

컴퓨터 분야 논문지에서 저자의 관심분야에 대한 소셜 네트워크 분석

이주연 · 박유현*

Social Network Analysis of author's interest area in Journals about Computer

Ju-Yeon Lee · Yoo-Hyun Park*

Department of Computer Software Engineering, Dongeui University, Busan 47227, Korea

요 약

최근 다양한 분야에서 소셜 네트워크 분석 기법을 통해 개체들간의 상호작용을 분석하는 연구들이 진행되고 있다. 본 논문에서는 한국정보통신학회 논문지의 저자약력에 기술되는 저자의 관심분야를 소셜 네트워크 분석기법을 통해 다양한 측면에서 분석하고자 한다. 분석결과 많은 저자들이 임베디드, 보안, 영상처리, 무선통신, 빅데이터, USN, 네트워크, RFID 등의 관심분야를 중심으로 연구하고 있음을 알 수 있다.

ABSTRACT

Recently there are many researches about analyzing the interaction between entities by social network analysis in various fields. In this paper, we are going to analyze the author's interests area at the biography section in Journal of the Korea Institute of Information and Communication Engineering by social network analysis. The results show that many authors in that journal are mainly focusing on embedded, security, image processing, wireless network, big data, USN, network, RFID.

키워드 : 저자 관심분야 네트워크, 소셜 네트워크 분석

Key word : author's interests area network, social network analysis

접수일자 : 2015. 10. 05 심사완료일자 : 2015. 10. 21 게재확정일자 : 2015. 11. 04

* **Corresponding Author** Yoo-Hyun Park(E-mail:yhpark@deu.ac.kr, Tel:+82-51-890-1737)

Department of Computer Software Engineering, Dongeui University, Busan 47227, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2016.20.1.193>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

복잡한 사실이나 현상을 분석하기 위한 방법으로 개별적인 개체에 집중하기 보다는 이들을 집단적인 형태로 묶어서 분석하는 통계적 방법이 유용하다. 통계적 분석을 활용하여 복잡한 사실이나 현상을 정량화된 수치로 변환하여 전체적인 흐름을 파악할 수 있다.

최근에는 기존의 통계적 분석방법의 부족한 부분을 보완하기 위해 개체들간의 상호작용을 계량적으로 분석하여 거시적, 미시적 관계 패턴을 분석하는 소셜 네트워크 분석방법(social network analysis: SNA)[1]이 유행하고 있다. 소셜 네트워크 분석은 전통적인 통계학 기법보다는 그래프 이론 접근방법을 이용하여 사람들 사이의 상호 행동을 연구하는 분석방법으로 관계 또는 관련 있는 데이터에 내재되어 있는 의미 있는 패턴을 찾아낼 수 있다. 즉, 소셜 네트워크를 구성하는 개체의 여러 관계를 이해하기 위해 네트워크에서의 개체를 시각적이고 수리적으로 나타내며 이를 통해 네트워크의 규모(size), 밀도(density), 중앙성(centrality) 등을 파악할 수 있으며, 네트워크를 이루는 행위자(node)들의 특성이나 관계를 어떻게 파악하느냐에 따라 지표들이 다양하게 사용된다[1]. 특히 중앙성은 권력과 영향력이라는 개념과 연결되어 가장 많이 쓰이는 지표로써 네트워크에서 한명의 행위자가 얼마나 많은 다른 행위자들과 연결되어 있는가를 나타내는 개념이다[1]. 다시 말해, 사회연결망 이론에서는 연결망 핵심부에 위치한 사람이 가장 많은 자원과 자원동원능력을 가지고 있다고 보고, 연결정도(degree), 사이(betweenness), 인접(closeness), 위세(eigenvector) 등으로 세분화하여 연결망의 특징을 설명할 수 있다[1].

소셜 네트워크 분석기법은 인문, 경제, 공학 등 다양한 분야에서 많은 연구가 진행되고 있다. [2]에서는 유전자 소셜네트워크를 이용해 동물모델에서 인간의 질환을 연구할 수 있는 새로운 신호전달경로 모델을 발굴하는 웹기반 예측시스템을 개발하였고, [1]에서는 소셜 네트워크 분석을 통해 팀 스포츠에서의 패스 네트워크를 연구하였다. [3]에서는 소셜 네트워크의 구조적 측정 지표를 활용하여 추천 성과에 관해 연구하였다.

