ICIC Express Letters
Volume **12**, Number **1**, January **2018**
 ISSN 1881-803X
pp. 71–78

RESEARCH TRENDS IN SOCIAL NETWORK ANALYSIS USING TOPIC MODELING AND NETWORK ANALYSIS

Yubin Kim[1] and Nam-Wook Cho[2,*]

[1]Department of Data Science
Graduate School
[2]Department of Industrial and Information Systems Engineering
Seoul National University of Science and Technology
232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea
[*]Corresponding author: nwcho@seoultech.ac.kr

Received July 2017; accepted October 2017

Abstract. *This study investigates the research trends in social network analysis (SNA) by examining related scholarly articles. Firstly, topic modeling is applied to a corpus composed of the title, abstract, keywords, and author information of 58,122 articles obtained from the Scopus database. Twenty topics and their ascending or descending trends are identified. Secondly, to explore the multidisciplinary nature of SNA, an academic field network is constructed based on co-authorship and affiliation information. Results show that physics, biology, and education have taken central position in the network. It was also interesting to see that sociology and psychology are less influential. The results of this study will be helpful for future researchers exploring SNA research topics and thereby improve the diversity of SNA research.*
Keywords: Research trend, Topic modeling, Social network analysis

1. **Introduction.** Social network analysis (SNA) is the process of investigating social structures by using networks and graph theory [1,2]. SNA is concerned with relationships and flows between nodes or actors that represent people, organizations, objects, and so on. In SNA, links show relationships or flows between the nodes. Unlike traditional social science studies, network analysis focuses on the relations among actors – not individual actors and their attributes [3]. Recently, due to the advancement of information technologies and the emergence of complex systems, SNA has attracted a fair amount of attention in the fields of social sciences, engineering, and natural sciences. Figure 1 shows the increasing number of research articles on SNA.

The increasing interest in SNA has revealed the multidisciplinary nature of SNA [4]. Although SNA originated in the field of sociology, numerous study topics in various academic fields are identified in SNA research. Thus, it is meaningful to investigate the research trends in SNA. Although a number of studies have applied SNA or network analysis to investigating research trends in various fields [5-8], it is worth investigating recent research trends in SNA.

This study analyzes research trends in SNA in 2000-2015. Topic modeling and network analysis are applied to a corpus composed of the title, abstract, keywords, and author information of 58,122 articles. The rest of this paper is organized as follows. Section 2 provides a research framework composed of topic modeling and network analysis. Section 3 explains the results of the topic model and network analysis. Finally, Section 4 discusses the benefits and limitations of our research.

