Asian Journal of Innovation and Policy (2017) 6.1:085-093
DOI: http//dx.doi.org/10.7545/ajip.2017.6.1.085

Collaborative Research Network and Scientific Productivity: The Case of Korean Statisticians and Computer Scientists

Ki-Seok Kwon*, Jin-Guk Kim**

Abstract This paper focuses on the relationship between the characteristics of network and the productivity of scientists, which is rarely examined in previous studies. Utilizing a unique dataset from the Korean Citation Index (KCI), we examine the overall characteristics of the research network (e.g. distribution of nodes, density and mean distance), and analyze whether the network centrality is related to the scientific productivity. According to the results, firstly we have found that the collaborative research network of the Korean academics in the field of statistics and computer science is a scale-free network. Secondly, these research networks show a disciplinary difference. The network of statisticians is denser than that of computer scientists. In addition, computer scientists are located in a fragmented network compared to statisticians. Thirdly, with regard to the relationship between the researchers' network position and scientific productivity, a significant relation and their disciplinary difference have been observed. In particular, the degree centrality is the strongest predictor for the scientists' productivity. Based on these findings, some policy implications are put forward.

Keywords Collaborative research, scientific network, social network analysis, Korea citation index, South Korea

I. Introduction

As science becomes bigger, collaborative research is an important way to accomplish successful scientific results. More importantly, the creativity of research results can be greatly enhanced by the participation of experts from various research areas. In a similar vein, a social network as a persuasive explanatory variable for individual researcher's activities has recently

Submitted, March 29, 2017; 1st Revised, April 25; Accepted, April 25
* Department of Public Policy, Hanbat National University, Daejeon, South Korea; kwon5861@gmail.com
** Corresponding, Department of Business Consulting, Pai Chai University, Daejeon, South Korea; jgkim947@gmail.com

emerged as an important issue not only in science policy and innovation studies, but also in other social sciences.

A decade ago, some scholars found that the network of collaborative researchers has ‘small world’ structure (Barabasi et al., 2001). In other words, the distribution of the number of links attached to node (i.e. the number of co-authorship) follows a power law in most disciplines. In Korea’s case, the collaborative research network of Korean academics is a scale-free network (i.e. Korean scholars live in a small world) (Kim et al., 2007). However, the scope of samples of the two studies is limited to the academics in natural science and engineering (i.e. mathematics and neuro-science in the former, IT and BT in the latter study).

In addition, in spite of the lack of empirical evidence, various characteristics of scientific network tend to be regarded as an influential determinant for scientists’ research performance. At least, this can be indirectly supported by a substantial body of literature reviews holding that the (internationally) co-authored papers are more academically excellent than those that are not (Bordons et al., 1996; Van Raan, 1998; Smeby and Try, 2005; Abramo et al., 2009). In order to identify the factors influencing research productivity, a number of studies have been carried out during the past few decades. Scientific productivity is related to individual characteristics (e.g. gender, age, and disciplines) as well as environmental conditions (locations, amount of funding, and institutional reputation) (e.g. Stephan and Levin, 1992; Stephan, 1996).

However, we hardly find prior research linking these two streams of studies (i.e. network and productivity). This is probably due to the fact that bibliometric data usually do not contain personal information on authors’ gender, age and disciplines. Moreover, it was not until recently that social network analysis has been applied to innovation studies (Van Der Valk et al., 2010). Therefore, this study, based on social network analysis, focuses on the relationship between various characteristics of network and scientific productivity of scientists involved in the network. Against this background, more specific research questions are as follows.

- Is there any disciplinary difference in the network of collaborative research network?
- Does the network centrality affect the scientists’ productivity?

II. Data and Methodology

In terms of data, first, we have collected a very large amount of bibliometric data of 670,000 published papers in journals monitored and supported by Korea Research Foundation, which is called KCI (Korean Citation Index) data from 2004 to 2009. This data set covers natural science, engineering, social science, and humanities fields.

In this paper, we focus on bibliometric data in two fields. Hence, statistics and computer science (6,614 and 26,510 publications respectively) are chosen, as a majority of papers (i.e. about 70%) in these two disciplines are published domestically (i.e. KCI journals) rather than internationally (i.e. SCI journals) according to NRF (2010). As the general academic language of Korean social scientists is Korean, most of their works are published in KCI journals (NRF, 2010). Therefore, the data set covers most research activities of Korean academics in the two fields.

