

시간적 공저 네트워크의 구조 분석에 관한 연구*

An Analysis on the Structure of Temporal Co-Authorship Networks

서선경 (SunKyung Seo)**

초 록

공저 네트워크에서는 문헌의 출판연도 정보로부터 시간이 지남에 따라 공저 관계가 형성되고 해체되는 것을 확인하여 시간적 네트워크를 구축할 수 있다. 이에 본 연구의 목적은 계량정보학 측면에서 시간적 네트워크 모델링을 목표로 데이터논문과 학술논문의 전체적인 연구 협업 네트워크를 진화적 관점으로 분석하고, 시간적 공저 네트워크의 구조적 특성을 규명하는 것이다. 이를 위해 데이터 출판에 위한 새로운 기제를 제안한 생물다양성 분야의 혼합형 데이터학술지 Biodiversity Data Journal에 게재된 공저 데이터논문 247건, 공저 학술논문 638건을 대상으로 시간적 공저 네트워크 분석을 수행하였다. 그 결과, 두 문헌 유형 모두, 네트워크가 진화하는 동안 무척도 네트워크 속성과 좁은 세상 현상이 나타났고, 국지적으로 응집된 저자들의 관계 구조가 시간이 경과 되어도 지속되고 있음을 파악했다. 또한, 본 연구에서 개발한 시간적 매개중심성 척도인 TTBC(Temporal Triadic Betweenness Centrality, 시간적 삼각매개중심성)를 적용하여 시간에 따른 중요한 저자들이 시점별로 확인되었고, 그 변화 추이를 파악할 수 있었다. 동적 접근으로 시각화 한 결과, 데이터논문과 학술논문의 시간적 공저 네트워크에서 드러나는 구조적 차이를 확인했으며, 이는 데이터를 다루는 연구자 집단의 협업 구조에 대한 특성으로 해석할 수 있다.

ABSTRACT

In co-authorship networks, temporal networks can be modeled by identifying the formation and dissolution (linking and removing) of co-authorship relationships over time from the publication year information of the papers. Therefore, this study seeks to analyze the overall research collaboration networks of data papers and articles from an evolutionary perspective for modeling the temporal network in terms of informetrics and investigating the dynamic and structural mechanisms of the temporal co-authorship network. For that purpose, Biodiversity Data Journal, a mixed data journal in the biodiversity domain was used as the unit of analysis in this study as this domain had proposed data paper as a new mechanism for data publication. In addition, bibliometric records of 247 data papers and 638 articles involving two or more researchers were collected from the Web of Science. The results indicated that the dynamic co-authorship networks of data papers and articles in the biodiversity domain exhibited the scale-free property of a complex network and the small-world property in the Watts-Strogatz sense during the network evolution. Also, both publication types kept the structure of locally cohesive author groups over time in the networks. The implementation of TTBC (Temporal Triadic Betweenness Centrality) has allowed for the examination and tracking of the evolutionary trends of important or influential time-dependent authors (nodes) by the target year. And last, visualization with a dynamic approach enabled a more effective identification of analysis results, such as the exhibited structural difference in the temporal co-authorship networks of data papers and articles in the biodiversity domain, which can be interpreted as the structural properties of the networks among collaborative researchers dealing with data.

키워드: 시간적 네트워크, 공저 네트워크, 시간적 공저 네트워크, 동적 네트워크, 네트워크 분석, 시간적 매개중심성, 네트워크 중심성, 데이터학술지, 데이터논문
temporal networks, co-authorship networks, temporal co-authorship networks, dynamic networks, network analysis, temporal betweenness centrality, network centrality, data journal, data paper

* 본 논문은 서선경의 박사학위논문 『시간적 공저 네트워크의 구조 분석에 관한 연구』(2023)를 요약·수정한 것임.

** 이화여자대학교 문헌정보학과 문헌정보학 박사, 이화사회과학원 비상임연구원(welloffssk@gmail.com)

■ 논문접수일자: 2024년 8월 22일 ■ 최초심사일자: 2024년 9월 4일 ■ 게재확정일자: 2024년 9월 10일
■ 정보관리학회지, 41(3), 381-409, 2024. <http://dx.doi.org/10.3743/KOSIM.2024.41.3.381>

© Copyright © 2024 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구의 배경 및 목적

