significant (p< .01)

period, and the family of indices to measure the productivity of themselves/their co-authors in the same/subsequent period are shown in Table 3. Significant values (p< .01) are marked with asterisk symbols (*).

5.1 Correlation between researchers' properties in the same period

The correlation between the properties, including productivity, of researchers themselves in the same period can be summarized as follows. ADJ_{r1} was very highly correlated with NOR_{r1} (r_s=0.769), indicating that substitution of the normal count for the adjusted count is somewhat reliable in the evaluation of research performance. With regard to STR_{r1}, however, the correlation coefficients with the other counting methods, NOR_{r1} and ADJ_{r1}, were not so high (r_s=0.412, 0.476). It was reconfirmed that one ought to use the straight count instead of the normal count or the adjusted count when focusing on the performance as the leader of research (the first author).

On the other hand, regarding the correlation between researchers' productivity and their importance in the network, DI_{r1}, CL_{r1}, and CL'_{r1} (i.e., indices measuring the importance as the leader) were very strongly correlated with STR_{r1} (r_s>0.9) while DO_{r1}, CF_{r1}, and CF'_{r1} (i.e., indices measuring the importance as the cooperator) were less well correlated with the number of papers for every counting method. That is to say, the number of produced papers cannot be a good indicator for evaluating researchers' activity from the latter viewpoint.

5.2 Correlation between researchers' properties and their subsequent productivity

Regarding the correlation between the properties of researchers and the subsequent productivity of themselves, all the values of the correlation coefficient were less than 0.42, that is, there are no combinations of indices that have a particularly high correlation. Compared with the other indices for the first period, NOR_{r1} was relatively highly correlated with all the indices measuring researchers' productivity for the second (subsequent) period, i.e., NOR_{r2}, ADJ_{r2}, and STR_{r2} (r_s is around 0.3 to 0.4). Although NOR_{r1} and STR_{r1} had similar values as regards the correlation with STR_{r2} ($r_s \approx 0.3$), the correlation coefficients of the latter with NOR_{r2} and ADJ_{r2} were low ($r_s = 0.208$, 0.214). Judging from these results, the normal count, which is the simplest method to count papers, was most useful as an indicator to generally predict the future performance of a researcher, including, in the case of predicting the productivity considering the proportion of contribution in research collaboration or predicting the productivity as the leader of research.

5.3 Correlation between researchers' properties and their collaborators' productivity

As for the correlation of researchers' properties with their co-authors' productivity in the same period, some combinations of indices had relatively high correlation coefficients. In particular, the correlation coefficient between CF_{r1} (i.e., the importance of researchers as the leader reflecting the global structures of co-authorship

Table 4 Correlation coefficients between indices, taking into account only the collaborators who have published co-authored papers where an object researcher is the first author ($N = 5{,}518$)

		NOR_{r1}	ADJ_{r1}	STR_{r1}	DI_{r1}	DO_{r1}	CL_{r1}	CF_{r1}	CL'_{r1}	CF'_{r1}
NOR_{r1}			0.858 *	0.756 *	0.268 *	0.777 *	0.421 *	0.772 *	0.323 *	0.768 *
ADJ_{r1}		0.858 *		0.696 *	−0.166 *	0.625 *	0.447 *	0.623 *	0.040 *	0.609 *
STR_{r1}		0.756 *	0.696 *		0.324 *	0.281 *	0.450 *	0.293 *	0.300 *	0.283 *
NOR_{r2}		0.464 *	0.412 *	0.369 *	0.148 *	0.390 *	0.264 *	0.391 *	0.211 *	0.386 *
ADJ_{r2}		0.453 *	0.412 *	0.367 *	0.129 *	0.373 *	0.257 *	0.374 *	0.193 *	0.367 *
STR_{r2}		0.336 *	0.317 *	0.328 *	0.117 *	0.225 *	0.221 *	0.231 *	0.155 *	0.226 *
NOR_{c1}	(mean)	0.359 *	0.281 *	0.306 *	0.141 *	0.271 *	0.784 *	0.291 *	0.704 *	0.287 *
NOR_{c1}	(min)	0.155 *	0.253 *	0.074 *	−0.310 *	0.148 *	0.526 *	0.162 *	0.325 *	0.155 *
NOR_{c1}	(max)	0.394 *	0.256 *	0.355 *	0.297 *	0.288 *	0.785 *	0.307 *	0.768 *	0.305 *
ADJ_{c1}	(mean)	0.327 *	0.415 *	0.296 *	−0.157 *	0.231 *	0.800 *	0.253 *	0.435 *	0.241 *
ADJ_{c1}	(min)	0.084 *	0.381 *	0.030	−0.663 *	0.081 *	0.441 *	0.094 *	−0.061 *	0.081 *
ADJ_{c1}	(max)	0.387 *	0.365 *	0.365 *	0.095 *	0.266 *	0.824 *	0.287 *	0.589 *	0.279 *
STR_{c1}	(mean)	0.297 *	0.214 *	0.221 *	0.182 *	0.257 *	0.424 *	0.270 *	0.392 *	0.264 *
STR_{c1}	(min)	0.074 *	0.175 *	−0.028	−0.304 *	0.119 *	0.196 *	0.127 *	0.024	0.118 *
STR_{c1}	(max)	0.328 *	0.209 *	0.268 *	0.284 *	0.269 *	0.441 *	0.283 *	0.448 *	0.278 *
NOR_{c2}	(mean)	0.246 *	0.202 *	0.225 *	0.123 *	0.186 *	0.525 *	0.205 *	0.477 *	0.200 *
NOR_{c2}	(min)	0.036 *	0.170 *	-0.020	−0.336 *	0.064 *	0.330 *	0.076 *	0.128 *	0.069 *
NOR_{c2}	(max)	0.298 *	0.189 *	0.293 *	0.288 *	0.211 *	0.535 *	0.229 *	0.559 *	0.227 *
ADJ_{c2}	(mean)	0.245 *	0.208 *	0.229 *	0.111 *	0.180 *	0.510 *	0.199 *	0.453 *	0.193 *
ADJ_{c2}	(min)	0.034	0.173 *	−0.020	−0.347 *	0.061 *	0.322 *	0.072 *	0.112 *	0.065 *
ADJ_{c2}	(max)	0.296 *	0.197 *	0.296 *	0.270 *	0.203 *	0.522 *	0.222 *	0.532 *	0.218 *
STR_{c2}	(mean)	0.226 *	0.158 *	0.217 *	0.192 *	0.157 *	0.310 *	0.169 *	0.303 *	0.165 *
STR_{c2}	(min)	0.021	0.127 *	-0.037 *	-0.277 *	0.053 *	0.137 *	0.061 *	-0.019	0.055 *
STR_{c2}	(max)	0.257 *	0.158 *	0.260 *	0.279 *	0.171 *	0.324 *	0.183 *	0.350 *	0.180 *