In addition, these bibliometric data are linked to another one containing the 50,000 Korean academics' personal profiles such as gender, age, disciplines and the characteristics of the universities with which they are affiliated as well as their scientific performances. That is to say, the 1,096 statisticians' and the 4,781 computer scientists' individual characteristics (e.g. gender and age) are collected. Thereafter, we have combined the data on the academics' position in a network with the data on their academic performance. This makes our data very unique, because ISI database does not provide personal details of the authors of the publications.

After investigating the overall characteristics of the research network (e.g. the number of links and nodes, the number of components, density and mean-distance), we carried out an econometric analysis on whether the network centrality as measured by degree, betweenness, and closeness centrality is related to the scientific productivity[1]. In particular, in order to explore whether the research network of Korean academics is a scale-free network or not, we investigate the distribution of the number of links (i.e. the frequencies of the academics' co-authorships). Moreover, various characteristics of the network are investigated according to disciplines. Finally, the network measures such as individual scientists' centrality are included as an independent variable in a regression model estimating the influence of the position on the scientists' productivity.

[1] Degree centrality is a simple centrality measure that counts the number of neighbors a node has. Betweenness centrality measures the extent to which a node lies on paths between other nodes. Closeness centrality is the mean distance from a node to other nodes.

III. Results and Discussions

1. The Characteristics of Networks of Korean Scientists

1.1 Statisticians

Some characteristics of the network of 1,096 Korean statisticians are identified. The number of links totals 1,208, and the average number of links is 1.102. The density of the network is 0.002. The number of components is 323. Finally, the mean-distance between the nodes is 9.238. Figure 1 shows the distribution of the number of nodes in the network. The figure below (right) shows that the distribution follows a power's law, which means that the academics are connected very closely (i.e. located in a small world).

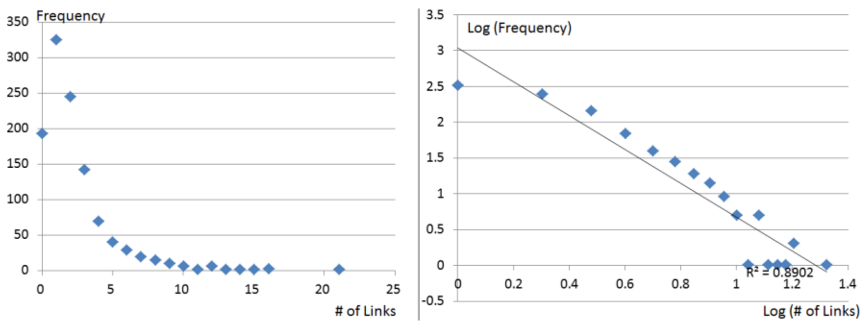

Figure 1 Distribution of the number of nodes in statisticians' research network

1.2 Computer Scientists

We have found various network properties of computer scientists. The number of links (i.e. co-authorship) is 4,720, and the average degree is 0.987, because the number of nodes (i.e. computer scientists) is 4,781. The density of the network is 0.0002; ten times lower than that of statisticians' network. The number of components is 1,674, and mean-distance between the nodes is 14.099. Figure 2 shows a distribution of the number of links in the network. According to this, the distribution follows the power's law. In other words, the Korean computer scientists work in a "small world".