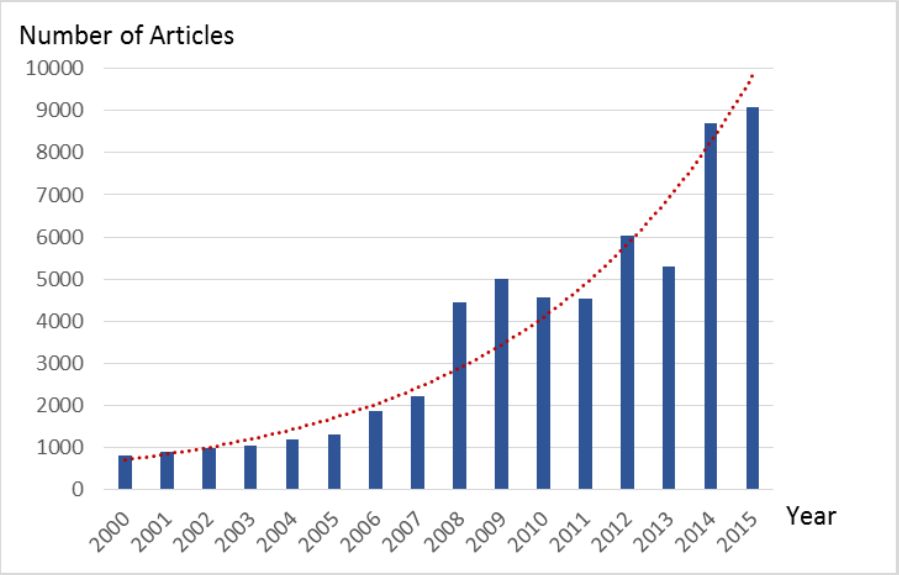

FIGURE 1. Number of articles on SNA

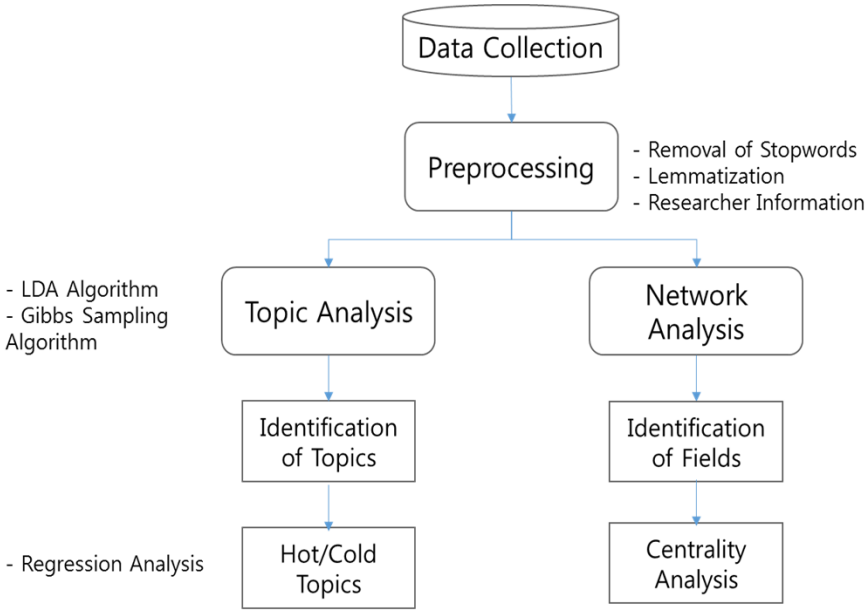

FIGURE 2. Research framework

2. Methods.

2.1. **Research framework.** Figure 2 shows the overall research framework. Firstly, research articles related to "social network analysis" or "network analysis" published between 2000 and 2015 were searched in the Scopus database. As a result, 58,122 journal articles were collected. In preprocessing, non-English articles and articles with missing information were removed from the collected articles, which resulted in 56,951 articles. Then, topic modeling and network analysis were conducted.

2.2. **Topic modeling.** Topic modeling is a type of statistical model for discovering the abstract topics that occur in a collection of documents, called a corpus. Among various topic modeling algorithms, latent Dirichlet allocation (LDA) is one of the most commonly used algorithms [9-11] and is also used in this study. Prior to the application of LDA, preprocessing is conducted, which involves tokenization of text data, removal of stop words, and lemmatization. After the preprocessing, the LDA algorithm is applied to the

preprocessed corpus. In this research "tm" package and "topic models" package in R were used.

2.3. **Network analysis.** In this research, the objective of network analysis is to explore the multidisciplinary nature of SNA research. As network analysis focuses on the relations among actors (node), and not individual actors and their attributes, the definition of actors and relations is important. In this study, the node is defined as a field of research and link is defined as co-authorship. As shown in Figure 3, the majority of research articles have multiple authors whose affiliations are often different from each other. For example, the "doc1" article has two co-authors: one's affiliation is "education" and the other's affiliation is "history". Therefore, a link between "education" and "history" can be generated. The original matrix in the left of Figure 3, (document * affiliation) relationships can easily be transformed into the (affiliation * affiliation) matrix.

	doc1	doc2	doc3
physics	0	0	3
education	1	0	2
history	1	2	0
pathology	0	0	1

	physics	education	history	pathology
physics	3	5	0	4
education	5	3	1	3
history	0	1	0	0
pathology	4	3	0	1

FIGURE 3. Co-authorship matrix

3. Results.

3.1. **Topic modeling.** Table 1 summarizes the results of topics identified by topic modeling. Twenty topics and eight topic words for each topic are identified. Table 2 further categorizes the topics into six categories. As shown in Table 2, the majority of topics belong to the natural sciences and technologies. Even though SNA originated in the field of social sciences, the topics of SNA have moved toward the natural sciences and technologies, which will be further analyzed in the network analysis in Section 3.2.