이러한 분야 중에서도 특정 분야 논문지에 나타난 저자, 키워드 등을 소셜 네트워크 분석을 통해 최근 연구 분야의 경향을 계량적으로 분석하여 거시적 흐름을 도

출하는 연구들도 있다[4-11].

한편 논문지 중에서 논문 저자의 약력을 기술하는 부분에서 저자의 관심분야를 추가적으로 나타내는 논문지들이 다수 존재한다. 컴퓨터 분야를 포함한 공학 분야에서 이러한 경향이 더욱 많이 나타난다. 이러한 저자의 관심분야는 과거와 현재의 관심분야 뿐만 아니라 가까운 미래에 집중적으로 연구할 부분까지 포함되는 경우가 많아 게재된 논문에만 국한하여 관계된 논문 키워드와는 성격이 다르다.

본 논문에서는 이러한 저자의 관심분야를 소셜 네트워크 분석기법을 통해 다양한 측면에서 분석하고자 한다. 이러한 분석을 통하여 얻어진 결과를 토대로 특정 논문지에서 주로 다루는 연구분야와 각 연구분야들 간의 관련정도, 특정 관심분야에 관한 연구자 그룹화 등에 활용될 것으로 기대된다.

II. 관련연구

2.1. 소셜 네트워크 분석

소셜 네트워크는 개인적인 인간관계가 확산되어 형성된 사람들 사이의 관계형태나 유형 혹은 구조를 말한다[3]. 소셜 네트워크는 노드와 관계로 구성되는데, 노드는 소셜 네트워크에 참여하는 각각의 참여자들을 의미하고 관계는 참여자들 사이에 가지는 사회적인 연결 관계를 의미한다[3].

소셜 네트워크의 구조를 파악하기 위한 방법으로 중심에 위치하는 참여자에게 얼마나 집중되어 있는가를 나타내는 중심성 개념이 대표적으로 사용되고 있다[3]. 중심성은 노드간의 관계 중 어느 측면에 초점을 맞추어 중심을 측정하는가에 따라서 연결정도, 근접, 사이 중심성이 사용되고 있다[9].

2.2. 공저자 및 키워드 네트워크

최근 학술 연구는 이공계뿐만 아니라 인문사회과학 분야에서도 연구자들 간의 공동연구는 갈수록 증가하고 있다[8]. 연구자들의 공동연구는 각 연구자들을 하나의 노드로 표현하는 소셜 네트워크로 나타낼 수 있으며 이러한 소셜 네트워크를 공저자 네트워크라고 한다[8]. 공저자 네트워크는 소셜 네트워크 분석을 통하여 개별 연구자의 사회적/학술적 관계와 학술적 영향력 정도 및

공저자 네트워크의 형태와 특성들을 파악할 수 있다[8]. 지금까지의 공저자 네트워크에 대한 연구는 공저자 네트워크에 나타난 일반적인 기초통계, 네트워크의 구조적 속성, 학술적 영향력 지표로 중심 영향력 지표인 연결정도 중심성, 근접 중심성, 매개중심성, 위세영향력을 나타내는 고유벡터 중심성, 신뢰 영향력을 나타내는 페이지랭크 등을 제시하고 그 변수들 간의 상관관계를 제시하는데 있었다[8].

이와 같은 공저자 및 키워드 네트워크에 대한 분석은 체육학[1], 생물학[2], 로봇 및 제어자동 분야[6], 지리공간정보학[7], 통계학[8], 물리학[9], 문헌정보학[10], 간호학[11] 등 대부분의 학문분야에서 수행되었다.

III. 저자 관심분야 네트워크 분석

본 논문은 한국정보통신학회에 2014년 1년간 게재된 논문들의 저자약력에서 저자의 관심부분을 소셜 네트워크 분석 기법을 통해 분석하여 저자 관심분야 네트워크를 시각화 하고 영향력이 있는 주요 관심분야들을 식별하고 네트워크의 그룹들을 나누어 살펴보고자 한다.

3.1. 분석 대상 데이터

본 논문에서 분석하고자 하는 기초 데이터는 한국정보통신학회에 2014년 1년간 투고된 논문을 대상으로 한다. 2014년 한 해 동안 한국정보통신학회에 게재된 논문은 총 414편이었고, 이 논문들의 총 저자 수는 927명, 총 관심분야 3,202개이었으며, 저자들은 평균 3.5개의 관심분야를 기술하고 있었다. 중복 저자를 제거 한 저자는 641명, 공백, 영문국문 변환, 표기법 통일 등으로 정제된 관심분야 수 1,006개로 정리되었다.