최근 네트워크 분석 영역에서는 진화적 역학을 조사하기 위해 시간을 고려하는 동적 네트워크 분석에 관한 요구가 크게 증가하고 있다 (Farine, 2018). 동적 네트워크는 진화하는 네트워크(evolving networks)와 시간적 네트워크(temporal networks)로 구분할 수 있다. 진화하는 네트워크 분석은 정적 네트워크 분석을 위한 사회 네트워크 분석 지표들을 활용하여, 실증적 데이터 분석 결과의 규칙성을 기반으로 진화하는 네트워크의 속성을 파악하고 이를 특성화하는 연구들이 대부분이다. 진화하는 네트워크는 시간적 속성(temporal property)을 중심으로 접근할 수 있는데, 진화하는 그래프는 시간적이지만, 시간적 그래프는 항상 진화할 필요는 없으며 동적일 수 있다는 점에 유의한다. 즉, 연결이 활성상태인 시간을 명시적으로 나타내는 네트워크인 시간적 네트워크에서 노드들을 추가하거나 삭제할 필요가 없긴 하지만 동적일 수 있다는 점으로, 그 시간적 속성은 때에 따라서 링크들하고만 관련이 있다(Tabassum et al., 2018).

과학계량학 분야에서는 과학적 협업의 추세와 패턴을 분석하기 위해 공저 문헌에 기반한 서지 데이터를 활용하여 연구를 수행해 왔다. 공저 네트워크는 과학적 협업을 기반으로 구축된 사회 네트워크로 볼 수 있으므로 공저 네트워크 분석은 협업을 연구하는 다양한 방법론 중에서 가장 효과적인 분석 방법이라 할 수 있다(이수상, 2010). 과학적 협업에서 네트워크

행위자(노드)는 저자이고, 관계(링크)는 그들 사이의 공동저술이며, 적어도 하나의 공동저술을 한 문헌이 있는 경우 두 저자들 사이에 관계가 존재하는데, 이 공저 네트워크는 시간이 지남에 따라 신규 저자와 신규 링크의 추가를 통해 성장하는 동적 네트워크이다. Barabási et al.(2002)은 네트워크 위상(토폴로지) 지도에서 네트워크 역학을 밝히기 위해서는 노드와 링크가 그 네트워크에 추가된 시간을 아는 것이 중요하며, 공저 네트워크는 문헌의 출판연도를 통해 시간의 진화를 추적할 수 있는 협업 네트워크로써 동적 및 구조적 메커니즘 추론에 적합함을 주장했다. 정리하면, 공저 네트워크는 ‘저자 A가 포함된 사회적 환경은 무엇인지?’와 같은 개인 수준에서 ‘저자들 간의 연결들에서 발생하는 전체적인 구조는 무엇인지?’와 같은 집단 수준으로 확장할 수 있게 한다. 이 관계에서 시간적 요소가 있으며, 출판연도 정보로부터 그 관계가 시간이 지남에 따라 형성되고 해체되는 것을 확인하여, 시간적 네트워크 구축이 가능하다.

시간적 네트워크는 초기에 정적 네트워크 개념을 시간적 네트워크로 단순히 확장하는 연구들에서 시간적 네트워크의 고유한 이론들로 이동하는 흐름이다(Holme & Saramäki, 2019). 시간적 네트워크에서 노드와 링크는 그 자체로 동적 개체들이며, 그들의 역학은 시간적 네트워크 모델링에 대한 주요한 목표이고, 실증적 데이터에 의해 알려질 수 있다. 시간성(temporality)은 두 노드 사이의 경로와 같은 네트워크 과학의 가장 기본적인 개념조차 근본적으로 변경하고 해당 연구에는 추가적인 계산 및 수학적 도구들이 필요로 되기 때문에 시간적 네트워크

는 정적 네트워크보다 훨씬 더 복잡한 대상이다. 따라서 시간적 네트워크와 그 동작에 대한 적절한 설명에는 시간적 네트워크의 구조적 및 동적 개체들의 통합적인 모델링이 필요하다(Masuda & Lambiotte, 2016).