Table 3 presents "hot" and "cold" topics and their statistical significance. Hot topics include online community, biology, interdisciplinary research, detection system and methodology, education, network methodology, developmental psychology, and population pattern, whereas electric current flow, information theory, and applied mathematics are identified as cold topics. The topic analysis shows that topics related to IT and network methodology have emerged, while theoretical research has slowed down. In particular, online community and biology are identified as the hottest topics. It is also interesting to note that interdisciplinary research is identified as a hot topic, which exhibits the multidisciplinary characteristics of SNA research.

3.2. **Network analysis.** This section presents the results of network analysis. The network is constructed based on authors' affiliations and their co-authorship as explained in Section 2. As shown in Table 4, the physics, biology, and education fields take central position in terms of degree and betweenness centrality. Surgery and food fields have relatively higher betweenness centrality than degree centrality, which implies their significant mediation roles in SNA research.

Figure 6 provides a visual representation of the entire network. Four clusters are identified in the graph: physics, biology, education, and psychology. Two clusters are related

TABLE 1. Topic modeling results

Topic	Topic Name	Topic Words
topic 1	Population Pattern	population, pattern, group, contact, individual, diver, genetic
topic 2	Applied physics	electric, field, conduct, property, experiment, temperature, theoretic, layer
topic 3	Cultural politics	culture, politics, article, intern, state, identity, migrant, immigration
topic 4	Education	group, learn, student, participation, education, program, professional
topic 5	Power system	distribution, cost, optimization, generate, control, energy, improve, efficient
topic 6	Regional community	community, capital, policy, local, resource, access, urban, community
topic 7	Online community	information, medium, online, communication, mobile, site, web, Facebook
topic 8	Network methodology	structure, connect, node, link, graph, cluster, complex, degree
topic 9	Medical sociology	health, care, woman, age, family, life, child, old
topic 10	Regional development	change, region, flow, spatial, area, duration, environment, increase
topic 11	Applied mathematics	measure, predict, level, factor, compare, rate, correlation, variable
topic 12	Developmental psychology	relationship, influence, behavior, individual, positive, tie, friend, peer
topic 13	Detection system and methodology	detect, applicable, search, feature, recommend, tool, compute, real
topic 14	Technology management	knowledge, manage, market, organization, project, innovation, product, technology
topic 15	Electric current flow	current, voltage, frequency, line, phase, control, mode, signal
topic 16	Interdisciplinary research	research, author, collaboration, field, review, science, publish, article
topic 17	Process theory	process, understand, theory, concept, integral, discuss, framework, context
topic 18	Biology	gene, express, protein, cell, identify, pathway, cancer, regular
topic 19	Information theory	compute, dynamic, state, order, linear, parameter, point, matrix
topic 20	Medicine and pathology	patient, risk, treatment, hive, association, drug, intervention, man

to the social sciences and the other two clusters are related to the natural sciences. Education and psychology represent the clusters in the social sciences, while physics and biology fields take a central position in the SNA research. It is interesting to observe that sociology could not take a central position in the SNA research.