3.2. 관심분야별 네트워크 구성 방법

공저자 네트워크의 경우, 분석을 위해 데이터의 구성을 주저자에서 공동연구자로 연결되도록 데이터를 구성하고, 공동연구자간의 연결은 고려하지 않는 경우가 많다[9]. 이는 주저자 A와 공동연구자 B, C가 있는 경우 네트워크에서의 연결은 A-B, A-C로의 연결만을 고려했음을 의미한다[9].

하지만, 이와 같은 공저자 네트워크에서 방향성의 방향을 인식할만한 충분한 근거나 기준을 마련하는 것은

쉽지 않다[10]. 즉, 저자A에서 저자B로의 방향성 있는 링크를 가지는 경우 저자 A는 초급 연구자 저자B는 중진 이상의 연구자를 나타낸다고 볼 수도 있다는 연구가 있으나 동급 수준의 연구자들이 협력할 수 있으며, 초급 연구자들도 단독으로 논문을 작성할 수 있다는 점에서 방향성의 기준을 설정하는 것은 타당하지 않아 보인다[10]. 또한, 저자의 여러 관심분야에서 하나의 관심분야를 주요 관심분야로 선택하는 것은 자의적으로 해석될 수 있기 때문에 본 논문에서는 저자의 관심분야들을 모두 상호관계가 있는 것으로 표현하였다. 즉, 저자1의 관심분야가 A, B, C라면, A-B, A-C, B-C의 연결을 모두 고려하였다. 또한, 저자 2의 관심분야가 A, D, F라면 기존의 네트워크에서 A-D, A-F를 추가하는 방식으로 네트워크를 구성하였다. 이렇게 저자 네트워크를 구성하면 저자 관심분야 노드(node)의 수는 1,006이며, 이들 사이 연결된 에지(edge)의 수는 2,718개로 나타났다.

3.3. 시각화 및 주요 컴포넌트 분석

본 논문에서는 저자의 관심분야 네트워크를 시각화 하기 위한 도구로 Gephi 0.8.2[12]를 사용하였다. [그림 1]은 Gephi를 통해 저자 관심분야 네트워크를 시각화한 그림이다. 전체 네트워크에는 총 80개의 그룹이 존재하며 이들 각각은 다른 색깔로 표시하였다. 하지만 [그림 1]의 왼쪽 그림을 살펴보면, 네트워크의 바깥쪽에 가운데 네트워크와 연결되지 않는 독립그룹들이 다수 존재하고 있다. 이들 독립그룹은 모두 65개로 나타났으며, 저자들이 본인의 관심분야를 너무 세밀하게 기술하여 다른 저자들의 관심분야와 관계를 맺지 못한 경우를 나타낸다. 이들은 2~11개의 관심분야들끼리만 연결되어 독자적인 그룹을 이루고 있으며 이를 제거하여 시각화한 결과가 [그림 1]의 오른쪽 그림이다.

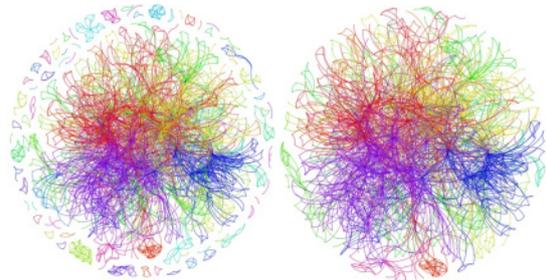


Fig. 1 Visualization of author's interesting area network

[그림 2]는 저자 관심분야 네트워크에서 그룹들의 분포를 보이고 있다. 가로축은 그룹의 번호이며, 세로축은 해당 그룹에 속한 노드(저자 관심분야)의 수이다. 그림에서와 같이 대부분의 그룹이 10개 이하의 노드를 포함하고 있으며 이들은 다른 네트워크와 연결되지 못하는 독립그룹들이다. 독립그룹들을 제거한 후 저자 관심분야 네트워크는 [그림 1]의 오른쪽과 같이 모두 762개의 저자 관심분야가 15개의 그룹으로 나타났다.