한편, 학술커뮤니케이션 영역에서는 데이터 공유를 통해 여러 연구자 집단이 기존의 데이터를 자유롭게 이용함으로써 후속연구에 도움을 주고 새로운 연구를 이끌어낼 수 있다는 인식이 자리 잡았다. 데이터가 연구의 주요한 결과물이자 학술커뮤니케이션의 통로로써 주목받기 시작하면서 데이터에 대한 오픈엑세스 요구가 지속적으로 성장하고 촉진되었다. 이에 데이터를 공유하고 그 이용을 지원하는 등 데이터 관리를 위한 대체 방안으로 과학적 출판 모델을 반영하는 데이터학술지가 새롭게 제기되었다(Chen, 2016). 데이터학술지에 게재되는 데이터논문은 연구에 대한 분석이나 결론을 작성하는 것이 아니라 데이터에 대한 사실을 상세하게 기술하는 것으로 전통적인 학술논문과 차이가 있다. 이러한 데이터논문은 학술적으로 가치가 있는 연구데이터를 공식적으로 출판한 것이며, 그 주제 분야의 데이터를 생산·수집하고 조직, 평가, 보존, 이용, 인용을 위한 맥락들을 기술한 것으로 기존의 과학적 문헌 유형인 학술논문과는 또 다른 의미를 지닌 연구 협업의 공식적인 성과물이자, 직접적인 내용물이라 할 수 있다. 이에 여기에 참여한 저자들의 공저 네트워크를 분석한다면, 과학적 연구데이터를 중심으로 연구 협업의 구조적 행태를 명시적 관점에서 파악할 수 있을 것이다.

따라서 본 연구는 데이터와 논문들을 기반으로 연구자 사이의 연결 패턴에 대한 역동적인

변화를 식별하기 위해 계량정보학 측면에서 시간적 네트워크 모델링을 목표로 공저 네트워크를 진화적 관점으로 분석하고, 시간성이 추가된 중심성 척도를 개발하여 시간적 공저 네트워크의 구조적 특성을 규명하고자 하였다. 본 연구의 목적을 위해 선정된 도메인은 데이터 출판에 관한 상당한 논의와 다각적인 연구가 수행되는 생물다양성 분야이며, 분석 대상 학술지는 Biodiversity Data Journal(이하 BDJ)이다. BDJ는 데이터논문과 일반 학술논문이 모두 게재된 혼합형 데이터학술지로 데이터 자원에 대한 출판을 지원하기 위해 세계 생물다양성 정보기관인 GBIF(Global Biodiversity Information Facility)와 학술출판사 Pensoft가 함께 협력하여 2013년에 창간했다. Seo와 Kim(2020)이 WoS에 색인된 데이터학술지 93종 중 데이터논문 비율이 20% 이상인 데이터학술지 9종을 대상으로 동료심사 시스템 유형과 심사기준, 편집위원회 구성을 조사한 결과, BDJ는 동료심사를 관할하는 주제 편집위원들(Subject editors Position)을 195명 보유하고 있는 동료심사된 오픈엑세스 데이터학술지로 확인됐다.

이와 같이 데이터학술지 출판을 위한 체계가 마련되어 발간된 BDJ를 분석 대상으로 선정하였으며, 본 연구의 목적 및 범위에 따른 구체적인 연구 문제는 다음과 같다.

첫째, 생물다양성 분야의 데이터논문 공저 네트워크와 학술논문 공저 네트워크에서 발견되는 구조적 현상은 시간이 지남에 따라 어떻게 변화하는가?

둘째, 생물다양성 분야의 데이터논문 기반 시간적 공저 네트워크와 학술논문 기반 시간적 공저 네트워크에서 식별된 영향력 있는 저자들은

시간에 따라 어떻게 변화하는가?

1.2 선행연구

데이터논문과 학술논문을 대상으로 시간적 네트워크 분석 기법을 활용하여 공저 관계에 기초한 생물다양성 분야의 구조적 특성을 규명하고자 본 연구에서 검토한 선행연구는 크게 세 그룹으로 구분할 수 있다. 첫 번째 그룹은 공저 네트워크에 관한 연구이고 두 번째 그룹은 시간적 네트워크에 관한 연구이며, 세 번째 그룹은 데이터학술지나 데이터논문에 관한 연구이다.

우선, 진화적 관점에서 공저 네트워크 구조의 기본 메커니즘을 다룬 국외 연구들은 다음과 같다. 공저 네트워크의 역학과 진화를 조사하기 위해 시계열 분석 실시한 Barabási et al. (2002)은 공저 네트워크 변화 과정에서 네트워크가 점차 더 좁아지는 경향이 있고, 새로운 노드들이 이미 많은 노드들과 연결된 노드에 링크를 선호하는 현상인 선호적 연결이 증가하며, 군집화 계수가 낮아짐을 밝혔다. Chen et al. (2013)은 1978년부터 2010년까지 Scientometrics 학술지에 논문들 2,541건을 기반으로 개별 저자들(미시), 기관들(중간), 국가들(거시)의 3개 단계에서 협업 네트워크의 진화를 분석했다. 그 결과, 공저 네트워크에서는 기존 저자들의 삭제와 신규 저자들의 추가를 경험하는 반면에, 기관과 국가의 협업 네트워크에서는 주요 노드들의 비교적 안정적인 구조를 가지고 있음을 지적했다. 또한, Savić et al. (2019)은 컴퓨터 과학 분야와 수학 분야에 초점을 맞춰 공저 네트워크의 전형적인 구조적, 진화적 특성을 정