4. **Conclusion.** In this study, research trends in SNA have been analyzed by topic modeling and network analysis. The results show that the majority of topics in SNA are related to the natural sciences and technologies rather than the social sciences. Through

Table 2. Categories of topics

Category	Topic
Humanity & Culture	Population pattern; Cultural politics; Education; Regional community; Regional development; Developmental psychology
Biology	Medical sociology; Biology; Medicine and pathology
Technology	Online community; Technology management
Methodology	Network methodology; Detection system and methodology; Process theory; Information theory
Energy	Power system; Electric current flow
Science	Applied physics; Applied mathematics; Interdisciplinary research

Table 3. Hot and cold topics

Topic	Name	p-value	Hot/Cold
1	Population pattern	0.000568**	Hot
2	Applied physics	0.584079	–
3	Cultural politics	0.82194	–
4	Education	0.003257**	Hot
5	Power system	0.106885	–
6	Regional community	0.285536	–
7	Online community	5.88E-07**	Hot
8	Network methodology	0.000691**	Hot
9	Medical sociology	0.262996	–
10	Regional development	0.572487	–
11	Applied mathematics	1.89E-06**	Cold
12	Developmental psychology	0.032471*	Hot
13	Detection system and methodology	0.00035**	Hot
14	Technology management	0.057714	–
15	Electric current flow	0.000212**	Cold
16	Interdisciplinary research	1.32E-07**	Hot
17	Process theory	0.999285	–
18	Biology	2.06E-07**	Hot
19	Information theory	0.000421**	Cold
20	Medicine and pathology	0.739142	–

**: $p < 0.01$ *: $p < 0.05$

Table 4. Centrality analysis

	Degree Centrality		Betweenness Centrality	
Rank	Field	Value	Field	Value
1	Physics	0.756	Physics	0.097
2	Biology	0.733	Biology	0.065
3	Education	0.700	Education	0.055
4	Life	0.656	Surgery	0.042
5	Psychology	0.600	Food	0.041
6	Epidemiology	0.567	Life	0.038
7	Economics	0.556	Economics	0.038
8	Food	0.522	Psychology	0.030
9	Sociology	0.522	Epidemiology	0.030
10	Statistics/Surgery	0.511	Sociology/Pharmacy	0.025

