An Analysis of the Collaborative Network Mechanism Based on the Dynamic Network

Hyejin Lee, Seok-Hyoung Lee, Kwang-Young Kim
Dept. of Information Convergence Research
Korea Institute of S&T Information
Daejeon, Rep. of Korea
E-mail: hyejin@kisti.re.kr, skyi@kisti.re.kr, glorykim@kisti.re.kr

***Abstract*—This ongoing research aims to explore factors contributing to collaboration in various time intervals using the Stochastic Actor-Oriented Model (SAOM), a dynamic network analysis method. This model can be used to measure correlations between the consequences of an individual's choice and the network structure in a time series, thereby permitting simultaneous observations of a network at the micro-level and macro-level. To accomplish the research purpose, three mechanisms related to collaboration were established as hypotheses based on the literature review: reciprocity, hierarchy, and similarity. These mechanisms were combined cumulatively to determine whether they were significant for collaboration. Accordingly, the results of a pilot experiment showed that reciprocity and similarity did not have significant effects individually across time intervals, but their explanatory power about collaborative relations increased when these two variables were used in combination.**

Keywords-Social Network Analysist; Dynamic Network; Co-author Network; Stochastic Actor-Oriented Model

I. INTRODUCTION

In the context of this study, collaboration is interaction between researchers and a long-standing scientific practice for advances associated with a discipline. Interaction between researchers is recognized as a part of the research process in which the research community effectively communicates and shares information about various investigations [1].

Researchers benefit from the following effects of collaboration. During the process of exchanging ideas and tacit knowledge about a research topic and reaching an agreement regarding anticipated findings and analytical methods, they can share research equipment or engage in formal/informal communications. In addition, continuous collaboration in research activities permits access to a wider range of information sources, new knowledge, and an increased pace for dissemination of research findings based on the various backgrounds of participating researchers. Further, collaboration reduces research expenses through the shared use of large-scale equipment or facilities and helps expand research activity opportunities [1] [2].

Researcher collaboration carries it own important meaning, but it also has many intrinsic values from a metrology-based data analysis perspective. Not only can researcher collaboration produce material for data analysis, but the meaning of collaboration can also be interpreted in many ways. The most frequently used metrology-based collaboration analysis method is the co-author network analysis method, which refers to social networks. It is applicable to researcher collaboration because the data can measure the scope of a collaboration easily by utilizing co-authors' unique features; additionally, it is objective, demonstrable, and stable over time [3].

Scientific findings are the product of massive collaborations; in other words, papers having multiple co-authors could be produced in massive numbers [4]. The latest research trends in "big science" and "data science" have allowed the number of collaboration-based, co-authored papers to increase explosively. For instance, the Thomson Reuters's Web of Science (WoS) indicates that—as they relate to physics—120 papers with more than 1,000 co-authors and 44 papers with more than 3,000 co-authors were published in 2011[5].

With the consistent growth in publication of co-authored papers, the co-author network continues to change dynamically, repeatedly evolving and differentiating into various forms. For this reason, a scientific network is often categorized as a dynamic network [6]. Accordingly, an analysis of the co-author network from the perspective of a dynamic network is needed [7]. The Stochastic Actor-Oriented Model (SAOM), one of the dynamic network analysis methods, is made according to processes of individual choice [8]. Also, the individual choice affects the overall collaborative network structure. This is why we adopted the SAOM in this study.

This research defines a co-author network as a dynamically changing network and tracks network changes over time based on the longitudinal analysis of a co-author network. The aim of this research is to identify empirically a relationship between individual factors and the network structure for significant factors used by individual

 ISBN: 978-1-61208-578-4

researchers to select their co-authors and establish a collaboration network.

II. STOCHASTIC ACTOR-ORIENTED MODEL(SAOM)

To verify the research hypotheses, the SAOM—a social network analysis method—was adopted. Social network analysis focuses on a social entity—a relationship between an actor and entity—to not only identify a relationship between two individuals but also expand and explain a social relationship between individual entities. The SAOM is a computational model focused on the actor(s). It initiates a complex system based on a small virtual world consisting of many interacting actors. This virtual world is made up of the actor, the system in which the actor acts and interacts, and the external environment that affects the system. This model can be used to research the interactions between various actors and better reflect and analyze the actual world in which the characteristics or choices of an actor (an individual) could affect another individual. It is especially effective in research explaining how micro-level interaction, such as a personal relationship, affects macro-level interaction.

