

학술지 단위 서지결합분석을 통한 빅데이터 연구분야의 학제적 구조에 관한 연구

A Study on Interdisciplinary Structure of Big Data Research with Journal-Level Bibliographic-Coupling Analysis

이보람 (Boram Lee)*

정은경 (EunKyung Chung)**

초 록

현대사회의 다양하고 복잡한 문제들을 해결하기 위해 학문영역을 넘나드는 학제적 연구가 등장하게 되었다. 본 연구에서는 최근 다양한 영역에서 주목 받고 있는 빅데이터 분야를 대상으로 학제성을 규명하고 학제적 구조를 파악하고자 하였다. 이를 위해 빅데이터를 다룬 학술지 총 1,083종의 데이터를 수집하였다. 이 중 420종(38.8%)의 학술지에 둘 이상의 Web of Science SC범주가 부여되었고, 239종(22.1%)에 부여된 SC범주는 상이한 학문영역에 속하여 빅데이터 분야의 비교적 높은 학제성을 확인할 수 있었다. 이와 함께 논문 게재 상위 56종의 학술지를 대상으로 서지결합분석 네트워크를 생성한 결과 총 10개의 군집이 나타났다. 10개 군집 중 7개 군집이 컴퓨터공학 분야에 해당하여 대부분의 연구가 빅데이터의 저장, 처리, 분석 등 기술적인 부분에 집중되어 있었다. 이외에도 군집분석을 통해 과학기술, 공학, 커뮤니케이션, 법학, 지리학, 생명공학 등 다양한 분야에서 빅데이터의 분석과 활용에 관한 연구가 이루어지고 있음을 확인할 수 있었다. 마지막으로 네트워크에서 매개중심성, 최근접중심성, 삼각매개중심성을 측정한 결과 컴퓨터공학 분야의 학술지들이 네트워크에 미치는 영향력이 크고 주제적 연관성이 강한 것으로 나타났다.

ABSTRACT

Interdisciplinary approach has been recognized as one of key strategies to address various and complex research problems in modern science. The purpose of this study is to investigate the interdisciplinary characteristics and structure of the field of big data. Among the 1,083 journals related to the field of big data, multiple Subject Categories (SC) from the Web of Science were assigned to 420 journals (38.8%) and 239 journals (22.1%) were assigned with the SCs from different fields. These results show that the field of big data indicates the characteristics of interdisciplinarity. In addition, through bibliographic coupling network analysis of top 56 journals, 10 clusters in the network were recognized. Among the 10 clusters, 7 clusters were from computer science field focusing on technical aspects such as storing, processing and analyzing the data. The results of cluster analysis also identified multiple research works of analyzing and utilizing big data in various fields such as science & technology, engineering, communication, law, geography, bio-engineering and etc. Finally, with measuring three types of centrality (betweenness centrality, nearest centrality, triangle betweenness centrality) of journals, computer science journals appeared to have strong impact and subjective relations to other fields in the network.

키워드: 빅데이터, 학제성, 서지결합 분석, 네트워크 분석

big data, interdisciplinarity, bibliographic coupling analysis, network analysis

* 이화여자대학교 일반대학원 문헌정보학과(rajeunir@gmail.com) (제1저자)

** 이화여자대학교 사회과학대학 문헌정보학 전공 부교수(echung@ewha.ac.kr) (교신저자)

■ 논문접수일자: 2016년 8월 17일 ■ 최초심사일자: 2016년 8월 31일 ■ 게재확정일자: 2016년 9월 5일
■ 정보관리학회지, 33(3), 133-154, 2016. [http://dx.doi.org/10.3743/KOSIM.2016.33.3.133]

1. 서론

정보기술의 발전과 함께 정보의 원천으로서 빅데이터의 중요성은 지속적으로 증가하고 있는 추세이다. 빅데이터가 이전에 존재하지 않다가 최근에 갑자기 등장한 새로운 개념으로 여겨질 수 있으나 실제로는 그 이전부터 개념적, 실체적으로 존재해 온 개념이다(이정미, 2013; Park & Leydesdorff, 2013). 빅데이터라는 용어를 사용하기 이전에도 데이터를 수집, 분석, 처리하는 다양한 기술 및 관련 연구가 존재해 왔다. 그러나 인터넷의 대중화와 모바일 기기의 보급으로 인해 데이터의 양이 폭발적으로 증가하고 그 유형이 다양화되면서 이전의 방식으로는 이러한 대규모의 데이터를 수집, 저장, 처리하는 것이 불가능해졌다. 따라서 이 비정형적이고 복잡하고 방대한 데이터를 수집, 저장, 처리하기 위한 도구 및 기술의 개발과 관련 연구가 활발히 일어났다. 또한 빅데이터를 분석하고 활용함으로써 데이터를 기반으로 한 해당 현상·시점의 이해 및 미래 예측을 위한 맥락적 근거를 얻을 수 있어 학계뿐 아니라 산업, 언론, 정부 등 다양한 영역에서 빅데이터에 대한 관심을 나타내고 있다. 국내를 비롯하여 미국, 유럽, 일본 등 해외 각국에서 공공정보 개발 및 공공서비스 개혁 등에 국가차원의 공공 빅데이터 정책을 추진하고 있으며, 민간부문에서도 마케팅, 비즈니스 차원에서 빅데이터 연구를 근거로 한 예측 및 전략수립 등을 진행 중이다. 빅데이터는 정보시스템, 인공지능, 하드웨어, 소프트웨어를 포함하는 컴퓨터공학에서 인간공학, 뇌과학, 언어학까지 망라되는 여러 분야가 적용되는 분야이며, 정치, 사회, 경제, 문

화, 과학기술 등 전 영역에 걸친 학제적 성격을 지닌다. 따라서 빅데이터 분야를 이해하기 위해서는 컴퓨터공학, 통계학, 경제학, 수학, 정보학 등 학제적 이해가 필요하며 학문적인 지식뿐 아니라 통합적 사고와 직관력이 요구된다(김현영, 지현수, 이화순, 지종덕, 2014).

학제성(interdisciplinarity)은 학문영역을 넘나드는 연구 활동을 의미하며(Morillo, Bordons, & Gómez, 2003), 현대사회에서 개별 학문의 전문적 지식만으로 해결할 수 없는 다양하고 복잡한 문제들을 해결하기 위해, 다른 학문분야의 지식이나 연구방법론을 빌려 새로운 시각에서 문제를 고찰하고 이들을 통합하여 적절한 해결방안을 개발하고자 하는 시도들이 등장하게 되었다. 이렇게 학문 간의 융·복합이 빈번하게 이루어지는 환경에서 여러 학문분야의 학술 활동을 개별적으로 파악하는 것에서 벗어나 전체 학문분야의 활동과 연계 상황 등을 종합적으로 파악하는 학제성의 연구는 학문의 진보, 흐름, 경향, 예측 등을 가능하게 함으로써 미래 교육과 연구 활동의 기초가 된다. 특히 컴퓨터공학, 경영·경제학, 수학, 문헌정보학 등 다양한 학문영역과 밀접한 관계를 갖는 빅데이터 분야의 학제성 연구를 통해 그 학문적 특성을 밝히고 지적구조를 규명하며 앞으로의 연구 방향을 예측할 수 있을 것이다.

학문분야 간 학제성을 측정하려는 시도에는 다양한 접근 방법이 존재해 왔으나, 계량정보학적 방법이 가장 널리 사용되어 왔다(Wagner, Roessner, Bobb, Klein, Boyack, Keyton, Rafols, & Börner, 2011). 학제성을 측정하는 계량정보학적 방법으로는 연구자들의 협업 또는 공저 행태를 분석하거나 학술지 또는 논문에 부

여된 주제어 및 주제범주를 분석하고 문헌 간 인용 행태를 분석하는 방법 등이 있다. 따라서 본 연구에서는 빅데이터 분야의 학제적 특성의 규명을 연구의 목적으로 하고 이를 달성하기 위해 다음의 두 가지 구체적인 연구 문제를 설정하였다. 첫째, 빅데이터 분야의 학제성을 측정하기 위해 빅데이터를 주제로 다룬 논문이 게재된 학술지의 주제범주를 분석한다. 둘째, 빅데이터 관련 학술지 간 서지결합분석 및 네트워크 생성을 통해 그 지적구조를 도출하고 네트워크의 매개중심성, 최근접중심성, 삼각매개중심성 등 다양한 중심성 분석을 실시하여 빅데이터 분야의 학제적 구조를 파악하고자 한다.

2. 관련 연구

2.1 빅데이터 주제분야 연구

빅데이터는 새롭게 등장한 개념이 아니라 기존의 기술로는 적절하게 처리하지 못하고 방치해둘 수밖에 없었던 데이터를 의미한다. 이러한 데이터는 소프트웨어, 하드웨어, 네트워크 등 정보기술이 발전하면서 처리, 통제 및 분석이 가능해짐에 따라 새롭게 관심의 대상이 되었다(이정미, 2013; Park & Leydesdorff, 2013; Rousseau, 2012). 일반적으로 빅데이터의 특징은 3Vs로 데이터의 방대한 규모(Volume), 다양한 유형(Variety) 그리고 빠른 생성·처리 속도(Velocity)를 든다(이정미, 2013; McAfee & Brynjolfsson, 2012). Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, Byers(2011)는 데

이터베이스의 규모에 초점을 맞춰 빅데이터를 일반적인 데이터베이스 소프트웨어가 저장, 관리, 분석하기에 그 범위를 넘어서는 대규모의 데이터라고 규정하였고, Gartner(2012)는 향상된 시사점을 제공하고 더 나은 의사결정을 위해 사용되는 데이터 가운데 비용 효율이 높고, 혁신적이며, 대용량, 고속 및 다양성의 특징을 가진 정보자산이라고 정의하였다.