리했다. 이 연구에서는 컴퓨터 과학과 수학 분야의 공저 네트워크에서 연결정도 분포가 멱함수 법칙을 준수하는 무척도 네트워크 속성을 확인했고, 좁은 세상 속성이 나타났으며, 거대 연결 컴포넌트의 존재를 확인했다. 그리고 공저 네트워크의 진화에서 자주 관찰되는 경향 중 하나로 평균 노드 연결정도의 증가를 확인하여 시간이 지남에 따라 협업이 증가하는 것을 파악했다. 국내 연구로는 Kim et al. (2016)이 1948년부터 2011년까지 대한민국의 모든 분야 협업 네트워크 구조와 진화를 조사하기 위해 네트워크 분석을 수행한 결과, 국내 연구 커뮤니티에서 저자의 생산성과 협업이 매년 증가했으며, 거대 연결 컴포넌트가 식별되었고, 저자들이 평균 6단계 내에서 서로 협업할 수 있음을 파악하여 좁은 세상 특성이 확인됨을 밝혔다.

두 번째 그룹은 동적 네트워크 모델링을 위해 시간 정보에 따른 시간적 네트워크를 다루는 연구들이다. Kempe 등(2002)은 많은 네트워크 문제들이 시간과 관련된 근본적인 관계에 기초하는 것에 주목하여, 각 연결들에 교신(communicated)한 시간을 지정하는 시간 레이블로 주석이 달린 그래프 모델을 제안하고 이러한 그래프를 시간적 네트워크라 제시했다. 시간 순서를 중시하는 레이블들을 갖는 경로들에서만 네트워크의 정보가 흐름(flow)이라는 개념을 모델링하기 위해 이 연구에서는 경로를 시간에 관한(time-respecting)이라 명명했다. 이와 같은 시간에 관한 경로는 네트워크를 통해 정보가 전파되는 방식을 이해하는 데 중요한 구조임을 강조했다. Blonder 외(2012)는 생태학 및 진화 생물학 영역에서 관찰 데이터의 네트워크 분석에 시간적 역학을 통합해야하는 필요성을 강조

하며, 동적 네트워크 분석의 중요한 기본 개념들을 정리하고, 시간-순서(time-ordered) 네트워크 분석에 대한 가이드를 제공했다. 그들은 이 연구에서 동적 네트워크는 흐름 프로세스(flow process)와 기본적인 네트워크 위상(토폴로지)의 변화와 같은 주요한 두 가지 유형에 의존함을 언급했다. 여기서 흐름 역학(flow dynamics)은 주어진 노드들과 연결들의 집합에 대하여 네트워크 노드들 사이에 자원들(resources), 관계들(contacts), 정보들, 질병들이 전파되는 방식이다. 그리고 위상(토폴로지) 역학(topological dynamics)은 네트워크 구조가 링크들의 가중치 또는 그 정체성(identity)의 변동을 통해 변화하는 방식을 의미하며, 재배선(rewiring)이라고도 한다. Nicosia et al.(2013)은 시간적 네트워크에서는 노드 인접성(node adjacency)과 도달가능성(reachability)에 대한 개념이 그 연결들의 정확한 시간 순서에 결정적으로 의존하는 것임을 지적했다. 이에 그들은 기존의 정적인 복잡계 네트워크 특성에 맞게 제안되고 사용된 모든 개념과 매트릭스(metrics)가 인과성에 대한 시간 순서의 영향을 고려할 수 있도록 다시 정의되거나 확장되어야 함을 주장하며, 시간 가변 그래프에서 적용할 수 있는 핵심적인 중심성 척도들을 제시했다.

