# Estimating the Influence of Researchers' and their Collaborators' Performance on their Future Productivity

Fuyuki Yoshikane<sup>1</sup>, Takafumi Suzuki<sup>2</sup> and Shuntaro Kawamura<sup>3</sup>

<sup>1</sup> fuyuki@niad.ac.jp National Institution for Academic Degrees and University Evaluation, 1-29-1 Gakuen-nishimachi, Kodaira, Tokyo 187-8587 (Japan)

> <sup>2</sup> qq16116@iii.u-tokyo.ac.jp, <sup>3</sup> zuiken2514@yahoo.co.jp University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033 (Japan)

#### Introduction

This study investigates and clarifies the correlation between researchers' precedent and subsequent performance. Not only productivity but also the importance in collaboration networks is considered as researchers' performance. Our analysis, which estimates the influence of researchers' and their collaborators' performance on their future productivity, serves to grasp the characteristics of research domains in terms of knowledge production.

Although a large number of studies have analyzed the synchronic correlation of properties between authors and their co-authors (e.g., Bozeman & Corley, 2004), the diachronic correlation of properties, that is, the correlation between their subsequent and precedent activity, has not yet been sufficiently studied. Yoshikane et al. (2007) analyzed the relationship between the researcher's productivity and their collaborator's precedent activity. However, their analysis was limited in scope of the covered targets, i.e., they focused only on the relationship between newcomers and their senior collaborators. Instead, this study covers all researchers and examines contribution of researchers' and their co-authors' performance to their subsequent productivity, on the basis of not only the comparison of the correlation coefficient between variables but also multiple regression analysis.

# Data

Our investigation targeted the domain of computer science. The object researchers are those who have published at least one co-authored paper between 1996 and 2000. In order to grasp the object researchers' and their co-authors' properties during a given period (1996-2000) and during the subsequent period (2001-2005), we extracted from SCI the bibliographic data of papers published 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. 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. Furthermore, for very frequently appearing authors, we manually checked their affiliations for identification.

There are 14,483 authors who have published at least one paper in the first period. Of them, 13,059 are the object researchers in this study, that is, those who have published at least one "co-authored" paper.

# Methodology

For counting the number of papers, we adopted the normal count, the adjusted count, and the straight count. As for network indices, we used indegree and outdegree. In addition, *CL* and *CF* proposed by Yoshikane et al. (2007) were adopted to measure the importance in the global structure including indirect ties. For each researcher  $n_i$ , the importance as the leader  $CL^{<n_i>}$  and that as the cooperator  $CF^{<n_i>}$  are obtained in a manner analogous to the HITS algorithm (Kleinberg, 1998), which is shown below.  $a_{i,j}$  represents the value in cell (i,j) of the adjacency matrix of the network where the ties are oriented from secondary authors to the first author for each paper.

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

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

Iterate (G,k)

G: a collection of g researchers

k: a natural number

Let z denote the vector  $(1,1,1,...,1) \in \mathbb{R}^{g}$ .

Set CL(0):=z.

- Set CF(0):=z.
- For i=1,2,...,k

Apply the (1) operation to (CL(i-1), CF(i-1)),

obtaining new *CL*-weights  $CL^{new}(i)$ .

Apply the (2) operation to  $(CL^{new}(i), CF(i-1))$ ,

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., 2007)

We examined which index of the researcher's performance has a strong correlation with productivity of himself/his collaborators in the subsequent period. First we calculated the correlation coefficients between the indices. As well, we applied multiple regression analysis using the researcher's subsequent productivity as the response variable.

#### Results

Regarding the correlation between the performance of researchers and "their own subsequent productivity", the values of the correlation coefficient r were not more than around 0.4. On the other hand, as for the correlation with "the subsequent productivity of their collaborators", it was *CL* (i.e., an index reflecting the global structures of co-authorship networks) that showed the highest correlation. *CL* was highly correlated with both productivity measured by the normal count and that measured by the adjusted count (r is around 0.5 to 0.6).



# Figure 1. The scatter plot of expected values vs observed values of subsequent productivity.

Figure 1 represents the scatter plot of observed values against estimated values of subsequent productivity measured by the adjusted count, and Table 1 shows the summary of the results by multiple regression analysis. Standardized partial regression coefficients in this model were significant for both of two explanatory variables, that is, their own past productivity (the adjusted count) and *CL* of their collaborators (p<.001).

 Table 1. The summary of the results of multiple regression analysis.

Multiple correlation coefficient	0.470
$R^2$	0.221
$R^2$ adjusted for degrees of freedom	0.220
<i>p</i> -value	<i>p</i> <.001

#### Conclusions

This study investigated the correlation between researchers' precedent and subsequent performance for the purpose of clarifying which aspect of researchers' performance has a strong influence on their subsequent productivity. It was found that the importance as the research leader where the global structures of collaboration networks are considered. namely, CL, was relatively highly correlated with collaborators' subsequent productivity. 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 leading active cooperators". This implies the possibility that such achievements or experiences of researchers affect the subsequent productivity of their collaborators.

As the topic is very human/person oriented, we will make cooperation with social scientists and derive further interpretations in future.

# Acknowledgments

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.

### References

- Bozeman, B. & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, 33, 599-616.
- Kleinberg, J.M. (1998). Authoritative sources in a hyperlinked environment. In Proceedings of 9th Annual ACM-SIAM Symposium on Discrete Algorithms (pp. 668-677). New York: Association for Computing Machinery.
- 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* (pp. 783-791). Madrid: Spanish Research Council.