빅데이터 관련 연구는 매우 활발히 이루어지고 있으며 많은 양의 보고서와 논문이 출판되고 있지만 대부분은 빅데이터 처리·분석에 필요한 하드웨어, 소프트웨어, 분석 방법 등 기술적 측면과 관련한 연구이거나 앞으로 어디서 어떻게 활용해야 하는 것인가에 대한 전망에 대한 내용이 주를 이룬다. 반면 빅데이터 자체를 대상으로 분석을 시도한 연구는 아직 많이 이루어지고 있지 않다(김원중, 2014). 연구분야로서 빅데이터의 학문적 성격을 규명하기 위한 시도를 살펴보면, Yang(2013)은 중국 내 빅데이터 관련 연구의 성격을 파악하기 위해 동시출현단어분석, 군집분석, 다차원척도분석 등 계량정보학 분석기법을 활용하여 논문생산 추이, 저자 및 연구기관, 관련 학술지의 분포와 출현빈도가 높은 키워드 등을 분석하였다. 그 결과 중국의 빅데이터 분야 논문의 생산성은 Price 법칙에, 저자분포는 Lotka 법칙에 부합한다는 것을 밝혀냈으며 관련 주제는 IT 분야와 문헌정보학 분야에 집중되어 있어 아직은 초기 연구단계에 머물러 있는 상태임을 알아냈다. Shiri(2014)는 패킷분석을 통해 빅데이터를 data type, environment, people, operations and activities, analytics, metadata의 6개 패킷과 각각의 하위 패킷으로 분류하여 그 성격

을 규명하고자 하였다.

빅데이터와 관련한 주제분야를 파악하고자 시도한 연구들을 살펴보면, 김원중(2014)은 동시출현단어분석을 통해 빅데이터 분야의 네트워크를 생성하여 연구동향과 관련 연구영역을 알아보고자 하였다. 그 결과 빅데이터 분야에는 전반적인 IT 분야뿐 아니라 사회과학, 경영, 생물학, 의학 등 다양한 영역의 주제어들이 포함되어 있으나 물리, 화학, 재료과학 등과 관련한 주제어는 출현하지 않음을 확인하였다. Huang, Zhang, Youtie, Porter, & Wang(2016)은 국가 차원에서 학제적 성격이 강한 연구를 어떻게 지원하든지 이해하기 위해 다양한 학문적 특성을 지닌 빅데이터 분야를 대상으로 연구를 실시하였다. 해당 연구는 미국의 National Science Foundation(NSF)과 중국의 National Natural Science Foundation of China(NSFC)에서 지원을 받은 빅데이터 관련 연구들을 대상으로 해당 기관에서 부여한 주제범주를 분석하고 연구계획서 제목에 나타난 키워드의 동시출현분석 및 Web of Science(WoS)에서 검색한 관련 논문의 WoS 주제범주 분석을 실시하였다. 연구 결과, 미국과 중국 모두 빅데이터와 관련하여 가장 큰 비중을 차지하는 주제분야는 정보학으로 나타났으며, 미국은 공학과 수학·자연과학이, 중국의 경우 경영학과 지구과학이 그 다음으로 큰 비중을 차지하여, 미국은 공학과 응용에, 중국은 이론과 경영에 중점을 두고 있는 것을 확인하였다. 두 나라에서 우선순위를 차지하는 주제분야는 각기 다르게 나타났지만 빅데이터 분야에서 다루어지는 주제분야들은 유사하게 나타났다. 또한 네트워크 분석을 통해 빅데이터 연구가 주로 컴퓨터공학에 의해

이루어지고 있지만 다른 다양한 학문분야에서도 폭넓게 실시되고 있음을 밝혀냈다.

2.2 학제성 규명 연구

학제간 융합, 복합, 통섭 등 다양하게 표현되는 학제성은 그 개념의 정의와 관련하여 아직 명확한 합의가 이루어지지 않았다. 이것은 학제성의 개념이 학문의 구조나 성격과 같은 인식론적 문제뿐 아니라 사회적, 정치적, 문화적 상황과 복잡하게 얽혀있기 때문이다(김대현, 강이화, 2010). 미국의 학제성 관련 공동 연구 프로그램인 NAKFI(National Academies Keck Futures Initiative)는 학제적 연구를 두 개 이상의 전문 지식 체계나 연구 관행으로부터 관점·개념·이론 또는 도구·기법 또는 정보·자료를 통합하여 팀 또는 개인이 수행하는 연구 유형이라고 정의하였다. 그리고 학제적 연구의 목적은 단일 분야의 범위 안에서 해결할 수 없는 문제를 해결하고 학제적 연구에 대한 기본적 이해를 증진시키는 것이라고 밝혔다(National Academies, 2005). 김대현과 강이화(2010)는 학제성 개념을 분석하여 다음과 같은 핵심요소를 제시하였다. 첫째, 관련된 학문영역의 수는 둘 이상이며, 둘째, 여러 학문들이 분리된 상태가 아닌 통합된 상태를 지향하고, 셋째, 학제성은 결과가 아닌 과정이며, 넷째, 학제적 연구의 최종 목표는 학제적으로 설명되는 새로운 인지적 향상이나 문제 해결이다. 이를 바탕으로 이들은 학제성을 기존의 하나의 학문으로는 풀 수 없거나 여러 학문영역에 걸쳐 탐구할만한 주제를 가지고 여러 학문의 관점을 종합·통합하고 나아가 학제적 이

해를 도모하는 교육과 연구의 과정으로 정의하였다.

OECD(1998: Morillo, Bordons, & Gómez, 2003에서 재인용)는 학문분야 간 통합의 수준에 따라 학제성을 세 가지 유형으로 나누었다. 첫째, 다학문성(multidisciplinarity)은 단일 학문분야의 지식만으로는 분석하기 어려운 복잡한 문제를 해결하기 위해 상이한 학문분야의 지식, 정보, 방법론 및 관점을 공유하는 것으로 학문분야 간 통합이 성립되기 이전 단계의 연구를 의미한다. 둘째, 간학문성(interdisciplinarity)은 복잡한 문제들의 전체적인 관점이나 일반적인 이해를 위해 두 개 이상의 상이한 학문분야들의 데이터, 방법론, 도구, 개념, 이론 등을 통합하는 것을 말한다. 다학문성이 개별 학문분야의 단순한 합을 의미한다면, 간학문성은 이와 달리 개념적, 방법론적으로 어느 정도 통합이 이루어진 상태를 뜻한다. 셋째, 초탈학문성(transdisciplinarity)은 다학문적 연구나 간학문적 연구의 수준을 넘어 학문분야 간 수렴이 이루어지고 학문분야의 인식론적 상호 통합이 수반되는 연구를 의미한다.

학문분야 간 학제성을 측정하려는 시도에는 다양한 접근 방법이 존재해 왔으나, 계량정보학적 방법이 가장 널리 사용되어 왔다(Wagner et al., 2011). 학제성을 측정하는 계량정보학적 방법으로는 연구자들의 협업 또는 공저 행태를 분석하거나 학술지 또는 논문에 부여된 주제어를 분석하고 문헌간 인용 행태를 분석하는 방법 등이 있다. 연구자 및 저자를 중심으로 한 학제성 측정 연구로 연구자의 지적이동 패턴 연구를 들 수 있다. Hargens(1986)는 미국의 박사학위 소지자를 대상으로 이전 전공 분야에

서 현재 학문분야로의 이동 패턴을 연구하였으며, Urata(1990) 역시 일본의 인문학과 사회과학 분야에 소속된 연구자들의 지적 이동 분석과 인용분석을 함께 수행하였다. 또한 학제적 연구에 관여하는 연구자들의 협력 행태 분석 또는 공저자 분석을 통해 연구의 학제성을 측정할 수 있다. Qin, Lancaster, & Allen(1997)은 학문분야 간 협력 및 학제성의 정도를 측정하였는데 협력의 정도는 저자의 수, 저자의 학회/학과 단위 소속의 수, 학과 단위 소속에 반영된 학문분야의 수 등 협력적 저술의 정도로 측정하고 학제성의 정도는 해당 논문들이 인용한 학술지의 학문분야의 수로 측정하였다. 정은경(2011)은 '디지털 도서관'이라는 주제가 갖는 학제적 성격을 규명하기 위해 공저자 분석을 실시하여 디지털 도서관이 다양한 학문분야에서 다루어지는 학제적 연구분야임을 밝혔다.