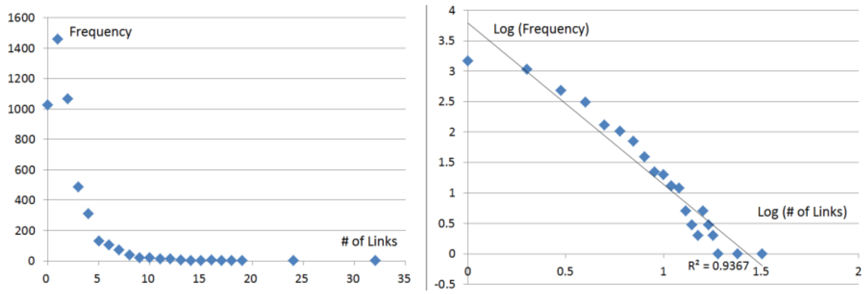

Figure 2 Distribution of the number of nodes in computer research network

2. Relationship between Centrality and Productivity

2.1 Statisticians

Table 1 presents details of the variables we measured that are included in our regression model. Firstly, 21% of academics in the field of statistics are female. Secondly, the average age of the statisticians is 48. They tend to be older than computer scientists. Thirdly, the average productivity of the scientists is measured as 6.39. Finally, the mean values of the three centralities (i.e. degree, betweenness, and closeness) are 0.002, 0.001, and 0.017, respectively.

Table 1 Descriptive statistics of statisticians

	N	Min	Max	Mean	SD
Gender	901	0	1	.79	.405
Age	899	26	72	48.04	8.621
No of Papers	975	1	55	6.39	6.368
Degree Centrality	1096	.00000	.01918	.0020130	.00209644
Betweenness Centrality	1096	.00000	.05702	.0010864	.00480972
Closeness Centrality	1096	.00000	.06271	.0170613	.01987110
No of Obs.	808				

Table 1 presents the relationship between network position and productivity. Degree and betweenness centralities are significantly related to the number of statisticians' publications. However, the directions of the relationship are opposite. In other words, degree centrality is a strong positive predictor for the productivity, whereas betweenness centrality is a strong negative predictor for the productivity. Furthermore, closeness centrality is not significantly related to the productivity. Next, the other individual characteristics are also related to the productivity of the Korean academics. Age and the square term of age are significantly related to the productivity, which means that, at a certain point in the life cycle, productivity starts to decrease. However, gender of academics is not a strong predictor for productivity in our empirical analysis.

2.2 Computer Scientists

Table 2 show a descriptive statistics of variables we adopt in our model. Firstly, 13% of computer scientists are female. Secondly, the average age of the scientists is 47. Thirdly, the average number of papers dependent variable in our model published by the scientists is 5.59. Finally, the degree, betweenness, and closeness centrality of the scientists are 0.0004, 0.0002 and 0.0067, respectively.

Table 2 Descriptive statistics of computer scientists

	N	Min	Max	Mean	SD
Gender	3926	0	1	.87	.341
Age	3903	26	90	46.93	8.508
No of Papers	3847	1	77	5.59	5.555
Degree Centrality	4782	.0000	.006693	.000413	.0004600
Betweenness Centrality	4782	.0000	.021500	.000243	.0014283
Closeness Centrality	4782	.0000	.030552	.006795	.0100219
No of Observation	3151				

2.3 Regression Results

Table 3 shows the relationship between academics' positions in the network (i.e. centrality) and academic productivity. Degree and closeness centralities are significantly related to the productivity, whereas betweenness centrality is weakly related to the productivity. In terms of the direction of the relationship, degree centrality is positively influencing productivity, which means that collaboration with a larger number of co-authors encourages publication of papers. However, somewhat contrary to our intuition, closeness centrality is negatively related to the productivity. Other control variables such as age and the square term of age are significant predictors for the productivity. The direction of relationship (i.e. the signs of coefficients) confirms the 'inverse-U' relationship between age and scientific productivity, which is in the same vein with other previous studies. However, in our study, gender does not turn out to be a significant predictor.

Table 3 Estimation of scientific productivity of statisticians and computer scientists

Productivity Dep. Var.	Statisticians	Computer Scientists
Gender	-.077 (.081)	.055 (.051)
Age	.017 (.005)***	.153 (.019)***
Age^2	-.0001 (.00001)**	-.002 (.000)***
Degree Centrality	115.677 (15.397)***	523.713 (35.123)***
Betweenness Centrality	-15.417 (4.528)**	-10.246 (8.882)
Closeness Centrality	.523 (1.859)	-9.997 (1.806)***
Constant	.980 (.213)***	-2.107 (0.445)***
Log p-likelihood	-2311.1134	-8465.6604
Wald chi^2 (d. of f.)	97.53 (6) ***	344.33 (6)***
No. of Observation	808	3151

Note: * < .05, ** < .01, *** < .001

1. Age2 is the square term of age
2. According to Dispersion and Vuong test, negative binomial model is the fittest model.
3. In order to prevent excessive multi-collinearity between the explanatory variables, the variables with a high VIF are excluded. Furthermore, considering the possibility of a heteroscedasticity problem, robust standard errors are calculated.
4. Cases with missing value are excluded in the econometric model.