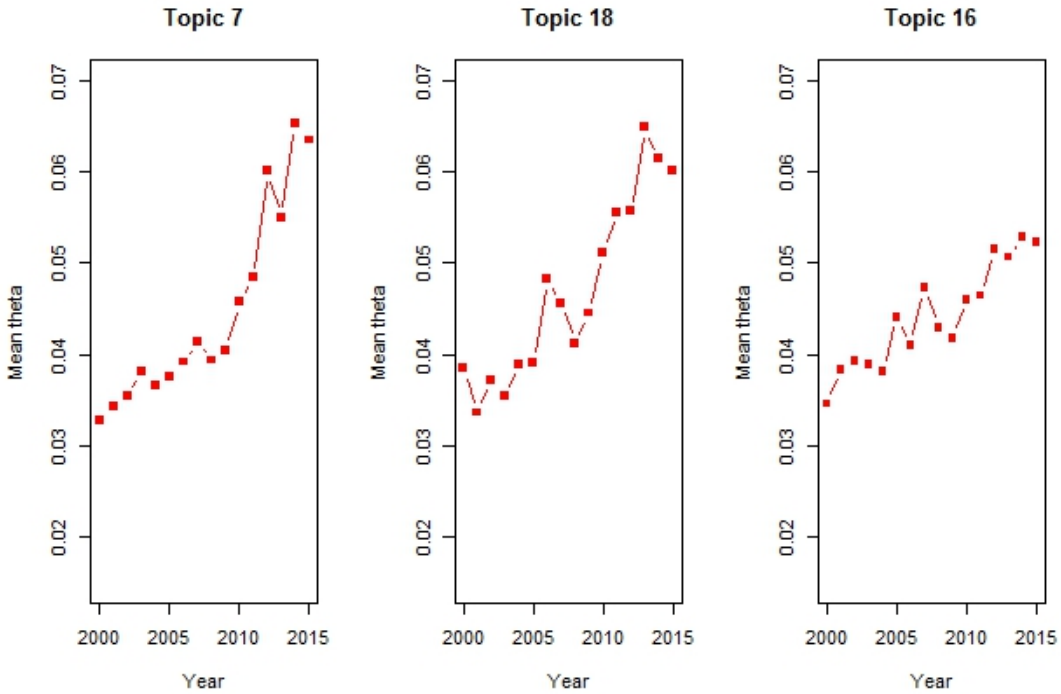

FIGURE 4. Hot topics

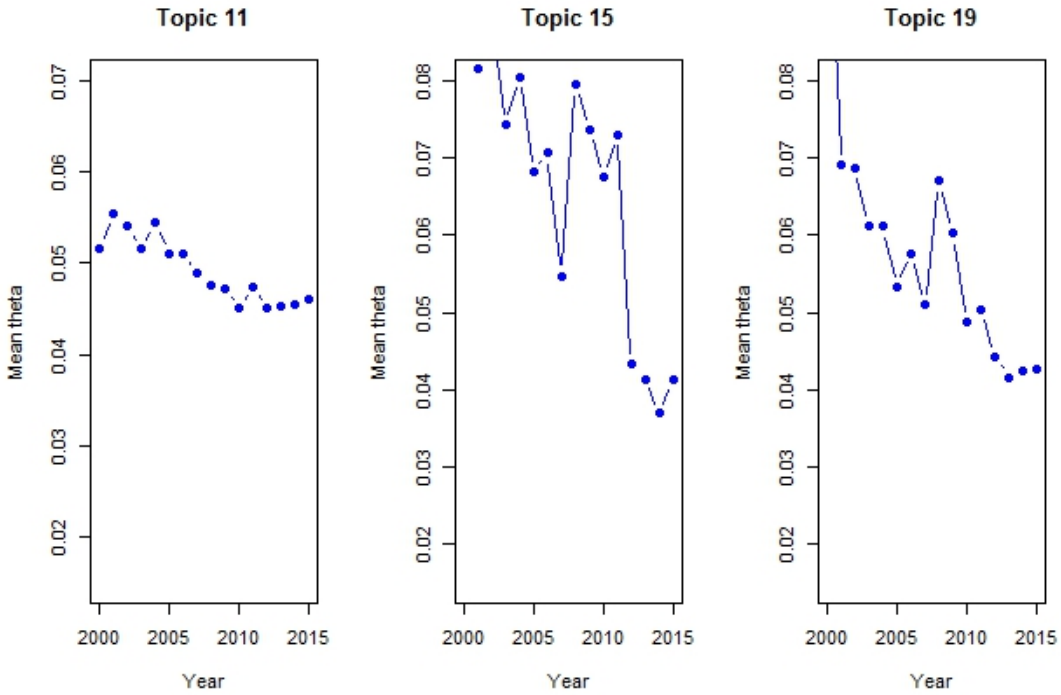

FIGURE 5. Cold topics

the analysis of hot and cold topics, we find that research trends are moving toward applications of SNA, rather than theories. Network analysis shows that the fields of physics, biology, and education take central position in the affiliation network. Four clusters of physics, biology, education, and psychology are also identified in the network. It is interesting to note that the centrality of sociology is lower than that of physics, biology, education, life, psychology, epidemiology, economics and food, which demonstrates the