숙련된 전문가에 의해 부여된 학술지의 주제범주를 일종의 분류체계로 보고 주제범주 간 연결성을 파악하여 학제성 분석연구의 대상 데이터로 채택한 연구들도 수행되었다. Morillo, Bordons, Gómez(2001)는 하나의 주제범주가 부여된 학술지보다 둘 이상의 주제범주가 부여된 학술지가 더 학제적인 성격을 띠 것이라는 전제 하에 화학분야의 학제성을 측정하였다. 학문분야 간 관계 및 지식교환의 강도를 나타내기 위해 동시출현단어 분석(co-word analysis)과 동시분류 분석(co-classification analysis)으로 학문분야의 학제성을 측정하기도 했다. 이재윤과 정주희(2006)는 국내 인지과학 분야의 학제적 구조를 파악하기 위하여 연구자의 소속기관 정보와 논문의 표제어의 동시출현빈도를 분석하였고, 김민지, 박정규, 이유아, 허은녕(2011)

은 태양전지 연구의 학제적 다양성을 규명하기 위해 ISI 주제범주를 대상으로 한 논문에 동시에 출현하는 주제범주 쌍의 빈도를 분석하였다.

인용은 계량정보학 분야에서 학제성을 규명하기 위한 방법들 중에서도 중요한 연구 수단으로 인식되어 다양하게 활용되어 왔다. Small (2010)은 동시인용 분석을 통해 학제성을 보여주는 과학지도를 작성하고 지도에 나타난 학제적 연결을 설명하였다. 그리고 인용된 주제분야의 수뿐만 아니라 주제분야의 인용 분포의 균등함을 함께 측정할 수 있는 다양성 지수인 브릴로앵 지수(Brillouin index) 또한 학제적 성격을 규명하기 위한 다양한 연구에 활용되었다. Steele과 Stier(2000)는 입학 분야 내 환경학 영역의 학제적 연구 영향력을 조사하기 위해 저자의 소속기관, 연구 주제, 인용문헌 등 세 가지 유형의 다양성을 브릴로앵 지수로 측정하여 비교하였다. 정연경(2012)은 국내 기록관리학의 학제적 성격을 분석하기 위하여 한국기록관리학회지의 연도별 브릴로앵 지수를 측정하여 시계열 분석을 실시하였다.

본 연구에서는 학제적 특성을 파악하는 여러 방법 중 복수의 주제범주가 부여된 학술지보다 단일 주제범주가 부여된 학술지보다 더 학제적인 성격을 띠는 것(Morillo, Bordons, & Gómez, 2001)이라는 가정 아래 학술지의 주제범주를 분석의 대상으로 빅데이터 분야의 학제성을 측정하고자 한다. 또한 인용을 기반으로 빅데이터 분야의 학술지 간 서지결합분석을 실시하여 네트워크상에서 매개중심성, 최근접중심성, 삼각매개중심성을 중심으로 빅데이터 분야의 학제적 구조를 파악하고자 한다.

3. 연구 방법

3.1 데이터 수집

빅데이터 분야의 학제적 특성을 파악하기 위해 본 연구에서는 톰슨 로이터(Thomson Reuters)에서 제공하는 인용색인 데이터베이스인 Web of Science(WoS)를 대상으로 2016년 5월 23일에 검색을 실시하였다. WoS 핵심컬렉션(Core Collection)의 주제(topic) 필드에서 “big data”를 질의어로 사용하였으며, 기간(timespan)은 특정기간으로 제한하지 않고 모든 범위(all years)로 설정하였다. 또한 문서유형은 학술논문(article)으로 언어는 영어(english)로 제한한 결과 총 2,681편의 논문이 검색되었다.

WoS로부터 각 논문의 서지 데이터와 참고문헌 데이터를 다운로드 받았다. 서지데이터 분석 프로그램인 Bibexcel을 사용하여 학술지의 주제범주 분석에 필요한 각 논문의 학술지명과 연구분야(Subject Categories)를 추출하였다. 그리고 서지결합분석을 위해 각 논문의 학술지명과 연구분야 외에도 Web of Science 범주와 참고문헌의 학술지명을 추출하였다.

3.2 분석 방법

빅데이터 분야의 학제성을 분석하기 위하여 본 연구에서는 두 가지 분석 방법을 시도하였다. 첫째, 빅데이터를 연구주제로 다루는 학술지에 부여된 주제범주를 분석하여 빅데이터 분야가 다루어지는 분야와 분야 간 연관성을 파악하고자 하였다. 둘째, 빅데이터 분야의 학제적 구조를 알아보기 위해 인용분석 기법 중 하

나인 서지결합분석을 학술지 단위로 실시하고 이를 기반으로 생성된 네트워크상에서 학제성을 측정하는 매개중심성과 지역중심성을 측정하는 최근접중심성, 전역중심성을 측정하는 삼각매개중심성을 구하여 그 의미를 해석하고자 하였다.

빅데이터 관련 학술지의 주제범주 분석을 위해 WoS로부터 검색된 논문 2,681편의 게재 학술지 1,083종에 WoS에서 부여한 주제범주를 추출하였다. WoS는 연구분야를 나타내는 SC (Subject Categories) 주제범주와 하위 주제분야를 나타내는 WC(Web of Science Categories)라는 두 가지 주제범주를 제공하는데, 학술지의 주제범주 분석에서는 둘 중 더 포괄적인 주제범주를 나타내는 SC를 선택하여 분석을 실시하였다. 검색된 학술지에는 총 130개의 고유한 SC가 부여되었다. 또한 SC를 대분류로 재분류하여 전체 학문영역에서 관련 학술지들이 어떤 주제영역에 분포하는지 파악하고자 하였다. 이를 위해 WoS가 2012년에 SC를 OECD의 과학기술분야 주제분류(Field of Science and Technology Classification)에 대입하여 작성한 분류표를 사용하였다. OECD 과학기술분야 주제분류는 OECD가 전체 학문영역을 6개의 대분류와 42개의 소분류로 구분한 분류체계로 OECD 회원국 외에도 많은 국가에서 활용되고 있으며, WoS의 주제분류와 달리 특정 데이터베이스에 제한되지 않고 보편적으로 활용 가능한 주제분류이다(Bartol, Budimir, Dekleva-Smrekar, Pusnik, & Juznic, 2014). 본 연구에서는 SC를 OECD 대분류인 농업과학, 공학/기술, 인문학, 의학/보건학, 자연과학, 사회과학 등 6개 영역으로 분류하였다. 이를 바탕으로 각 학

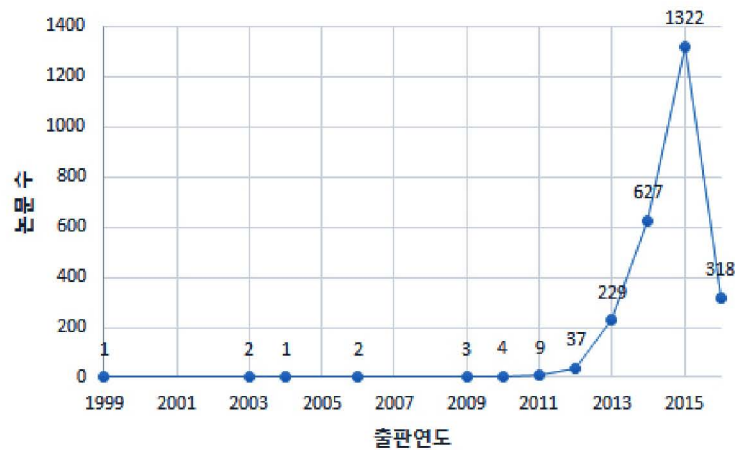
술지에 부여된 SC의 빈도를 파악하고 이 중 SC가 복수로 부여된 학술지를 대상으로 학술지 당 SC가 동일 OECD 영역 내에서 부여된 비율 및 영역 외에서 부여된 비율, 영역 내외 모두에서 부여된 비율을 산출하여 비교하였다.

빅데이터 분야의 학제적 구조 분석을 위한 네트워크를 생성하기 위해서 학술지 서지결합 분석을 실시하였다. 네트워크에서 표현하기 적합한 노드의 수를 결정하기 위해 각 학술지에서 빅데이터 관련 논문의 게재 빈도수를 계산하여 순위를 집계한 후 그 중 게재 빈도 9회 이상인 상위 46위(동순위 인정)에 해당하는 학술지 57종을 분석의 대상으로 선정하였다. 분석 대상인 57종의 학술지 목록은 [부록]에 첨부하였다. 해당 학술지 간 서지결합행렬을 이재운의 COOC 프로그램(ver. 0.4)에 입력한 결과 서지결합이 나타나지 않은 1종의 학술지(Harvard Business Review)를 제외한 56종 학술지의 서지결합빈도 행렬과 코사인 연관성 행렬(56 × 56)이 생성되었다. 본 연구에서는 네트워크 분석·시각화 소프트웨어인 NodeXL을 이용하여 코사인 연관성 행렬의 네트워크를 생성하고 각 노드의 매개중심성을 측정하였으며 이재운의 WNET 프로그램(ver. 0.4.1)을 이용하여 최근접중심성과 삼각매개중심성을 측정하였다.