세 번째 그룹의 연구는 데이터학술지나 데이터논문에 관한 것으로 국내 연구를 중심으로 살펴보면 다음과 같다. 김지현 외(2019)는 국내외 데이터학술지 현황과 발간 정책 등을 조사하고 데이터논문의 구성 요소, 데이터 품질 평가 정책, 레포지터리 연계 방안 등을 분석하여 국내 데이터학술지 창간을 위한 필요요건 및 발간 프로세스를 제안했다. 정은경(2019)은 다

학제적 분야의 데이터학술지인 Scientific Data에 게재된 총 713건의 논문을 대상으로 데이터 논문의 주제, 인용한 저작물, 저자 현황을 분석하여 데이터학술지의 현황과 특성을 규명했다. 이 연구에서 공저 네트워크 분석을 수행한 결과, 저자의 주된 주제 영역은 생명공학, 물리학 등으로 파악되고, 공저자 수는 평균 12명이며, 특정 연구자 그룹이 폐쇄적으로 공저 활동을 수행하는 것으로 나타났음을 밝혔다. Seo와 Kim (2020)은 데이터학술지의 출판 과정에서 주요한 이슈 사항들을 조사하여 Interactive Review, Community Review와 같은 변형된 동료심사 모델들이 채택되고 있음을 확인하고, 동료심사를 관할하는 편집위원들의 다양한 포지션을 식별했다. 이들은 데이터학술지 출판과 함께 그 핵심 과정에 점차 새로운 관행이 자리 잡을 것으로 예상한다고 하였다. 이상과 같이 데이터학술지 출간 이후, 미시적인 관점에서 데이터학술지를 대상으로 계량정보학적 측면에서 분석한 연구는 아직 초기 단계이다. 따라서 오픈 과학 환경을 반영하고 학술 생태계에서 요구되는 데이터 공유, 재이용, 인용을 촉진시키기 위해 출현한 데이터학술지를 대상으로 과학계량학적 접근 분석을 수행하여 그 특성을 규명할 필요가 있다.

2. 시간적 공저 네트워크 매트릭스(metrics)

2.1 시간적 네트워크 매트릭스

시간적 네트워크가 인코딩되면, 네트워크의

구조와 진화에 관한 질문을 시작할 수 있다. 시간적 네트워크 매트릭스를 적용하면 주어진 그래프 집합이 시간에 따라 정렬된다는 사실을 명시적으로 설명하고 다각적인 접근으로 분석을 수행할 수 있다. 시간적 네트워크에서는 하나의 노드가 다른 노드와 교신하는 데 걸리는 시간을 정량화할 수 있다. 노드 A와 노드 B 사이의 경로에 경로 지속 시간(duration of the path)을 추가로 기록하여, 시간적 경로거리(temporal path length)와 최단 시간적 경로거리(shortest temporal path length, shortest temporal distance, Latency)를 산출할 수 있다. 최단 시간적 경로거리는 두 노드 간에 가장 빠른 경로의 시간적 경로거리이며, 이동한 링크들의 수 또는 이동한 시간 단계 또는 두 노드 간 마지막 교신 사이의 시간 차이 등과 같이 계산하는 방법에 따라 다르게 측정된다(Sizemore & Bassett, 2018; Thompson et al., 2017). 특히 효율적인 시스템의 경우, 동적 네트워크 내에서 정보가 가장 짧은 경로, 더 정확하게는 가장 빠른 경로를 따라 이동할 것으로 예상할 수 있다.

최단 시간적 경로거리가 정의되면 네트워크 중심성의 개념을 동적 네트워크로 확장할 수 있다. 시간적 네트워크에서는 시간적 경로거리의 정의를 사용하여 지정된 시간 범위(time window)내에서 최단경로를 가장 빠른 경로로 바꿔서 생각한다. 시간적 매개중심성(temporal betweenness centrality)의 대체 정의로는 가장 빠른 시간에 관한 경로들(the fastest time-respecting paths)을 최단 위상적 시간에 관한 경로들(the shortest topological time-respecting paths)로 바꿔서 말할 수 있다(Holme &

Saramaki, 2012). 이는 주어진 시간 범위 안에서 특정 노드를 통과하는(거쳐) 가장 빠른 경로를 강조하고 있으며, 시간적 매개중심성이 높은 노드는 주어진 시간 범위 안에서 정보의 전달이나 흐름을 통제할 뿐만 아니라, 그 흐름의 속도를 통제하는 데 큰 영향력을 가질 수 있다는 것으로 해석할 수 있다.

검토된 시간적 네트워크 연구 영역에서 개발되거나 모델링되는 매트릭스는 최단 시간적 경로거리 등과 같이 흐름 역학 측면에서 정보의 전달이나 질병의 전파를 파악하는 데 그 의미를 얻을 수 있으며, 시간 간격(시간분해능, time resolution) 측면에서 연 단위인 종적 연구의 경우에는 최단 시간적 경로거리의 개념이 큰 의미를 지니지 않을 수 있다(Thompson et al., 2017). 그러나 공저 네트워크 분석은 공저 빈도수를 가중치로 인접행렬을 생성하는 동시출현(co-occurrence) 측면의 가중 네트워크이고, 분석 데이터세트는 대부분 출판연도를 시점으로 하는 연 단위 데이터세트라는 특성을 지닌다. 그러므로 본 연구는 시간적 네트워크에서 활용하는 매트릭스와 공저 네트워크에서 활용하는 매트릭스를 토대로 시간적 공저 네트워크 분석을 위한 새로운 중심성 척도를 고안하여 제시하고자 한다.