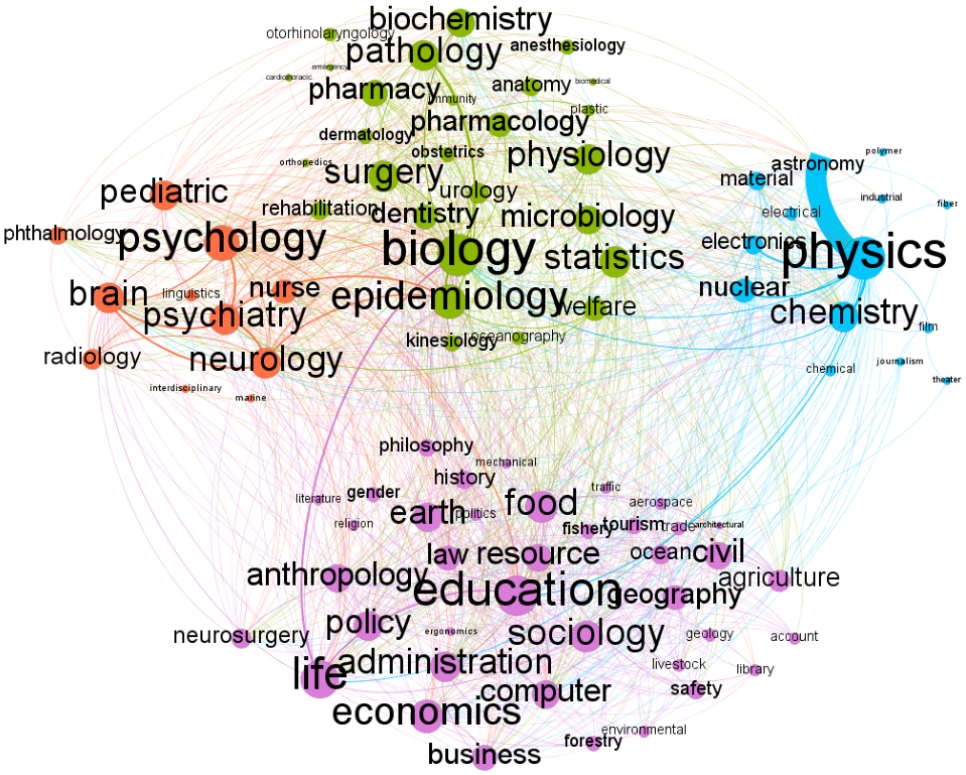

FIGURE 6. Author affiliation network of SNA research

relatively weak multidisciplinary roles of sociology. The results of this study will help future researchers explore SNA research topics and contribute to the diversity of SNA research.

In spite of its contribution, this paper only presents a static approach to the research trend analysis of SNA. Further, network analysis is applied to the research fields, which are identified from authors' affiliation information. A study on dynamic aspects of network structures would be a suitable future research topic. In addition, as authors' information also contains nationality and affiliation, analyzing collaboration networks based on authors' nationality and affiliation could be another future research topic.

REFERENCES

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.

[2] E. Otte and R. Ronald, Social network analysis: A powerful strategy, also for the information sciences, *Journal of Information Science*, vol.28, no.6, pp.441-453, 2002.

[3] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods*, Riverside, CA, http://faculty.ucr.edu/∼hanneman/, 2015.

[4] C. Thovex and T. Francky, Static, dynamic and semantic dimensions: Towards a multidisciplinary approach of social networks analysis, *International Conference on Knowledge Science, Engineering and Management*, 2010.

[5] J. Kho, K. Cho and Y. Cho, A study on recent research trend in management of technology using keywords network analysis, *Journal of Intelligence and Information Systems*, vol.19, no.2, pp.101-123, 2013.

[6] D. R. Kim, K. S. Kwon and S. B. Jeong, Knowledge structure analysis on business administration research using keyword network analysis, *Journal of Decision Science*, vol.23, no.1, pp.111-125, 2015.

[7] E. Yan and Y. Ding, Applying centrality measures to impact analysis: A coauthorship network analysis, *Journal of the Association for Information Science and Technology*, vol.60, no.10, pp.2107-2118, 2009.

[8] M. A. Kim, C. Y. Hwang and C. K. Suh, Topic detection in Korean SCM research using latent Dirichlet allocation, *Proc. of the 24th International Academic Conferences*, Barcelona, 2016.

[9] D. M. Blei, Probabilistic topic models, *Communications of the ACM*, vol.55, no.4, pp.77-84, 2012.
[10] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, vol.41, no.6, pp.391-407, 1990.
[11] Y. Zhang, G. Zhang, H. Chen, A. L. Porter, D. Zhu and J. Lu, Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research, *Technological Forecasting and Social Change*, vol.105, pp.179-191, 2016.