4. 연구 결과

4.1 빅데이터 분야 개요

본 연구에서 수집한 총 2,681편의 논문에 대한 출판연도별 게재 빈도 분포는 <그림 1>과 같



〈그림 1〉 출판연도에 따른 빅데이터 관련 논문 게재 빈도

다. 처음으로 관련 논문이 출현한 시기는 1999년이며, 그 때부터 현재까지 17년 동안의 빅데이터 분야의 논문을 살펴보면 2012년부터 급격한 증가추세가 나타남을 확인할 수 있다. 1999년부터 2011년까지 한자리 수로 나타나던 논문 빈도수가 2012년 37편, 2013년 229편, 2014년 627편으로 증가하였고 2015년에는 1,322편으로 가장 높게 나타났다. 이는 2012년 시작된 미국의 빅데이터 연구 지원 프로젝트인 “Big Data Research and Development Initiative”와 NSF의 막대한 연구비 지원(약 45백82만 달러, 한화 약 511억 원)의 영향으로 보인다(Huang et al., 2016). 2016년의 경우 2016년 5월 23일까지 Web of Science에 포함된 논문만을 집계하였으므로 2016년의 모든 논문을 집계할 2017년에는 더욱 증가한 수치가 나타날 것이라 예상된다. 따라서 빅데이터 분야에 대한 관심이 최근 3-4년간 크게 증가하여 활발한 연구가 수행되고 있음을 알 수 있다.

WoS로부터 검색된 논문 총 2,681편이 게재

된 학술지 1,083종을 논문 게재 빈도순으로 정렬하여 상위 10위에 해당하는 학술지를 〈표 1〉에 정리하였다. 상위 10위 학술지에 게재된 논문 수는 총 314편으로 전체 논문의 11.7%를 차지하며, 이 중 ‘PLOS ONE’이 47편의 논문을 게재해 가장 많은 빅데이터 관련 논문을 수록하고 있었다. 이는 ‘PLOS ONE’이 과학 및 의학 전 분야를 포괄하는 다학제적 성격을 지녀 빅데이터 관련 연구가 많이 다루어 졌기 때문이다. 해당 학술지에 부여된 모든 SC 주제범주를 살펴본 결과, 전반적인 과학기술 분야의 ‘PLOS ONE’과 생명공학 분야의 ‘BMC Bioinformatics’, 공학/지리학 분야의 ‘IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing’을 제외한 나머지 상위 학술지 모두 컴퓨터공학 분야에 해당하는 학술지들로 나타났다.

주제범주가 복수로 부여된 경우를 포함하여 산출한 결과 해당 논문들에는 총 130개의 WoS 연구분야(SC)가 4,130번 부여되어 논문 당 평

〈표 1〉 논문 게재 빈도 상위 10위 학술지

학술지명	게재논문 수
PLoS ONE	47
Future Generation Computer Systems-the International Journal of Escience	42
Big Data	41
Neurocomputing	32
Information Sciences	31
International Journal of Distributed Sensor Networks	26
IEEE Network	25
Cluster Computing-the Journal of Networks Software Tools and Applications	24
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	24
BMC Bioinformatics	22
총계	314

〈표 2〉 상위 10위 WoS 연구분야(SC)와 해당 OECD 영역

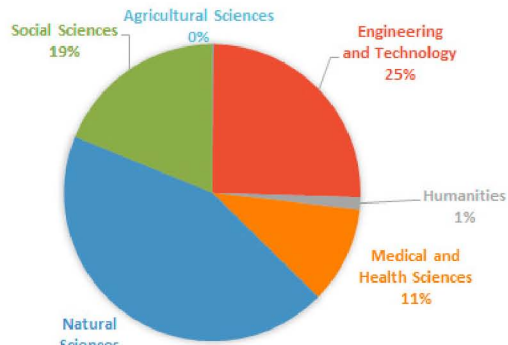
WoS 연구분야(SC)	OECD 영역	게재논문 수
Computer Science	Natural Sciences	1,120
Engineering	Engineering and Technology	442
Telecommunications	Engineering and Technology	197
Business & Economics	Social Sciences	172
Mathematics	Natural Sciences	136
Science & Technology - Other Topics	Engineering and Technology	121
Information Science & Library Science	Social Sciences	93
Environmental Sciences & Ecology	Natural Sciences	86
Biochemistry & Molecular Biology	Natural Sciences	85
Mathematical & Computational Biology	Natural Sciences	84
총계		2,536

균 약 1.5개의 SC범주가 부여되었으며, 〈표 2〉에서 보이듯이 컴퓨터공학이 전체 빈도인 4,130번 중 약 27.1%인 1,120번 부여되어 가장 많이 부여된 주제범주임을 알 수 있었다. 그 뒤로 공학, 통신, 경영/경제학, 수학, 과학기술, 문헌정보학, 환경학, 생화학/분자생물학, 수리생물학/전산생물학이 빅데이터 분야의 주요 주제범주로서 나타났다. 또한 WoS 연구분야(SC)를 OECD

과학기술분야 주제분류에 대입하여 대분류로 구분한 결과, 〈표 3〉과 〈그림 2〉와 같이 나타났다. 가장 높은 비율을 차지한 자연과학의 경우 WoS의 SC범주인 컴퓨터공학이 자연과학으로 분류되었기 때문으로 해석할 수 있으며, 공학, 통신 등이 포함된 공학/기술과 경영/경제학, 문헌정보학 등이 포함된 사회과학이 그 뒤를 이어 많이 나타났다.

〈표 3〉 OECD 영역 분석

OECD 영역	게재논문 수	%
Natural Sciences	1,809	43.80
Engineering and Technology	1,046	25.33
Social Sciences	778	18.84
Medical and Health Sciences	435	10.53
Humanities	56	1.36
Agricultural Sciences	6	0.15
총계	4,130	100



〈그림 2〉 OECD 영역 비율

4.2 주제범주에 따른 학제성 분석

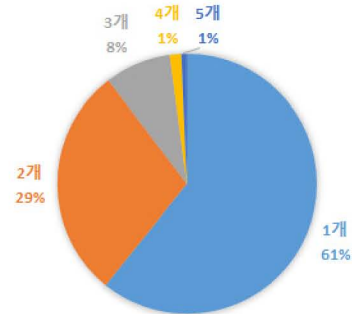
빅데이터 분야의 학제적 특성을 파악하기 위해 관련 학술지의 주제범주 분석을 실시하였다. 하나의 주제범주가 부여된 학술지보다 둘 이상의 주제범주가 부여된 학술지가 더 학제적인 성격을 띤다는(Morillo, Bordons, & Gómez, 2001) 가정에 따라 빅데이터를 다룬 학술지 총 1,083종에 부여된 WoS SC범주 130개가 각 학술지에 나타난 빈도를 산출하였다. 그리고 둘 이상의 SC범주가 부여된 학술지를 대상으로 동일한 OECD 영역에 해당하는 SC범주가 부여되었는지 동일하지 않은 OECD 영역에 해당하는 SC범주가 부여되었는지를 분석하여 학술지 내 주제범주의 다양성을 살펴보았다.

분석 결과 단일 SC범주가 부여된 학술지는 전체 1,083종 중 663종으로 전체의 61.2%를 차지했으며, SC범주가 2개 부여된 학술지는 310종, 3개 부여된 학술지는 88종, 4개 부여된 학술지는 317종, 5개 부여된 학술지는 5종으로 각각 28.6%, 8.1%, 1.6%, 0.5% 순으로 나타났다

(〈표 4〉, 〈그림 3〉 참조). 개별 학술지가 지닌 다양한 학문적 성격의 정도를 알아보기 위해 복수의 SC범주가 부여된 학술지 420종을 대상으로 개별 학술지의 SC범주들이 동일한 OECD 영역에 속하는 비율, 동일하지 않은 영역에 속하는 비율 그리고 동일한 영역과 동일하지 않은 영역 모두에 속하는 비율을 측정하였다. 〈표 5〉와 〈그림 4〉에서 볼 수 있듯이 420종 중 239종의 학술지가 서로 다른 OECD 영역에 해당하는 SC범주가 부여되었으며 84종의 학술지는 동일한 영역 및 동일하지 않은 영역 모두의 SC범주가 부여되어 동일 영역의 SC범주만이 부여된 97종에 비해 3배 이상의 학술지가 다양한 학문적 성격을 가지고 있음을 확인할 수 있었다. 다시 말해 빅데이터 분야는 30%에 해당하는 학술지가 다양한 학문적 성격의 주제범주를 포함하고 있어 학제적 성격이 비교적 잘 나타나는 분야라고 할 수 있다.