Research using this model in information science is still in an early stage, and no research reflecting various characteristics of the actor has been conducted to date. Kronegger et al. [9] conducted research on the community of Slovenian scientists using the SAOM, examining changes in the collaboration network structure for each time interval in four scientific fields. Ferligoj et al. [10] also applied the SAOM to Slovenian scientists in seven disciplines and explained factors changing the domestic and overseas collaboration network based on the cumulative performance expectation mechanism. Zinilli [8] used this model and investigated researcher collaboration in the Projects of National Interest (PRIN) performed in Italy. The research traced changes in network links based on four academic disciplines and used the h-index as an independent variable for choosing a partner.

III. RESEARCH DESIGN

To achieve the research purposes, the following four steps describe the research process: 1) establish the hypotheses, 2) select data, 3) extract data, and 4) analyze and verify.

A. *Research Process*

First, the hypotheses address factors for collaboration based on a literature review of existing research papers. By grouping research findings suggested by the papers, the researchers defined the top mechanisms for the formation of collaborations.

Second, data were selected to choose research subjects in the nanoscience field and set time intervals to measure networks. Subjects included both key researchers and co-authors in the field. Time intervals were set to range from 2001—the year when the National Nanotechnology Initiative was established—to 2015. To select specific fields of nanoscience and measured time intervals, researchers focused the analysis on nanoscience-related policy and used WoS and InCites databases to select data.

Third, the researchers are extracted data to generate attribute information and co-author relationship information about the research subjects. Accordingly, foundational data for data extraction came from WoS. Attribute information was based on individual characteristics derived from the hypotheses as collaboration factors, and co-author relationship information was based on the co-author relationships for published papers.

Fourth, this research incorporated a longitudinal analysis, which applied the SAOM to the extracted data and verified the hypotheses. An analysis of the overall co-author network structure and a pattern analysis of each measured time interval were performed. An analysis of factors pertaining to individual or network structure and affecting the formation of a co-author relationship was also conducted. Though NetMiner 4.0—which specializes in analyzing and visualizing overall structure and pattern—was used for the former, RSiena was used for the latter, as it specializes in longitudinal analyses of networks.

B. *Research Hypothesis*

First, research collaboration establishes a network based on interrelationships and trust.

Maglaughlin and Sonnenwald [11] regard trust in researchers' research capabilities, values, and academic knowledge as important factors in choosing collaboration partners. Hara et al. [12] interpreted the reason researchers choose partners based on prior collaborative relationships as the decreased chances for failure with research outcomes. Intimacy with collaboration partners could also enhance collaboration efficiency, and the collaboration process is easier when a researcher partners with others he or she personally knows [13].

H1: *Researchers' academic friendship and trust (whether they co-authored in the previous year or research career) would affect the formation of an academic collaborative network.*

Second, researchers tend to forge collaborative relationships with partners who are characterized as more esteemed. As actors whose cumulative performance is expected to be high have more connections in a network structure [14], nodes are concentrated on the nodes with high levels of centrality. Abbasi et al. [15] demonstrated that degree centrality, eigenvector centrality, average strength, and network efficiency were independent variables affecting the g-index, which indicates an individual researcher's research performance.

 ISBN: 978-1-61208-578-4

H2: Researcher hierarchy (organizational reputation) would affect the formation of an academic collaborative network.

Third, the similarity mechanism is known as homophily; it means that similar characteristics and qualities between actors promote intimacy and bonds, with a high likelihood for the formation of a network structure. Various research papers have cited results for homophily in researcher collaboration. Maglaughlin and Sonnenwald [11] viewed the research topic as an important factor when researchers collaborated with other researchers. Kronegger et al. [9] stated that a network was formed depending on similarities between individual researchers' organizations.

H3: Similarities between researchers (organizations) would affect the formation of an academic, collaborative network..

IV. DATA ANALYSIS AND PILOT TEST RESULTS

In order to achieve the purpose of this research, data analysis was carried out as follows and the results of the research were derived.

A. Data Analysis Process

To extract authors and co-authors in nanobiotechnology fields as data, papers in the fields were collected by subject area through WoS and its methodology. Initially, NT papers in Korea from 2001 to 2015 were used in the first subject search scope in accordance with the WoS search rule, and the scope of subjects were reduced to BT based on the search results. As a result, the total number of collected papers was 1,704. From the collected papers, five authors who consistently published papers within the designated time interval(s) and who also had high h-indexes were selected; then, their co-authors were extracted.