저자 관심분야 네트워크의 기본 측정값은 [표 1]과 같다. [표 1]에서와 같이 네트워크의 평균 연결정도(average degree)는 6.213로 각 관심분야가 평균적으로 이웃 관심분야와 연결되어 있는 수를 나타낸다. 네트워크 지름(network diameter)은 9로 네트워크 내 임의의 두 노드들 간에 설정되는 경로들 중에서 가장 많은 수의 링크들을 경유해야 하는 경우의 에지(edge) 수를 나타낸다. 즉, 저자 관심분야 네트워크의 노드 중에서 가장 멀리 떨어진 두 노드간의 거리가 9이다. 그래프 밀도(graph density)는 그래프에서 생성할 수 있는 모든 에지수에 대한 실제 에지수의 비율을 0과 1사이의 수로 표현한 것으로 저자 관심분야 네트워크의 그래프 밀도는 0.008이다.

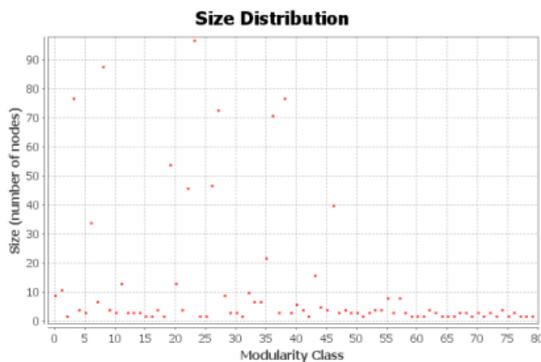


Fig. 2 Modularity analysis of author's interest area network

Table. 1 Network Overview

average degree	6.213
network diameter	9
graph density	0.008
modularity	0.631
average clustering coefficient	0.793
average path length	3.59

클러스터 집단화 계수(clustering coefficient)는 특정 노드와 이웃 노드가 연결되어 있을 확률로 0과 1사이의 수로 표현하는데 평균적인 저자 관심분야 네트워크의 클러스터 집단화 계수는 0.793이고, 임의의 두 노드 사이의 평균 거리는 3.59로 나타났다.

3.4. 중심성(Centrality) 분석

저자 관심분야 네트워크에 존재하는 핵심 관심분야를 파악하기 위해서는 중심성 분석이 매우 유용한 방법이다. 중심성은 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 측정하는 지표로 노드간의 관계 중 어느 측면에 초점을 맞추어 중심을 측정하는가에 따라서 연결정도, 근접, 사이 중심성이 대표적인 중심성 척도로 사용되고 있다[9].

연결정도 중심성(Degree Centrality)은 네트워크에서 특정 노드가 주변의 다른 노드들과 연결되어 있는 정도를 측정하는데, 이는 주로 링크의 개수를 합산하여 구한다[5]. [그림 3]은 연결정도 중심성이 높은 순으로 상위 20개의 저자 관심분야를 나타내고 있다. 임베디드의 경우 84개의 이웃 노드들과 연결되어 있어 연결정도 중심성이 가장 높았고, 보안, 영상처리, 무선통신 순으로 높은 수치를 나타냈다.

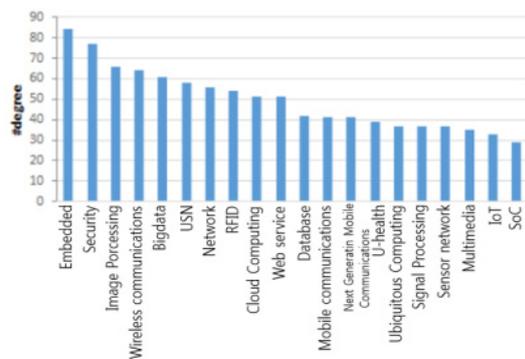


Fig. 3 Top 20 Degree Centrality

근접 중심성(Closeness Centrality)은 한 노드가 다른 노드들과 얼마나 평균적으로 가까이 있는지를 측정하는 전체 중심성을 표현하는 대표적인 지표로서, 다른 노드와의 경로거리로 측정할 수 있다[9]. 근접 중심성을 분석함으로써 네트워크 전역에서 가장 일반적인 영향력을 가지는 저자를 가늠할 수 있으며, 근접 중심성

이 높은 개인은 다른 모든 사람들과 가장 짧은 거리에 위치하여, 한 노드가 전체 네트워크의 다른 노드와 연결된 거리를 모두 고려할 수 있게 되므로 위치적 이점으로 인해 네트워크에서의 정보접근이 용이하고 영향력의 정도가 큰 노드라고 할 수 있다[9]. 이러한 근접 중심성은 노드간의 최단거리(경로거리)의 합의 역수로 계산되며, 각 노드로의 최단거리의 합이 가장 작은 노드가 전체 중앙성이 가장 높은, 즉 네트워크 전체의 중심을 차지하는 노드라고 할 수 있다[9]. [그림 4]는 근접 중심성 값이 가장 작은 상위 20개의 저자 관심분야를 나타내고 있다. 연결 중심성과 같이 임베디드가 가장 근접 중심에 있으며, 무선통신, 보안의 순으로 분석되었다.