〈표 4〉 학술지 당 SC범주 빈도

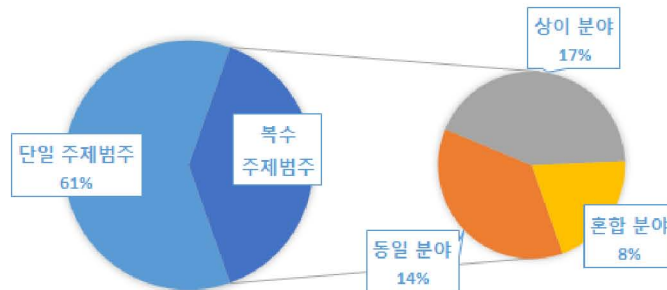
SC범주	학술지 수
1개	663
2개	310
3개	88
4개	17
5개	5
전체	1,083



〈그림 3〉 학술지 당 SC범주 비율

〈표 5〉 개별 학술지의 OECD 분야에 따른 SC범주 빈도 및 비율

	단일 범주	복수 범주			
		전체	동일 분야	상이 분야	혼합 분야
논문 수	663	420	97	239	84
비율	61.2%	38.8%	9.0%	22.1%	7.8%



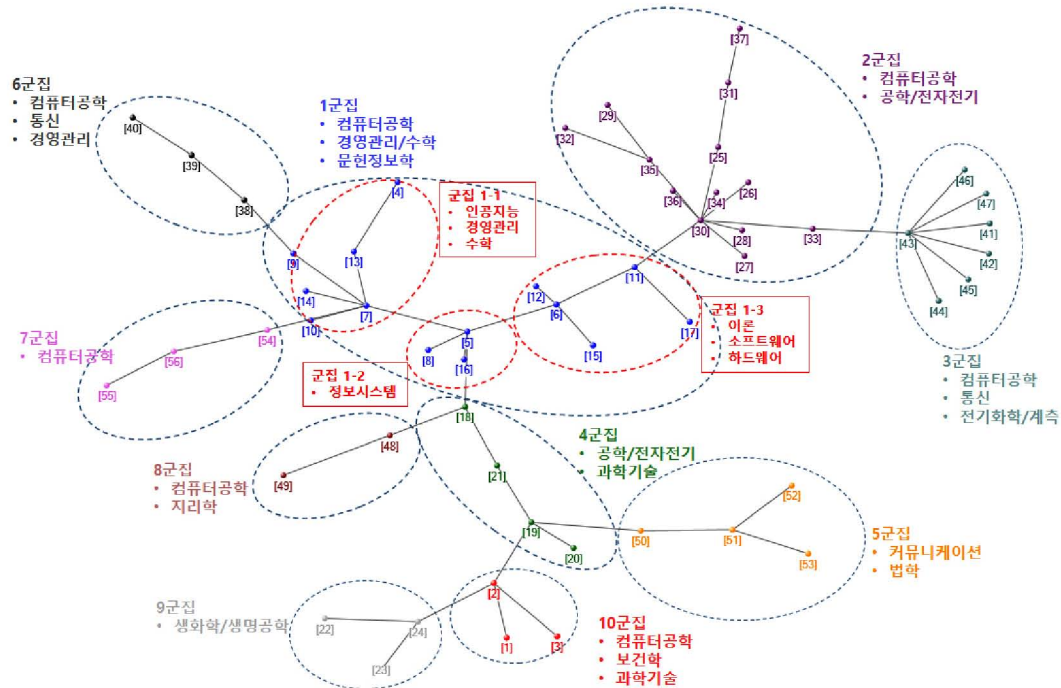
〈그림 4〉 개별 학술지의 OECD 분야에 따른 SC범주 비율

4.3 서지결합 네트워크 기반의 학제적 구조

4.3.1 서지결합 네트워크 군집 분석

본 연구에서는 빅데이터 분야의 학제적 구조를 분석하기 위해 논문 게재 빈도 기준 상위 57종의 서지결합분석을 실시하였다. 서지결합이 발생하지 않은 1종(Harvard Business Review)

을 제외한 학술지 56종 간 서지결합빈도를 이용하여 코사인 유사도 행렬을 작성하였다. 코사인 유사도 행렬에 패스파인더 네트워크 알고리즘을 적용하여 네트워크를 생성하였다. 그리고 이 재윤의 WNET 프로그램(ver. 0.4.1)을 사용하여 병렬 최근접 이웃 클러스터링(PNNC) 알고리즘을 적용한 결과 네트워크에 10개의 군집이 생성되었다. 〈그림 5〉와 같이 56개 학술지의 패



〈그림 5〉 56종 학술지 패스파인더 네트워크와 PNNC 클러스터링

스파인더 네트워크에 점선으로 PNNC 군집을 표시하였고, 해당 학술지의 모든 WoS SC범주를 기반으로 각 군집을 대표하는 주제를 부여하였다. 네트워크상의 노드는 학술지를 나타내고 각 학술지를 구분하기 위해 학술지명을 대신하여 숫자를 부여하였다. 각 숫자에 해당하는 학술지명은 [부록]에 첨부하였다.

제 1군집은 컴퓨터공학, 경영관리/수학, 문헌정보학을 다루는 총 14종의 학술지로 이루어졌으며, 제 2군집은 컴퓨터공학, 공학/전자전기 분야의 13종의 학술지로 이루어졌다. 제 3군집은 컴퓨터공학, 통신, 전기화학/계측 분야의 학술지 7종, 제 4군집은 공학/전자전기, 과학기술 분야의 학술지 4종, 제 5군집은 커뮤니케이션, 법학 분야의 학술지 4종, 제 6군집은 컴퓨터공

학, 통신, 경영관리 분야의 학술지 3종, 제 7군집은 컴퓨터공학 분야의 학술지 3종, 제 8군집은 컴퓨터공학, 지리학 분야의 학술지 2종, 제 9군집은 생화학/생명공학 분야의 학술지 3종, 제 10군집은 컴퓨터공학, 보건학, 과학기술 분야의 학술지 3종으로 구성되어 있다.

네트워크 군집 분석을 통해 컴퓨터공학이 빅데이터 분야에서 가장 많이 다루어지는 학문분야임이 나타났다. 전체 네트워크의 군집 10개 중 7개에 해당하는 군집이 컴퓨터공학 분야를 다루는 학술지들로 나타났으며 단지 3개의 군집만이 컴퓨터공학 외 분야에 해당했다. 주제가 컴퓨터공학으로 동일하게 나타난 7개의 군집의 경우 각 군집의 특징을 파악하기 어려워 SC범주보다 세부적인 주제를 보여주는 WC범

주를 통해 세부 주제를 파악하고자 하였다. 특히 네트워크의 중심에 위치하여 가장 큰 군집을 이루고 있는 제 1군집의 경우 군집 내 학술지의 WC범주로 세부 주제를 파악한 결과 <그림 5>와 같이 3개의 세부 군집으로 구분할 수 있었다. 가장 중심에 위치하는 군집 1-2는 컴퓨터공학 분야의 정보시스템을 다루는 학술지들로 이루어져 있었고, 왼쪽에 위치한 군집 1-1은 인공지능과 경영관리, 수학에 해당하는 학술지들로 이루어져 있었으며, 오른쪽에 위치한 군집 1-3은 이론, 소프트웨어, 하드웨어와 관련한 학술지들로 이루어져 있었다. 이들 제 1군집의 하위 군집들을 중심으로 네트워크의 왼쪽 상단에서 하위 군집 1-1과 연결된 제 6군집과 제 7군집은 인공지능이 그 세부 주제로 나타났다. 하위 군집 1-3과 연결된 네트워크 오른쪽 상단의 제 2군집과 제 3군집은 소프트웨어, 하드웨어, 정보시스템이 세부 주제로 나타났다. 다만 정보시스템과 관련한 하위 군집 1-1과 연결된 네트워크 하단에는 컴퓨터공학 분야에서 지리 시스템을 다루는 제 8군집과 전반적인 과학기술 분야를 다루는 제 10군집과 함께 컴퓨터공학 외 분야의 제 4군집, 제 5군집, 제 9군집 등의 군집들이 나타났다.

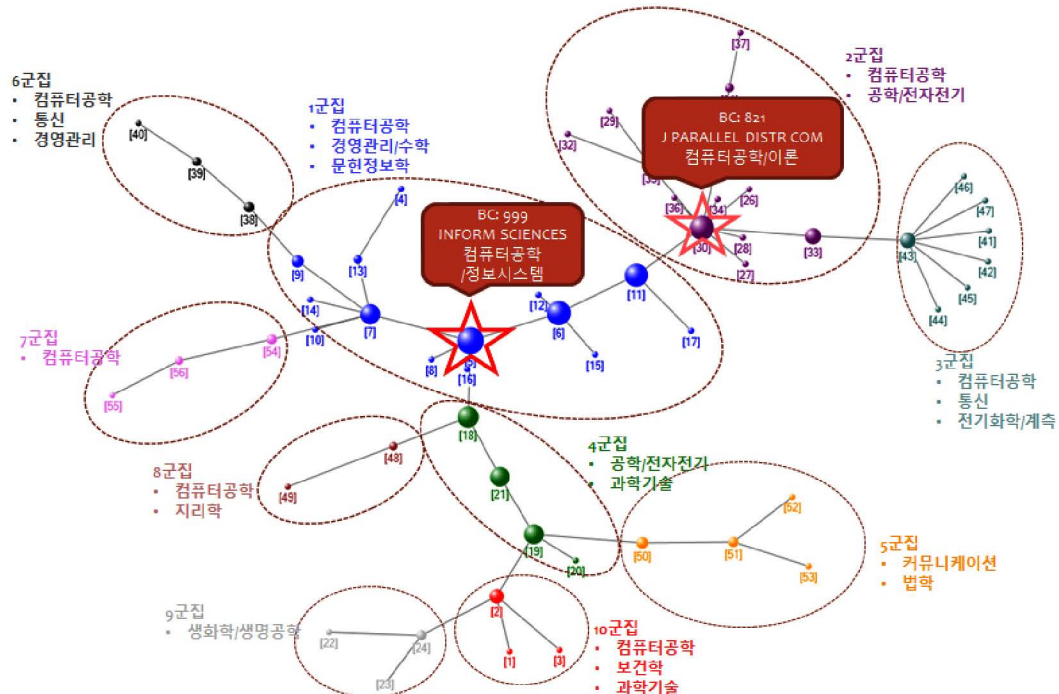
이러한 네트워크 군집의 특성을 살펴보면 빅데이터 분야는 정보시스템, 인공지능, 이론, 소프트웨어 등 컴퓨터공학 분야에서 가장 많은 연구가 이루어짐을 알 수 있다. 이는 빅데이터의 저장, 처리, 분석 등 빅데이터를 다루는 기술적인 연구가 주를 이루고 있기 때문으로 보인다. 특히 분석의 기간이 빅데이터 개념이 처음 등장한 1999년 이래 모든 기간을 포함하였기 때문에 초기에 활발히 수행된 빅데이터 관련 기술 및

도구 개발 연구가 많은 부분을 차지한 것으로 추측할 수 있다. 하지만 네트워크의 하단에 나타난 제 4군집, 제 5군집, 제 8군집, 제 9군집, 제 10군집을 보면 빅데이터를 처리하는 기술적 연구를 넘어서 과학기술, 공학, 커뮤니케이션, 법학, 지리학, 생화학/생명공학, 보건학 등 다양한 분야에 빅데이터를 분석하고 활용하는 연구가 이루어지고 있는 것도 확인할 수 있다.