IV. Conclusion

Several empirical findings can be summarized as follows. Firstly, the distribution of the number of links attached to nodes (i.e. the number of co-authorship) follows a power law in most disciplines. In other words, the collaborative research network of the Korean academics in the field of statistics and computer science is a scale-free network (i.e. Korean scholars live in a small world). This corroborates the findings from previous studies (e.g. Barabasi et al., 2001; Kim et al., 2007). Secondly, these research networks show a disciplinary difference. The network of statisticians is denser than that of computer scientists. In addition, computer scientists are located in a fragmented network compared to statisticians.

Thirdly, we have found a significant positive relationship between the network position and scientific productivity. In particular, degree centrality is the strongest predictor for the scientists' productivity, while the betweenness centrality is significantly and negatively related to the statisticians. In contrast, closeness centrality shows a strong relation to computer scientists. Possibly, the disciplines of academics affect the relationship. For example, hard science such as statistics tends to increase productivity when they are not located ‘between’ scientists. This is the opposite in the case of computer scientists. Additionally, various personal characteristics such as age are significantly related to the academics’ research productivity, while the gender is a significant predictor only for computer scientists.

Based on these findings, we put forward some policy implications. Firstly and most importantly, science policy needs to be changed as the research network is proven to be a ‘small world’. That is to say, scientific resources should be allocated not only on the basis of individual academics’ excellence, but also with regards to the network they are linked to. We may identify several key journals, organizations and individuals in the network. These actors in these multi-levels require special attention to invigorating the research targeted by the government. Secondly, various network indicators measuring the position in the collaborative research need to be generated and monitored. For example, network centrality indicators can be included in research proposals as criteria for funding. Thirdly, we have found some inter-disciplinary areas with a weaker collaborative research, which requires a more intensive encouragement of the government.

In future research, we ought to expand the scope of the data set horizontally and longitudinally. In terms of horizontal perspective, Korean scholars in the field of natural science and engineering prefer to publish in international journals, but the data set based on domestic journals in these fields do not cover all the papers produced by Korean scholars. Moreover, social sciences (e.g. economics and management studies) need to be included in this kind of research. In terms of longitudinal perspective, the accumulation of KCI dataset will make time series data available. Moreover, some qualitative data based on interview or documents can provide a richer explanation of the academics’ behaviors with regards to publication and research collaboration. The relationship between network positions and the quality of research is a promising research topic. To work on this, future studies will need to collect researchers’ citation numbers.

References

Abramo, G., D´Angelo, C.A. and Pugini, F. (2007) The measurement of Italian universities' research productivity by a non-parametric-bibliometric methodology, Scientometrics, 76(2), 225-244.

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002) Evolution of the social network of scientific collaborations, Physica A: Statistical Mechanics and its Applications, 311(3-4), 590-614.

Bordons, M., Go´mez, I., Ferna´ndez, M.T., Zulueta, M.A. and Me´ndez, A. (1996) Local, domestic and international scientific collaboration in biomedical research, Scientometrics, 37(2), 279-295.

Kim, Y.H., Yoon, J.R., Cho, H. and Kim, Y.J. (2007) Structure of collaboration network among Korean scientists -'small world' and position effect (in Korean), Korean Journal of Sociology, 41(1), 68-103.

Smeby, J. and Try, S. (2005) Departmental contexts and faculty research activity in Norway, Research in Higher Education, 46(6), 593-619.

Stephan, P.E. (1996) The economics of science, Journal of Economic Literature, 36, 1199-1235.

Stephan, P.E. and Levin S.G. (1992) Striking the Mother Lode: The Importance of Age, Place, and Time, New York and Oxford: Oxford University Press.

Van Raan, A.F.J. (1998) The influence of international collaboration of the impact of the research results, Scientometrics, 42(3), 423-428.