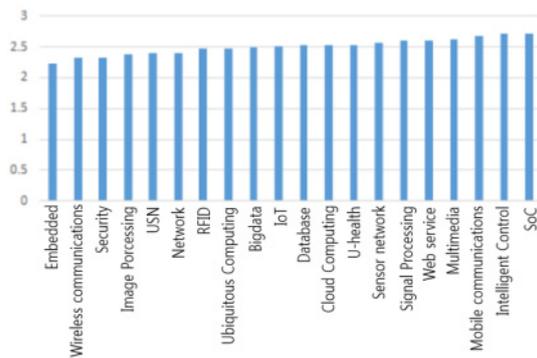


Fig. 4 Top 20 of Closeness Centrality

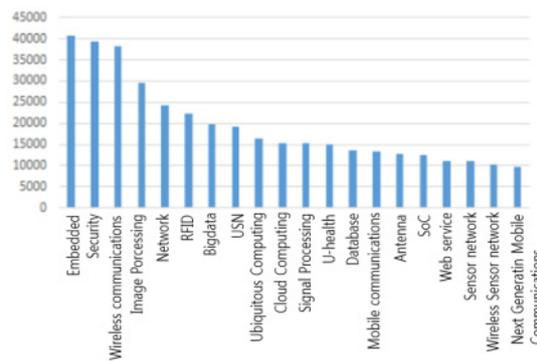


Fig. 5 Top 20 of Betweenness Centrality

매개 중심성이라고도 하는 사이 중심성(Betweenness Centrality)은 한 노드가 다른 노드들 사이의 최단경로

상에 위치하는 정도를 측정하는 전체 중심성을 측정하는 지표이다[9]. 노드가 최단 경로상에 위치하는 횟수가 많을수록 중심성은 높아지며, 상이한 집단 간을 연결하는 노드일수록 높게 나타난다[9]. [그림 5]는 사이 중심성 값이 가장 높은 상위 20개의 저자 관심분야를 나타내고 있다. 연결 중심성 및 근접 중심성과 같이 임베디드가 가장 근접 중심에 있으며, 보안, 무선통신의 순으로 분석되었다.

[표 2]는 저자들의 관심분야 네트워크에서 연결성, 근접, 사이 중심성을 기준으로 상위 20개를 정리한 표이다. 이들은 약간의 차이가 있지만 비슷한 결과를 보였고, 상위 10개에 속하는 임베디드, 보안, 영상처리, 무선통신, 빅데이터, USN, 네트워크, RFID는 한국정보통신학회 2014년에 논문을 게재한 저자들의 관심분야로 매우 큰 영향을 주고 있는 것으로 볼 수 있다.

Table. 2 Degree, Closeness, Between Centrality

	Degree	Closeness	Between
Embedded	1	1	1
Security	2	3	2
Image processing	3	4	4
Wireless communications	4	2	3
Bigdata	5	9	7
USN	6	5	8
Network	7	6	5
RFID	8	7	6
Cloud computing	9	12	10
Web service	10	16	17
Database	11	11	13
Mobile Communications	12	18	14
Next generation mobile Communications	13	28	20
U-health	14	13	12
Ubiquitous Computing	15	8	9
Signal Processing	16	15	11
Sensor network	17	14	18
Multimedia	18	17	21
IoT	19	10	23
SoC	20	20	16

IV. 결 론

최근 다양한 분야에서 소셜 네트워크 분석 기법을 통해 개체들간의 상호작용을 계량적으로 분석하여 거시적, 미시적 관계 패턴을 분석하는 시도들이 이루어지고 있다. 본 논문에서는 한국정보통신학회 논문지에 2014년 1년동안 게재된 논문의 저자약력에 기술되는 저자의 관심분야를 소셜 네트워크 분석기법을 통해 다양한 측면에서 분석하였다.