2.2 시간적 삼각매개중심성 TTBC 제안

공저 네트워크가 시간이 지남에 따라 등장하는 저자들이나 공저 관계로 형성된 링크들이 변화하는 동적 네트워크임에 초점을 맞추고, 계량서지적 분석 대상이 되는 서지정보를 통해 자료의 수집 단계에서부터 시간 정보인 출판연

도를 교신 시점으로 획득할 수 있음을 주목한다면, 시간 진화의 측면으로 공저 관계를 접근할 수 있다. 즉, 출판연도를 통해 교신 순서를 구성할 수 있으며, 이를 통해 시간에 관한 경로를 파악할 수 있다. 그러나 시간적 분해능 측면에서 분석 대상인 데이터셋이 연단위로 구분되어 해당 시점에 하나의 저자(노드)에서 여러 저자(노드) 간에 링크가 형성되므로 앞서 살펴 보았던 시간적 네트워크 이론 연구 영역에서 개발되고 적용되는 최단 시간적 경로거리 등과 같은 중심성 지수를 그대로 적용하기가 어렵다고 판단된다. 따라서 본 논문에서는 검토된 시간적 네트워크 매트릭스를 참고하고, 이재윤(2006)이 제안한 TBC에 시간 차원을 포함시켜 동적 접근 방식으로 공저 관계를 파악할 수 있도록 시간적 공저 네트워크의 전역중심성 지수인 시간적 삼각매개중심성(Temporal Triadic Betweenness Centrality, 이하 TTBC)을 제안한다.

본 논문에서 제안하는 TTBC는 측정의 기본 분석 단위를 3개의 노드로 구성된 삼각형으로 하고, 공저 네트워크 상에서 전체 저자(노드)

를 대상으로 공저 관계가 형성된 기간이 더 오래된(길었던) 저자(노드)가 다른 두 저자(노드)를 매개하는 경우의 수를 누적하여 산정한다. 기존 정적 네트워크 분석에 의한 공저 네트워크에서는 저자 사이의 공저 관계 형성 시점을 고려하지 않은 상태에서 공저 횟수만을 고려하는 중심성 분석이 수행되었다. 이를테면, 이재윤(2006)의 TBC가 공저 관계를 형성한 저자 간에 언제 공저를 시작했는지(since)를 모르는 상태에서 서로 공저를 몇 번 했는지(frequency)만을 가지고 매개가 되는 저자를 추론하는 방식이었다면, 본 논문에서 제안하는 TTBC는 교신 순서의 개념을 적용해서 공동 저자들 간에 그 관계가 형성된 시점을 파악하고, 이 최초 공저 시점의 순서대로 가중값을 부여하여 삼각관계에서 두 저자를 먼저 서로 알고 있었던(공저를 했었던) 저자가 매개가 되는 것으로 추론하는 방식이다. TTBC의 네트워크 시각화 및 조건을 정리하면 <그림 1>과 같다.

TTBC를 산출하기 위해서는 먼저, 분석 데이터셋의 분석기간을 기준으로 분석기간의 최종연도를 대상연도로 하고, 분석기간 중 저

	네트워크 시각화	조건	
TTBC		최초 교신 순서 가중값	$fcsw(X,Z) < \min(fcsw(Y,X), fcsw(Y,Z))$ $*fcsw = \text{대상연도} - \text{최초공저출판연도} + 1$
		최초 공저 출판연도	$fyear(X,Z) > \max(fyear(Y,X), fyear(Y,Z))$