4.3.2 중심성 분석

본 연구에서는 빅데이터 분야의 중심성 분석을 통해 네트워크상에서 중요한 영향력을 갖고 있는 학술지들을 알아보았다. 우선, 학제성을 측정하는 유용한 지수로서 입증된 네트워크의 매개중심성(Leydesdorff, 2007)을 측정하여 매개중심성이 높은 학술지를 확인하였다. <그림 6>과 같이 매개중심성이 가장 높은 학술지를 별 모양으로 표시하였다. 매개중심성이 가장 높은 학술지는 'Information Sciences'로 컴퓨터공학 분야에서 정보시스템을 다루는 학술지이다. 'Information Sciences'는 네트워크의 정중앙에 위치하여 왼쪽에 위치한 컴퓨터공학 분야의 인공지능과 오른쪽에 위치한 컴퓨터공학 분야의 이론, 소프트웨어, 하드웨어 등 컴퓨터공학 분야의 세부 주제영역을 연결하고 하단에 위치한 과학기술, 공학, 커뮤니케이션, 법학, 지리학, 생화학/생명공학, 보건학 등 다양한 분야를 연결하는 중개자 역할을 수행하고 있다.

두 번째로 매개중심성이 큰 학술지는 'Journal of Parallel and Distributed Computing'으로 제 2군집에 위치하여 가장 큰 군집인 제 1군집과 제 3군집을 연결하고 있다. 'Journal of Parallel and Distributed Computing'은 컴퓨터공학의 이론



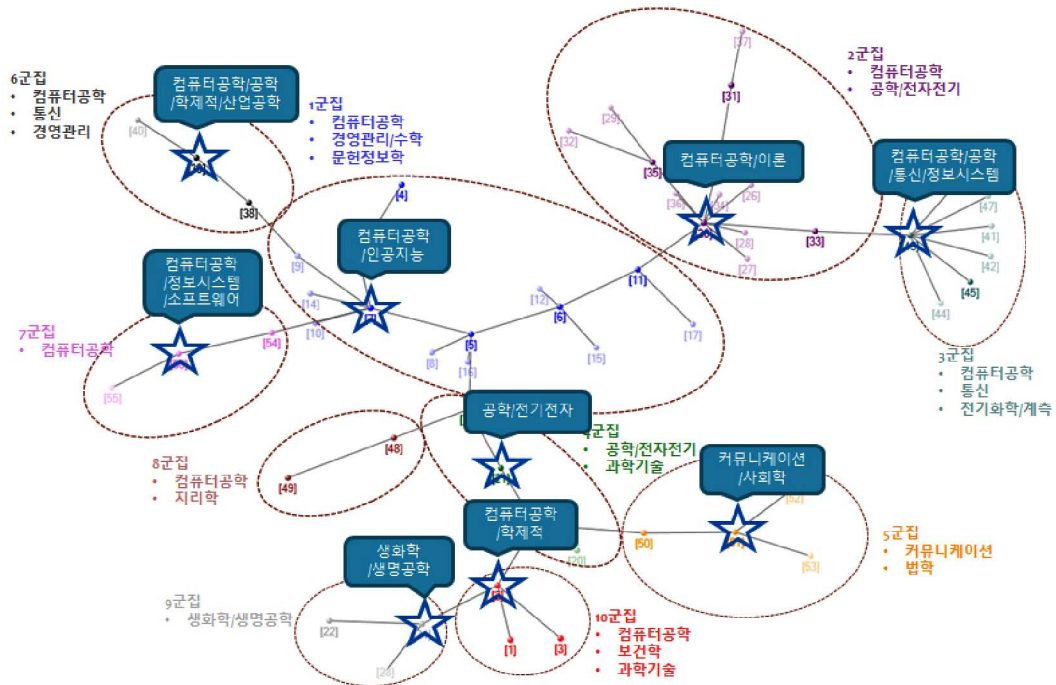
〈그림 6〉 빅데이터 분야 네트워크의 매개중심성

을 다루는 학술지로 컴퓨터공학 분야를 전반적으로 아우르는 제 1군집과 컴퓨터공학과 통신 분야의 제 3군집의 중개자 역할을 하고 있다. 'Journal of Parallel and Distributed Computing'의 매개중심성이 크게 나타나는 이유는 가장 많은 학술지가 포함되어 있는 제 1-3군집의 중심에 위치하여 세 군집을 연결해주기 때문이다.

본 연구에서 생성된 네트워크는 학술지의 서지결합빈도를 반영한 가중네트워크이므로 가중네트워크 중심성 척도인(이재운, 2006) 최근접 중심성과 삼각매개중심성을 WNET 프로그램(ver. 0.4.1)을 사용하여 측정하였다. 우선 지역 중심성을 나타내는 최근접중심성을 측정하여 네트워크의 군집별 대표 학술지와 주제분야를 살펴보았다. 〈그림 7〉에서는 최근접중심성의

확인용이하게 하도록 각 군집에서 상대적 최근접중심성이 가장 높은 학술지를 별모양으로 표시하고 최근접중심성의 값이 없는 학술지는 투명도를 조절하여 흐리게 표현하였다. 또한 두 학술지가 동일한 최근접중심성 값을 가지는 제 8군집은 제외하였다.

측정 결과 군집별로 최근접중심성이 가장 높은 학술지는 제 8군집을 제외한 제 1군집부터 제 10군집까지 순서대로 'Knowledge-Based Systems', 'Journal of Parallel and Distributed Computing', 'IEEE Networks', 'IEEE Journal of Selected Topics in Signal Processing', 'Information Communication & Society', 'Industrial Management & Data Systems', 'Sigmod Record', 'Bioinformatics', 'Big Data' 순으로

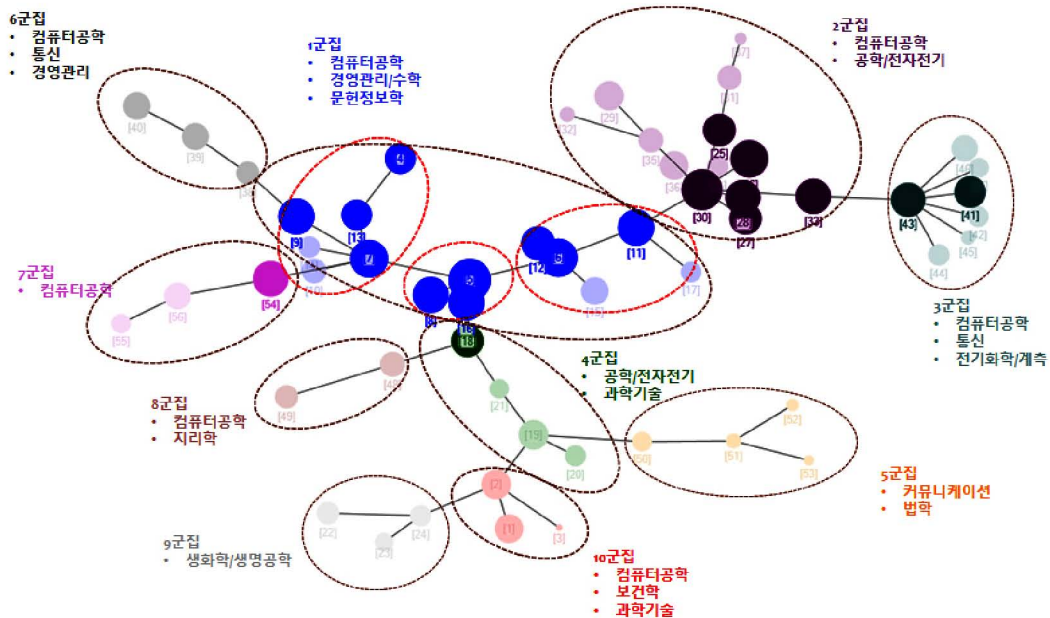


〈그림 7〉 빅데이터 분야 네트워크의 최근접중심성

나타났다. 각 군집에서 최근접중심성이 가장 높게 나타난 학술지의 주제범주와 군집을 대표하는 주제를 비교한 결과 대부분 일치하는 것을 확인하였다.