* significant ($p < .01$)

networks) and the mean of STR_{c1} (the productivity of their co-authors as the first author) was high (r_s =0.678). On the other hand, as for the correlation with the "subsequent" productivity of their co-authors, no combinations of indices had a high correlation (r_s<0.4).

Table 4 shows the correlation coefficients between indices focusing only on the relationships between researchers and their cooperators (secondary authors). It was CL_{r1} and CL'_{r1} (i.e., indices representing the importance as the leader reflecting the global structures of co-authorship networks) that showed the highest correlation with the subsequent productivity of their cooperators. More precisely, as for the mean values of productivity among all the cooperators, the correlation coefficients with CL_{r1} (where the co-authorship frequency is normalized) were relatively high while, as for the maximum values, the correlation coefficients with CL'_{r1} (where the co-authorship frequency is not normalized) were also relatively high. In particular, CL_{r1} was highly correlated with both NOR_{c2} (by the normal count) and ADJ_{c2} (by the adjusted count) (r_s>0.5). On the other hand, there was little correlation between indegree DI_{r1} of researchers and the subsequent productivity of their cooperators. The correlation coefficients between DI_{r1} and the mean values of NOR_{c2}, ADJ_{c2}, or STR_{c2} are no more than around 0.15. *DI*, as well as *CL*, is an index that measures the importance as the leader in the collaboration network but it only takes into account direct ties between researchers. The results imply that, if we are to predict, on the basis of properties of a researcher, the subsequent productivity of his cooperators, we should take into account the network's global structure.

6. Conclusions

This study investigated the correlation between properties to clarify which properties of a researcher have strong relationships with his own productivity or that of his collaborators. The results show that some of the indices representing researchers' properties were highly correlated with their own productivity in the same period or that of their collaborators. For example, *CL* proposed by Yoshikane et al. (2006) and indegree (i.e., indices measuring the importance as the leader) were very strongly correlated with the number of produced papers for the straight count (where only papers published as the first author are counted). In contrast, for subsequent productivity, we found no index with a similarly high correlation coefficient. However, *CL* was relatively highly correlated with collaborators' subsequent productivity, when we focused on the relationships between object researchers and "their cooperators", that is, the collaborators who have published co-authored papers with them "as the secondary author". *CL* is a measure that takes into account not only the number of collaborators but also their importance as the cooperator; in other words, it represents "the achievements of the object researcher in organizing researchers who are actively performing research as the cooperator". This implies the possibility that such achievements (or experiences) of researchers affect the subsequent productivity of their cooperators.

Since research styles, particularly customs concerning co-authorship, vary with the domain, the tendencies on correlation between researchers' properties and productivity may also vary with the domain. The tendencies shown in this study should be understood as the characteristics of computer science. We will analyze co-authorship networks of other domains and compare those results in further studies, for grasping the characteristics of each research domain.

Acknowledgement

I would like to thank Takafumi Suzuki of the University of Tokyo and Keita Tsuji of the University of Tsukuba. I would also like to thank Takayuki Nozawa of the Tokyo University of Agriculture and Technology and Susumu Shibui of the Kagoshima University for technical supports.