<그림 1> TTBC의 네트워크 시각화 및 조건

자들(노드들)의 공저 관계가 형성된 최초 공저 출판연도를 파악한다. 이를 위해서는 각 출판 연도별 공저 여부만을 확인하면 되므로 공저 빈도수로 가중값을 넣은 가중 네트워크가 아닌 이진 네트워크가 필요로 된다. 예를 들면, 분석 데이터세트의 분석기간의 시작연도가 2001년 이고, 최종 연도가 2022년 일 때, 저자 X와 저자 Y의 공저 링크 리스트가 (저자 X, 저자 Y, 2010), (저자 X, 저자 Y, 2015), (저자 X, 저자 Y, 2018)인 경우에 (저자 X, 저자 Y, 2010)만을 남기고 나머지 데이터는 삭제한다. 이 경우, 분석기간 내에서 저자 X와 저자 Y는 최초 공저 출판연도가 2010년이므로 최초 공저 출판연도가 확인되면 그 시점 이후의 링크는 필요로 되지 않는다. 이렇게 작성된 링크 리스트에 최초 교신 순서 가중값(First Contact Sequence Weight, 이하 FCSW)을 부여한다. FCSW는 대상연도에서 최초 공저 출판연도를 감산한 것에 1을 가산한 값으로 한다. 즉, 이는 최초 공저 출판연도 이후로 대상연도까지의 시간의 경과에 대해 가중치를 부여한 것으로 이 가중값(FCSW)은 두 저자가 최초 공저한 시점의 순서가 일찍일수록 (빠를수록) 더 큰 값을 갖게 된다.

TTBC는 공저 네트워크에서 앞서 설명한 FCSW를 부여해서 링크 리스트(행렬)를 작성하고, 분석 기준이 되는 노드를 저자 Y라고 할 때, 저자 Y를 제외한 다른 노드들의 조합 쌍 중에서 <그림 2> TTBC의 조건을 만족하는 노드 쌍(X,Z)의 비율로 산출한다. 예를 들면, <그림 2>에서 분석 데이터세트의 분석 기간이 2001년부터 2022년인 경우, 2022년을 대상연도라 한다. 대상연도인 2022년을 기준으로 저자 Y와

저자 X는 최초로 공저를 2001년에 했고, 공저 관계 Y-X의 FCSW 값은 22($FCSW = 2022 - 2001 + 1$)가 된다. 같은 계산 방식에 따라 저자 Y와 저자 Z의 최초로 공저를 2006년에 했을 때, 공저 관계 Y-Z의 FCSW 값은 17($FCSW = 2022 - 2006 + 1$)이다. 저자 X와 저자 Z는 최초로 공저를 2012년에 했으므로 공저 관계 X-Z의 FCSW 값은 11($FCSW = 2022 - 2012 + 1$)이 된다. 저자 X와 저자 Z는 대상연도에서 가장 가까운 시점에 공저를 수행하였으므로 TTBC의 조건에 따라 가장 부여받는 가중치가 가장 적은 값이 된다. 여기서 TTBC를 계산할 때, 동일한 논문에 3명의 공저가 있는 경우에는 TTBC의 조건식에서 부등호에 등호를 함께 넣어줄 수 있다. 그러나 저자 X, 저자 Y, 저자 Z가 모두 동일한 출판연도로 가중값을 부여받았을 때, 시간 지연을 정보에서 얻지 못하므로 불확실한 인과 관계를 포함시키지 않기 위해 조건에서 등호는 넣지 않도록 한다.

정리하면, TTBC는 공저 관계에 시간적 차원을 적용함으로써 분석 대상 데이터세트의 시간 범주 내의 출판연도를 통해 각 저자들 사이에 알고 지낸 기간을 수치로 계량화할 수 있는 이점을 포착하고, 여기에 가중값을 부여하여 작성된 공저 네트워크에서 측정된 전역중심성 척도이다. 이때, 최초 교신 순서 가중값인 FCSW는 분석 시간 범주 안에서 최초로 등장하는 공저 관계 시점인 출판연도가 가장 적은 순서대로 높은(큰) 값의 가중치를 부여한다는 의미이다. 기존에 TBC가 정적 데이터세트를 기반으로 공저 빈도수를 통해 결속력을 측정할 수 있는 맥락에서 매개중심성을 산출했다면, TTBC는 동적 접근 방식으로 분석 데이터세트의 시

간 지연을 수반하여 시간의 흐름에 따라 더 먼저, 더 오래 전에 알았던 저자가 그 이후의 공저 관계 형성에 영향을 끼칠 수 있다(다른 저자들을 서로 이어주는데 기여를 했다)는 논리이며, 이 중심성 척도를 적용하면 매개라는 개념에 부합되고 그렇게 추론된 결과에 타당성이 있을 것으로 생각한다.

