Investigation of the correlation between researchers' properties and productivity through analyzing co-authorship networks

Fuyuki Yoshikane*

National Institution for Academic Degrees and University Evaluation, 1-29-1 Gakuen-nishimachi, Kodaira, Tokyo, 187-8587, Japan.
Tel.: +81-42-353-1824; Fax: +81-42-353-1863; E-mail: fuyuki@niad.ac.jp

Takafumi Suzuki

University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan.
E-mail: qq16116@iii.u-tokyo.ac.jp

Keita Tsuji

University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8571, Japan.
E-mail keita@slis.tsukuba.ac.jp
* Corresponding author

Theme: Theme 1; Theme 3

Keywords: co-authorship network; research collaboration; research productivity; performance prediction

1 Background

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. Our analysis 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.

A large number of studies have analyzed the correlation of properties between authors and their co-authors [e.g., Kundra and Kretschmer, 1999; Bahr and Zemon, 2000]. 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. 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. 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.

2 Methodology

Data

Our investigation targeted the domain of computer science. We used SCI and regarded the 37 journals classified in the category of "computer science, theory & methods", as the core journals in computer science. The number of the object researchers in our investigation, who have published in those journals at least one co-authored paper between 1996 and 2000, is 13,059. In order to grasp the object researchers' and their co-authors' properties (i.e., 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 papers published over the ten-year-period.

Indices

We divided the data into two periods (1996–2000 and 2001–2005), and calculated the correlation coefficients r 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.

Table 1 Indices for measuring researchers' properties

			First period		Second 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}

To measure the importance in the network, we developed directed graphs where the ties are oriented from secondary authors to the first author for each paper. As basic indices considering only direct ties, we used indegree and outdegree. In addition, *CL* and *CF* proposed by Yoshikane et al. [2006] were adopted as indices of the importance in the global structure including indirect ties. Table 1 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*").

3 Resultls and conclusion

Some of the indices representing researchers' properties were highly correlated with their own productivity in the same period or that of their collaborators. For instance, the correlation coefficient between CF_{r1} (i.e., the importance of researchers as the cooperator reflecting the global structures of co-authorship networks) and STR_{c1} (the productivity of their collaborators as the first author) was about 0.7.

In contrast, for subsequent productivity, we found no index with a similarly high correlation coefficient. However, CL_{r1} 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". In particular, CL_{r1} was highly correlated with NOR_{c2} (by the normal count) and ADJ_{c2} (by the adjusted count) (r>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 of DI_{r1} with NOR_{c2} (by the normal count), ADJ_{c2} (by the adjusted count), or STR_{c2} (by the straight count) 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.

References

KUNDRA, R.; KRETSCHMER, H. (1999), A new model of scientific collaboration Part 2: Collaboration patterns in Indian medicine, Scientometrics, 46 : 519–528.

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: D. TORRES-SALINAS, H. F. MOED (Eds), Proceedings of 11th International Conference of the International Society for Scientometrics and Informetrics. Spanish Research Council, pp. 783–791.

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.