Table I shows the results from the collection of papers via WoS for the five authors (applying time intervals). Data extracted included numbers of co-authors and co-author relationships. The total number of identified authors was 1,154, and the number of authors identified in the last time interval was 644.

As a next step, the researchers are extracted co-authors for Time Interval 3 (2011 – 2016) based on the publication year and established a collaborative network. Based on the same authors in Time Interval 3, collaborative networks were established in the same manner for Time Interval 1 (2001 – 2005) and Time Interval 2 (2006 – 2010). Because the population of each network should remain the same to identify objectively the factors that cause a collaborative network to look a certain way in a given time interval, the number of nodes in Time Intervals 1 to 3 were made consistent at 644; then, experimental data were generated.

TABLE I. STATISTICS OF THE COLLECTION OF PAPERS

Number of papers	Number of authors in papers	Average number of authors per paper	Number of identified authors
597	3,224	5.49	1,154

B. Results of Pilot Test Analysis

1) Network Structure Analysis for Each Time Interval

The network structure analysis results showed that the average degree or the average number of edges, which represents the number of collaborative relationships per researcher, increased as time passed.

Density gets closer to 1 when all researchers are linked to one another, but in this research, it decreased over time, most likely because of the rapid growth of edges.

The level of fragmentation rose gradually over time. Decreasing density means that, even though subgroups were established, their internal density was low and there were few excessive disconnects between subgroups.

Clustering represents a subgroup's level of separation, and fragmented subgroups seem to be the result of segmentation of the subject area, differentiation of the key researcher, and convergence with other fields.

TABLE II. RESULTS OF NETWORK STRUCTURE FOR EACH TIME INTERVAL

Observance Time Interval	Degree	Mean of Degree	Density	Clustering
2001-2005	52	1.529	0.093	0.781
2006-2010	538	3.183	0.038	0.84
2011-2015	3,591	5.567	0.017	0.861

Figure 1 shows the co-authors' network by time interval. The network structure for each time interval is composed of three components.

2) Network Structure Analysis for Each Time Interval

This section longitudinally analyzes factors that influence collaborations between nanobiotechnology researchers in Korea. To examine the effect of each factor, this research will identify collaboration factors for individual independent variables as the first step and combine them phase by phase to examine the extent of increased explanatory power from each combination of variables based on changes of the random variable as the second step.

Therefore, network data 1, 2, and 3 (the N by N adjacency matrix) were used as dependent variables for each of the three intervals and dependent variables' changes in accordance with 4 independent variables in each time interval were confirmed using RSiena.

In the pilot test to analyze the collaboration mechanism involving network changes using RSiena, whether a network establishment factor was significant in the overall interval was examined. The overall maximum convergence ratio is an index for identifying whether a network establishment factor is significant in the overall interval; when the ratio is 0.25 or less, the factor is considered significant. Two out of five researchers and their co-authors were analyzed, and the results showed that the combination of two independent

 ISBN: 978-1-61208-578-4

variables—“whether researchers co-authored in the previous year” and “whether researchers belonged to the same organization”—were interpreted to have a significant effect on the dependent variables, as described in the Table Ⅲ.

V. FUTURE RESEARCH DESCRIPTION AND EXPECTED OUTCOMES

Research is now in progress to analyze significant factors regarding the five researchers’ selections of co-authors for each time interval and the structural characteristics affecting their individual choices. This research holds academic significance for its analysis of a co-author network phenomenon from a new angle based on application of the SAOM, a social network theory, to information science. More notably, the method that combined a dynamic network structure analysis and the probability theory is meaningful in expanding the traditional static network-based macro-level co-author analysis methodology and connecting the macro-level (structure) and the micro-level (actor) in a dynamic way for analysis. Moreover, the results of the analysis are worth using for service planning (e.g., information service recommendations or predictions).

REFERENCES

[1] S. Lee, and B. Bozeman, “The impact of research collaboration on scientific productivity,” Social studies of science, vol. 35, number5, 2005, pp. 673-702.

[2] G. Abramo, C. A. D’Angelo, and F. Di Costa, “Research collaboration and productivity: is there correlation?” Higher Education, vol. 57, no.2, 2009, pp. 155-171.

[3] DH. Lee, IW. Seo, HC. Choe, and HD. Kim, “Collaboration network patterns and research performance: the case of Korean public research institutions,” Scientomerics, vol. 91, 2012, pp.925-942.