References

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: evidence from US universities, 1981-1999. Research Policy 34, 259-285.

Arunachalam, S., Srinivasan, R., Raman, V., 1994. International collaboration in science: participation by the Asian giants. Scientometrics 30, 7-22.

Bahr, A.H., Zemon, M., 2000. Collaborative authorship in the journal literature: perspectives for academic librarians who wish to publish. College & Research Libraries 61, 410-419.

Bartle, S.A., Fink, A.A., Hayes, B.C., 2000. Psychology of the scientist: LXXX. attitudes regarding authorship issues in psychological publications. Psychological Reports 86, 771-788.

Batagelj, V., Mrvar, A., 2000. Some analyses of Erdös collaboration graph. Social Networks 22, 173-186.

Bonacich, P., 1987. Power and centrality: a family of measures. The American Journal of Sociology 92, 1170-1182.

Bonacich, P., 2007. Some unique properties of eigenvector centrality. Social Networks 29, 555-564.

Bozeman, B., Corley, E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. Research Policy 33, 599-616.

Bridgwater, C.A., Bornstein, P.H., Walkenbach, J., 1981. Ethical issues and the assignment of publication credit. American Psychologist 36, 524-525.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. In: Proceedings of 7th International World Wide Web Conference. Elsevier, Amsterdam, pp. 101-117.

Cason, J.A., 1992. Authorship trends in poultry science, 1981 through 1990. Poultry Science 71, 1283-1291.

Cronin, B., Shaw, D., Barre, K.L., 2003. A cast of thousands: coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. Journal of the American Society for Information Science and Technology 54, 855-871.

Hoen, W.P., Walvoort, H.C., Overbeke, A.J.P.M., 1998. What are the factors determining authorship and the order of the authors' names?: a study among authors of the Nederlands Tijdschrift voor Geneeskunde (Dutch Journal of Medicine). Journal of the American Medical Association 280, 217-218.

ICMJE (International Committee of Medical Journal Editors), 1997. Uniform requirements for manuscripts submitted to biomedical journals. Journal of the American Medical Association 277, 927-934.

Kleinberg, J.M., 1998. Authoritative sources in a hyperlinked environment. In: Proceedings of 9th Annual ACM-SIAM Symposium on Discrete Algorithms. Association for Computing Machinery, New York, pp. 668-677.

Kretschmer, H., 1994. Coauthorship networks of invisible-colleges and institutionalized communities. Scientometrics 30, 363-369.

Kretschmer, H., 1997. Patterns of behaviour in coauthorship networks of invisible colleges. Scientometrics 40, 579-591.

Kretschmer, H., Gupta, B.M., 1998. Collaboration patterns in theoretical population genetics. Scientometrics 43, 455-462.

Kundra, R., Kretschmer, H., 1999. A new model of scientific collaboration Part 2: collaboration patterns in Indian medicine. Scientometrics 46, 519-528.

Lindsey, D., 1980. Production and citation measures in the sociology of science: the problem of multiple authorship. Social Studies of Science 10, 145-162.

Narin, F., Stevens, K., Whitlow, E.S., 1991.

Scientific cooperation in Europe and the citation of multinationally authored papers. Scientometrics 21, 313-323.

Yasuda, Y., 2004. The Science of Networking. Nikkei, Tokyo [in Japanese].

Yoshikane, F., Nozawa, T., Tsuji, K., 2006. Comparative analysis of co-authorship networks considering authors' roles in collaboration: differences between the theoretical and application areas. Scientometrics 68, 643-655.

Yoshikane, F., Nozawa, T., Shibui, S., Suzuki, T., 2007. An analysis of the connection between researchers' productivity and their co-authors' past attributions, including the importance in collaboration networks. In: Torres-Salinas, D., Moed, H.F. (Eds.), Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics. Spanish Research Council, Madrid, pp. 783-791.

Zobel, J., 1999. Guideline on Research Practice in Computer Science. http://goanna.cs.rmit.edu.au/~jz/conduct.html.

（受稿日　平成20年７月２日）
（受理日　平成20年９月24日）

［要旨］

論文生産性および研究協力ネットワークに関する指標の相関分析

芳鐘　冬樹*

研究者のどの属性が，本人や研究協力相手の論文生産性と高い関連性があるかを特定し，指標の代替可能性やパフォーマンスの予測可能性についての知見を得ることを目的として，計算機科学分野を対象に，論文生産性および研究協力ネットワークに関する指標の相関について調査した。同時期の発表論文数に関しては，高い相関を示す指標も存在したが，後の発表論文数に関しては，本人のものにも，研究協力相手のものにも，特に高い相関を示す指標はなかった。ただし，自身が第1著者であるときの共著者に限定した場合，ネットワークの大域的構造を反映した中心性指標が，共著者のその後の発表論文数と，比較的高い相関を持つことが明らかになった。

*　大学評価・学位授与機構　評価研究部　助教