가중네트워크에서 특정 노드가 다른 노드들을 결속시키는 정도를 나타내는 전역중심성인 삼각매개중심성을 측정하여 전체 네트워크에 큰 영향을 미치는 학술지를 확인하였다. 〈그림 8〉에서 보이듯이 각 학술지의 삼각매개중심성을 측정하여 상대적 삼각매개중심성 값에 따라 노드의 크기에 반영하고 삼각매개중심성이 높은 노드들을 강조하기 위해 상대적 삼각매개중심성 값이 높은 순서대로 상위 20개의 노드를 제외한 나머지 노드의 투명도를 높여 흐리게 나타냈다. 네트워크 전체에서 제 1군집에 속하

는 대부분의 노드들이 삼각매개중심성이 높은 것으로 나타났으며, 제 2군집과 제 3군집에 속하는 일부 노드들도 삼각매개중심성이 높은 것으로 나타났다. 따라서 제 1군집의 컴퓨터공학 학술지들과 제 2군집의 컴퓨터공학 이론, 소프트웨어, 하드웨어 학술지들 그리고 제 3군집의 컴퓨터공학, 통신 관련 학술지들이 전체 네트워크에 미치는 영향력 및 주제적 연관성이 강한 것으로 파악되었다. 이외 나머지 군집 중 특히 커뮤니케이션과 법학 분야의 제 5군집과 생화학, 생명공학 분야의 제 9군집에 속하는 학술지들의 삼각매개중심성이 낮게 나타나 이들 분야가 빅데이터 분야에 미치는 영향력이 크지 않음을 확인하였다.



〈그림 8〉 빅데이터 분야 네트워크의 삼각매개중심성

5. 결론

본 연구는 최근 국내의 학계, 산업, 언론, 정부 등 다양한 영역에서 관심을 받고 있는 빅데이터 분야의 학제성을 규명하고 학제적 지적구조를 파악하고자 하였다. 이를 위해 빅데이터 관련 학술지의 WoS SC주제범주를 분석하여 빅데이터가 얼마나 다양한 주제영역에서 다루어지는지를 확인하였다. 또한 학술지 간 서지결합분석을 통해 빅데이터 분야의 패스파인더 네트워크를 생성하고 군집 분석을 실시해 빅데이터를 연구하고 있는 관련 주제분야를 파악하였다. 또한 매개 중심성, 최근접중심성, 삼각매개중심성 등 네트워크의 중심성 척도를 측정하여 빅데이터 분야에서 큰 영향력을 지닌 학술지를 통해 세부주제 분야의 영향력을 규명하였다.

빅데이터 관련 학술지의 WoS 주제범주를 분석한 결과 전체 1,083종의 학술지에 총 130개의 고유한 SC범주가 부여되었으며, 이 중 38.8%의 학술지 420종이 둘 이상의 SC범주를 부여받았음을 확인하였다. 그리고 22.1%에 해당하는 학술지 239종에 부여된 SC범주들은 서로 다른 OECD 학문영역에 속하였고 7.8%의 84종 학술지는 동일하거나 상이한 OECD 학문영역에 해당하는 SC범주를 부여받았다. 따라서 전체의 약 30% 이르는 빅데이터 분야 개별 학술지들에 다양한 대분류 학문영역의 주제분야가 동시에 부여되어 있어, 빅데이터 분야가 학제적 성격을 띠고 있음을 확인할 수 있다.

빅데이터 관련 학술지 56종을 대상으로 서지결합분석을 통해 패스파인더 네트워크를 생성하고 병렬 최근접 이웃 클러스터링을 적용한

결과 총 10개의 군집이 나타났다. 군집 내 학술지들의 WoS SC범주를 기반으로 군집을 대표하는 주제를 표시하였다. 10개의 군집 중 7개의 군집이 컴퓨터공학 분야에 해당하였고 나머지 3개 군집만이 공학과 과학기술, 커뮤니케이션과 법학, 생화학/생명공학 등 컴퓨터공학 외 분야에 해당하였다. 컴퓨터공학 분야로 나타난 군집의 세부 주제를 파악하기 위해 7개 군집 내 학술지의 WC범주를 확인한 결과 정보시스템, 인공지능, 이론, 소프트웨어, 하드웨어 등의 주제가 나타났다. 이러한 군집의 특성에 따라 빅데이터 분야는 정보시스템, 인공지능, 이론, 소프트웨어 등 컴퓨터공학 분야에서 가장 많은 연구가 이루어짐을 알 수 있다. 이는 빅데이터의 저장, 처리, 분석 등 빅데이터를 다루는 기술적인 연구가 주를 이루고 있기 때문으로 보인다. 제 4, 5, 8, 9, 10군집을 보면 빅데이터를 처리하는 기술적 연구 외에도 과학기술, 공학, 커뮤니케이션, 법학, 지리학, 생화학/생명공학, 보건학 등 다양한 분야에 빅데이터를 분석하고 활용하는 연구가 이루어지고 있음을 확인할 수 있다.

빅데이터 분야 네트워크에서 영향력이 큰 학술지를 확인하기 위하여 다양한 중심성을 측정하였다. 네트워크에서 노드 간 중개 역할을 하는 노드를 보여주는 매개중심성의 경우 'Information Sciences'와 'Journal of Parallel and Distributed Computing'이 가장 높게 나타났다. 'Information Sciences'는 컴퓨터공학 분야의 정보시스템을 다루는 학술지로 정보시스템이 네트워크의 다른 노드들을 연결하는 역할을 하고 있음을 보여준다. 'Journal of Parallel and Distributed Computing'은 컴퓨터공학 분야의 이론을 다루

는 학술지로 가장 규모가 큰 제 1, 2, 3군집 사이를 연결하며 컴퓨터공학 분야 내 정보시스템, 인공지능, 통신 등을 연결하는 역할을 하고 있음을 보여주었다. 그리고 각 군집 내에서 가장 영향력이 높은 노드를 보여주는 최근접중심성이 가장 높게 나타난 학술지의 주제범주와 군집을 대표하는 주제를 비교한 결과 대부분 일치하는 것을 확인하였다. 따라서 컴퓨터공학 분야에서 빅데이터를 가장 많이 연구해왔으며 타 분야 중에는 과학기술 전반, 공학, 커뮤니케이션, 법학, 생화학/생명공학 등의 분야에서 연구가 수행되고 있음을 확인할 수 있다. 마지막으로 네트워크 전반에 큰 영향력을 갖는 노드를 나타내는 삼각매개중심성의 측정 결과 삼각매개중심성이 높은 학술지는 제 1, 2, 3군집에 모여 있었다. 이는 해당 군집들에 속한 컴퓨터공학 전반, 이론, 소프트웨어, 하드웨어, 통신 관련 학술지들이 전체 네트워크에 미치는 영향력 및 주제적 연관성이 강한 것으로 설명될 수 있다. 이외에 커뮤니케이션과 법학, 생화학/생명공학 분야 학술지들의 삼각매개중심성은 비교적 낮게 나타나 이들 분야가 빅데이터 분야에 미치는 영향력이 크지 않음을 확인하였다.

본 연구는 빅데이터가 최근에 등장한 주제임을 고려하여 관련 학술지들을 대상으로 서지결합분석을 실시하였다. 하지만 저자서지결합분석, 동시인용분석, 단어동시출현분석, 공저자분석 등 다양한 계량정보적 분석을 실시하고 각 분석결과를 비교하여 그 차이점과 공통점을 도출한다면 더욱 의미 있는 결과를 얻을 수 있을 것으로 예상된다. 또한 본 연구에서는 빅데이터 연구가 시작된 1999년부터 현재까지의 모든 관련 학술지를 전체적으로 분석하였으나 이를

더욱 세분화하여 시간의 추이에 따른 빅데이터 분야의 학제성 변화양상을 파악하는 것도 의미가 있을 것이다. 마지막으로 본 연구는 학술지에 WoS에서 부여한 SC범주와 WC범주를 적용하여 학술지 및 관련 군집의 주제적 성격을

규정하였으나 학술지의 실질적인 주제를 제대로 반영하지 못하는 사례가 나타나는 한계를 지녔다. 따라서 학술지 단위보다 더욱 세밀하게 문헌 단위로 분석하는 것도 의미가 있을 것이다.