[4] G. Melin, “Pragmatism and self-organization Research collaboration on the individual level,” Research Policy, vol. 29, 2000, pp.31-40.

[5] J. Adams, “Collaborations: The rise of research networks,” Nature, vol. 490, no. 7420, 2012, pp.335-336.

[6] J.S. Katz, and B.R. Martin, “What is research collaboration?” Research Policy, vol. 26, no.1, 1997, pp.1-18.

[7] A. Perianes-Rodriguez, L. Waltman, and N. J. van Eck, “Constructing bibliometric networks: A comparison between full and fractional counting,” Journal of Informetrics, vol. 10, no. 4, 2016, pp.1178-1195.

[8] A. Zinilli, “Competitive project funding and dynamic complex networks: evidence from Projects of National Interest(PRIN),” Scientometrics, vol. 108, 2016, pp. 633-652.

[9] L. Kronegger, F. Mali, A. Ferligoj, and P. Doreian, “Collaboration structures in Slovenian scientific communities,” Scientometircs, vol. 90, 2012, pp.631-647.

[10] A. Ferligoj, L. Kronegger, F. Mali, T. A. Snijders, and P. Doreian, “Scientific collaboration dynamics in a national scientific system,” Scientometrics, vol. 104, 2015, pp.985-1012.

[11] K. L. Maglaughlin, and D.H. Sonnenwald, "Factors that impact interdisciplinary natural science research collaboration in academia," 10th International Conference of the International Society for Scientometrics and Informetrics, Jul. 2005, pp. 499-508. Stockholm.

[12] N. Hara, P. Solomon, S. L. Kim, and D. H. Sonnenwald, “An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration,” Journal of the American Society for Information science and Technology, vol. 54, no. 10, 2003, pp.952-965.

[13] D. Stokols, S. Misra, R. P. Moser, K. L. Hall, and B. K. Taylor, “The ecology of team science: Understanding contextual influences on transdisciplinary collaboration,” American Journal of Preventive Medicine, vol. 35, no. 2, 2008, pp.S96-S115.

[14] M. Ruef, H.E. Aldrich, and N.M. Carter, “The Structure of Founding Teams: Homophily, Strong Ties, and Isolation Among U.S. Entrepreneurs,” American Sociological Review, vol. 68, no. 2, 2003, pp. 195-222.

[15] A. Abbasi, J. Altmann, and L. Hossain, “Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regressionanalysis of performance measures and social network analysis measures,” Journal of Informetrics, vol. 5, no. 4, 2011, pp. 594‐607.

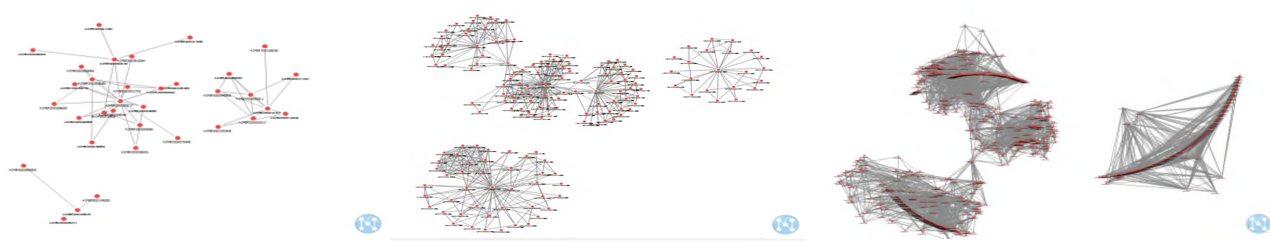

Figure 1. Results of network structure for each time interval

TABLE III. RESULTS OF SAMPLE TEST

Variable	Measurement Method	Researcher A t-conv.max value (N=234)	Researcher B t-conv.max value (N=154)
Researcher friendship	Whether researchers co-authored in the previous year	0.6888**	0.1161*
Researcher career	Researcher’s research career in the measured time interval	1.7564	1.8215
Organization reputation	The organization’s contributions to the field (number of papers published)	0.9429**	0.9920**
The same organization	Whether researchers belonged to the same organization	0.8954**	0.1108*

 ISBN: 978-1-61208-578-4

Researcher friendship + the same organization	Whether researchers co-authored in the previous year + whether researchers belonged to the same organization	0.1953*	0.1002*

 ISBN: 978-1-61208-578-4