분석결과, 14개의 그룹으로 나누어진 네트워크가 존재하였고 연결성, 근접, 사이 중심성을 기준으로 판단할 때 임베디드, 보안, 영상처리, 무선통신, 빅데이터, USN, 네트워크, RFID 등의 분야들이 게재된 논문들 중 다수 논문들이 다루었던 주제임을 보여 주었다. 이러한 분석을 통하여 얻어진 결과를 토대로 특정 논문지에서 주로 다루는 연구분야와 각 연구분야들 간의 관련정도, 특정 관심분야에 관한 연구자 그룹화 등에 활용될 것으로 기대된다.

ACKNOWLEDGMENTS

This Work was supported by Dong-eui University Foundation Grant (2015)

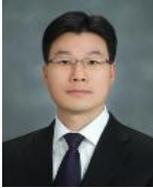
REFERENCES

- [1] Hyeojin Kim, "Notational Analysis of Sports using Social Network Analysis," *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*, vol. 9, no.1, pp.99-112, Jan. 2007.
- [2] Sohyun Hwang, Eiru Kim, Sunmo Yang, Edward M. Marcotte, Insuk Lee, "MORPHIN: a web tool for human disease research by projecting model organism biology onto a human integrated gene network," *Nucleic Acids Research*, vol.42, pp.147-153, May. 2014.
- [3] Sung-Mahn Ahn, In Hwan Kim, Byounggu Choi, Yoonho Cho, Eunhong Kim, Myeong-Kyun Kim, "Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis," *Journal of Society for e-Business Studies*, vol.17, no.2, pp. 129-147, May, 2012.
- [4] Injai Kim, Jaewon Choi, Kihwan Kim, Geumyoung Min, "Social Network Analysis of Professional Groups based on Co-author and Review Networks," *Journal of the Korea society of IT services*, vol.13 no.1, pp.181-196, Mar. 2014.
- [5] Yong-Gu Lee, Yun-Hee Woo, "A Study on Analyzing Co-authorship Networks of Library and Information Science in Taiwan," *Journal of the Korea Society for Information Management*, vol. 32, no.2, pp.167-192, Jun. 2015.
- [6] Suk Lee, Byung Kyu Jeong, "Research Trends in Robotics, Control, and Automation based on 2011 ~ 2014 IEEE ICRA Proceedings(Part1 ~ 3)," *J. Korean Soc. Precis. Eng.*, vol.32, no.3, pp. 233-241, Mar. 2015.
- [7] Byeongsun Kim, Minwoo Jeong, Sangeum Jeon, Dongbin Shin, "Global Research Trends on Geospatial Information by Keyword Network Analysis," *Journal of Korea Spatial Information Society*, vol.23, no.1, pp. 69-77, Feb. 2015.
- [8] Heuiju Chun, "The comparison of coauthor networks of two statistical journals of the Korean Statistical Society using social network analysis," *Journal of the Korean Data & Information Science Society*, vol.26, no.2, pp.335-346, Mar. 2015.
- [9] Hye-Sun Lim, Tai-Woo, "A Study on Co-authorship Network in the Journals of a Branch of Logistics," *IE Interfaces*, vol.25, no.4, pp. 458-471, Dec. 2012.
- [10] Soo-Sang Lee, "A Analytical Study on the Properties of Coauthorship Network Based on the Co-author Frequency," *Journal of Korean Library and Information Science Society*, vol.42, no.2, pp.105-125, Jun. 2011.
- [11] Soo-Kyoung Lee, Senator Jeong, Hong-Gee Kim, Young-Hee Yom, "A Social Network Analysis of Research Topics in Korean Nursing Science," *J Korean Acad Nurs*, vol.41 no.5, pp.623-632, Oct. 2011.
- [12] Gephi(The Open Graph Viz Platform) [Internet]. Available: <http://gephi.github.io/>



이주연(Ju-Yeon Lee)

2013년 동의대학교 문헌정보학과 문헌정보학사
2014년 ~ 현재 동의대학교 컴퓨터소프트웨어공학과 석사과정
※관심분야 : 데이터베이스, 정보검색



박유현(Yoo-Hyun Park)

1996, 1998, 2008년 부산대학교 전자계산학과 이학사, 이학석사, 이학박사
2000년 한국국방연구원(KIDA) 연구원
2001년 ~ 2009년 한국전자통신연구원(ETRI) 선임연구원
2009년-현재 동의대학교 컴퓨터소프트웨어공학과 부교수
2012년-2014년 동의대학교 부산T융합부품연구소 부소장
※관심분야 : 인터넷 시스템, 클라우드 시스템, 빅데이터, 소프트웨어 품질, IT 융합 서비스