참 고 문 헌

- 김대현, 강이화 (2010). 고등교육에서 학제성의 개념과 유형에 관한 고찰. *교육사상연구*, 24(3), 31-46.
- 김민지, 박정규, 이유아, 허은영 (2011). Co-Classification 방법을 이용한 태양전지 연구의 학제간 다양성 분석. *신재생에너지*, 7(1), 36-44. <http://dx.doi.org/10.7849/ksnre.2011.7.1.036>
- 김완중 (2014). 동시출현 단어분석을 활용한 빅데이터 관련 연구동향 분석. *한국정보관리학회 학술대회 논문집*, 17-20.
- 김현영, 지현수, 이화순, 지종덕 (2014). 빅 데이터에 따른 지적정보의 효율화 방안 연구. *한국지적정보학회지*, 16(1), 29-48.
- 이재윤. COOC. (Version 0.4) [Computer Software].
- 이재윤. WNET. (Version 0.4.1) [Computer Software].
- 이재윤 (2006). 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. *한국문헌정보학회지*, 40(3), 191-214. <http://dx.doi.org/10.4275/kslis.2006.40.3.191>
- 이재윤, 정주희 (2006). 연구자 소속과 표제어 분석을 통한 국내 인지과학 분야의 학제적 구조 파악. 제 13회 *한국정보관리학회 학술대회 논문집*, 127-134.
- 이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. *한국비블리아학회지*, 24(4), 53-73. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.053>
- 정은경 (2011). 디지털도서관 분야의 학제적 공동연구 분석에 관한 연구. *정보관리학회지*, 28(2), 37-51. <http://dx.doi.org/10.3743/kosim.2011.28.2.037>
- 정연경 (2012). 국내 기록관리학 분야 학술지에 나타난 학제성 연구. *한국기록관리학회지*, 12(2), 7-27.
- Bartol, T., Budimir, G., Dekleva-Smrekar, D., Pusnik, M., & Juznic, P. (2014). Assessment of research fields in Scopus and Web of Science in the view of national research evaluation in Slovenia. *Scientometrics*, 98(2), 1491-1504. <http://dx.doi.org/10.1007/s11192-013-1148-8>
- Gartner (2012). Gartner Identifies the Top 10 Strategic Technology Trends for 2013. Retrieved

- from <http://www.gartner.com/newsroom/id/2209615>
- Hargens, L. L. (1986). Migration patterns of U.S. Ph.D.s among disciplines and specialties. *Scientometrics*, 9(3-4), 145-164. <http://dx.doi.org/10.1007/bf02017238>
- Huang, Y., Zhang, Y., Youtie, J., Porter, A. L., & Wang, X. (2016). How does national scientific funding support emerging interdisciplinary research: A comparison study of big data research in the US and China. *PLoS ONE*, 11(5), e0154509. <http://dx.doi.org/10.1371/journal.pone.0154509>
- Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303-1319. <http://dx.doi.org/10.1002/asi.20614>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-66.
- Morillo, F., Bordons, M., & Gomez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, 51(1), 203-222.
- Morillo, F., Bordons, M., & Gomez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237-1249.
- National Academies Committee on Facilitating Interdisciplinary Research, Committee on Science, Engineering, Public Policy (COSEPUP) (2005). *Facilitating interdisciplinary research*. Washington: National Academies Press.
- OECD (1998). *Interdisciplinarity in Science and Technology*. T. Directorate for Science, and Industry. Paris: OECD. Quoted in: Morillo, F., Bordons, M., & Gomez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237-1249.
- Park, H. W., & Leydesdorff, L. (2013). Decomposing social and semantic networks in emerging "big data" research. *Journal of Informetrics*, 7(3), 756-765. <http://dx.doi.org/10.1016/j.joi.2013.05.004>
- Qin, J., Lancaster, F. W., & Allen, B. (1997). Types and levels of collaboration in interdisciplinary

- research in the sciences. *Journal of the American Society for Information Science*, 48(10), 893-916. [http://dx.doi.org/10.1002/\(sici\)1097-4571\(199710\)48:10<893::aid-asi5>3.0.co;2-x](http://dx.doi.org/10.1002/(sici)1097-4571(199710)48:10<893::aid-asi5>3.0.co;2-x)
- Rousseau, R. (2012). A view on big data and its relation to Informetrics. *Chinese Journal of Library and Information Science*, 5(3), 12-26.
- Shiri, A. (2014). Making sense of big data: A facet analysis approach. *Knowledge Organization*, 41(5), 357-368.
- Small, H. (2010). Maps of science as interdisciplinary discourse: Co-citation contexts and the role of analogy. *Scientometrics*, 83(3), 835-849.
<http://dx.doi.org/10.1007/s11192-009-0121-z>
- Steele, T. W., & Stier, J. C. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science*, 51(5), 476-484.
[http://dx.doi.org/10.1002/\(sici\)1097-4571\(2000\)51:5<476::aid-asi8>3.0.co;2-g](http://dx.doi.org/10.1002/(sici)1097-4571(2000)51:5<476::aid-asi8>3.0.co;2-g)
- Urata, H. (1990). Information flows among academic disciplines in Japan. *Scientometrics*, 18(3-4), 309-319. <http://dx.doi.org/10.1007/bf02017767>
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14-26.
<http://dx.doi.org/10.1016/j.joi.2010.06.004>
- Yang, R. (2013). Bibliometrical analysis on the big data research in China. *Journal of Digital Information Management*, 11(6), 383-390.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

- Chung, Eun-Kyung (2011). Interdisciplinary collaborations in the domain of digital libraries. *Journal of the Korean Society for Information Management*, 28(2), 37-51.
<http://dx.doi.org/10.3743/kosim.2011.28.2.037>
- Chung, Yeon-Kyoung (2012). A study of interdisciplinarity in *Journal of Korean Society of Archives and Records Management*. *Journal of Korean Society of Archives and Records Management*, 12(2), 7-27.
- Kim, Dae-Hyun, & Kang, Ew-Ha (2010). Conceptualizations and typologies of interdisciplinarity in higher education. *Journal of Korean Educational Idea*, 24(3), 31-46.

- Kim, Hyun-Young, Ji, Hyun-Soo, Lee, Hwa-Soon, & Ji, Jong-Duck (2014). Big data point according to the study on the efficiency methods of cadastral spatial information. *Journal of the Korean Cadastre Information Association*, 16(1), 29-48.
- Kim, Min-Ji, Park, Jung-Kyu, Lee, You-Ah, & Heo, Eun-Nyeong (2011). Co-classification analysis of inter-disciplinarity on solar cell research. *Journal of the Korean Society for New and Renewable Energy*, 7(1), 36-44. <http://dx.doi.org/10.7849/ksnre.2011.7.1.036>
- Kim, Wan-Jong (2014). The research trends about the big data using co-word analysis. *Proceedings of the Korean Society for Information Management*, 17-20.
- Lee, Jae-Yun. COOC. (Version 0.4) [Computer Software].
- Lee, Jae-Yun. WNET. (Version 0.4.1) [Computer Software].
- Lee, Jae-Yun (2006). Centrality measures for bibliometric network analysis. *Journal of the Korean Society for Library and Information Science*, 40(3), 191-214. <http://dx.doi.org/10.4275/kslis.2006.40.3.191>
- Lee, Jae-Yun, & Jung, Ju-Hee (2006). Examining the interdisciplinary structure of Korean cognitive science through analyzing author affiliations and title words. *Proceedings of the 13th Korean Society for Information Management*, 127-134.
- Lee, Jeong-Mee (2013). Understanding big data and utilizing its analysis into library and information services. *Journal of the Korean Biblia Society for Library and Information Science*, 24(4), 53-73. <http://dx.doi.org/10.14699/kbiblia.2013.24.4.053>

[부록] 서지결합 분석 대상 57종 학술지와 해당 네트워크 번호

번호	학술지명	N #	번호	학술지명	N #
1	PLoS ONE	[1]	26	International Journal of Communication	[53]
2	Future Generation Computer Systems-the International Journal of Escience	[6]	26	Journal of Supercomputing	[11]
3	Big Data	[2]	32	AI Magazine	[36]
4	Neurocomputing	[4]	32	Computer Law & Security Review	[52]
5	Information Sciences	[5]	32	IEEE Transactions on Services Computing	[33]
6	International Journal of Distributed Sensor Networks	[41]	32	Information Communication & Society	[51]
7	IEEE Network	[43]	32	Proceedings of the IEEE	[19]
8	Cluster Computing-the Journal of Networks Software Tools and Applications	[25]	32	Tsinghua Science and Technology	[29]
8	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	[18]	38	Annals of the American Academy of Political and Social Science	[50]
10	BMC Bioinformatics	[22]	38	Fujitsu Scientific & Technical Journal	[37]
11	Computer	[55]	38	Harvard Business Review	-
12	Concurrency and Computation-Practice & Experience	[12]	38	ISPRS International Journal of Geo-Information	[49]
12	IEEE Access	[8]	38	Journal of Internet Technology	[45]
14	IBM Journal of Research and Development	[27]	38	Journal of Parallel and Distributed Computing	[30]
15	Knowledge-Based Systems	[7]	38	Software-Practice & Experience	[31]
16	Expert Systems with Applications	[9]	38	Transportation Research Part C-emerging Technologies	[14]
16	Mathematical Problems in Engineering	[13]	46	ACM Sigplan Notices	[32]
18	IEEE Transactions on Knowledge and Data Engineering	[26]	46	AD Hoc Networks	[44]
19	Biomed Research International	[23]	46	China Communications	[46]
19	Health Affairs	[3]	46	Frontiers of Computer Science	[35]
19	IEEE Journal of Selected Topics in Signal Processing	[21]	46	Industrial Management & Data Systems	[39]
19	IEICE Transactions on Information and Systems	[34]	46	International Journal of Geographical Information Science	[48]
19	International Journal of Advanced Computer Science and Applications	[17]	46	IT Professional	[47]
19	Journal of Computer Science and Technology	[28]	46	Knowledge and Information Systems	[54]
19	Proceedings of the National Academy of Sciences of the United States of America	[20]	46	KSII Transactions on Internet and Information Systems	[40]
26	Big Data Research	[16]	46	Scientometrics	[10]
26	Bioinformatics	[24]	46	Sensors	[42]
26	Decision Support Systems	[38]	46	Sigmod Record	[56]
26	IEEE Transactions on Parallel and Distributed Systems	[15]			