2.3 시간적 삼각매개중심성 TTBC 구현

전역중심성 척도인 시간적 삼각매개중심성 TTBC를 구현하기 위해 python으로 TTBC 프로그램을 개발하였다. TTBC를 산출하기 위해서는 입력데이터인 최초 교신 순서 가중값 FCSW 리스트 또는 행렬이 필요하다. FCSW 리스트나 행렬을 분석 대상연도별로 생성하려면 데이터세트가 많기 때문에 이를 위한 작업도 python으로 프로그램을 개발해서 수행하였다. 산출된 FCSW 행렬을 입력데이터로 하여 TTBC 값을 산출하기 위해서 고려된 삼각관계 유형은 무방향 네트워크에서의 폐쇄형 삼각관계(Closed Triad)와 개방형 삼각관계(Open Triad)까지만 이다. 폐쇄형 삼각관계는 해당 시점별로 이미 공저 관계가 형성된 상태를 기반으로 조건에 부합하면 점수를 부여하는 것이고, 개방형 삼각관계는 해당 시점별로 제3의 공유 저자(노드)를 갖는 두 저자(노드)가 아직 연결되지 않았지만 연결될 수 있게 해주는 가능성을 갖는 매개 저자(노드)를 식별하여 조건에 부합하면 점수가 부여되는 것을 의미한다.

따라서 본 연구에서 개발된 TTBC에서는 시간이 지나서 공저 관계에 직접적인 영향을 끼친 것으로 식별되는 폐쇄형 삼각관계에서 조

건에 부합하는 노드에게 1점을 부여하고, 개방형 삼각관계에서 조건에 부합하여 추후 영향을 끼칠 가능성으로 식별되는 노드에게 0.1점을 부여하도록 설계하고 이를 구현했다. 이처럼 시간 차원이 통합된 시간적 공저 네트워크에서 폐쇄형 삼각관계는 노드와 링크가 재배선된다는 점에서 위상(토폴로지) 역학 측면으로 볼 수 있다. 또한 TTBC는 교신 시점 순서가 고려되기 때문에 흐름 역학 측면의 접근이기도 하다. 여기서 각 노드들의 초기 TTBC 값은 0으로 시작하고, TTBC 조건식에 등호를 넣지 않고 모두 동일한 출판연도를 제외한 것과 같이 개방형 삼각관계에서도 두 저자의 링크가 동일한 출판연도에 생성된 경우일 때는 시간 지연 정보를 얻지 못하므로 이를 제외하도록 한다.

3. 연구 설계

3.1 연구 방법

연구 문제에 대한 고찰을 위해 본 연구에서는 연구 방법으로 시간적 네트워크 개념과 분석 방법을 활용하였으며, 다음과 같은 방법론적 차별성을 강조하여 연구 과정을 진행하였다.

첫째, 생물다양성 분야의 공저 패턴에 대한 시간적 진화를 추적하기 위해 동적 공저 네트워크 분석을 수행하였다. 이를 위해 거시적 측면으로 전체 분석 데이터세트를 시간 단위로 분할하여 공저 네트워크에 대한 계량서지학적 매트릭스의 시간 변화를 조사하였다. 이는 결

과적으로 시간이 지남에 따라 펼쳐지는 생물다양성 분야 공저 관계의 구조적 메커니즘 식별이 가능하다.

둘째, 흐름 역학과 위상(토폴로지) 역학 측면으로 시간적 공저 네트워크에 기반한 중심성 분석을 수행하였다. 네트워크 연구의 미시적 분석 수준에서는 노드 간 연결 관계에 나타나는 특성, 그리고 노드와 링크가 형성되는 다양한 역할과 행위에 대한 분석 및 해석이 가능하다. 여기서 시간적 공저 네트워크 분석은 시간 정보를 고려할 수 있으며, 본 논문에서 제안한 TTBC를 적용하여 매개 영향력이 높은 저자들을 식별하고 그 추세를 파악하였다.

셋째, 데이터논문과 학술논문을 구분하고 각 데이터세트 별로 분석을 수행하여 각 네트워크 구조에 대한 비교분석을 실시하였다. 이는 데이터를 다루는 연구자와 일반 연구자들의 협업 양상을 파악할 수 있다.

3.2 데이터 수집과 처리

3.2.1 Biodiversity Data Journal의 기초 통계데이터

BDJ는 2018년에 JCR에 등재되었으며, WoS에는 출판연도가 2015년인 것부터 서지레코드로 색인되어 있다. 이를 기준으로 2015년부터 2021년까지 지난 7년 동안 BDJ의 각 연도별 저널 영향력 지수(JIF), 데이터논문 건수, 학술논문 건수, 전체(데이터논문 건수+학술논문 건수) 대비 데이터논문 비율, 공저 문헌 건수, 공저 문헌 비율 등을 정리하면 <표 1>과 같다.

2015년부터 2021년까지 BDJ에 게재된 데이터논문 258건과 학술논문 689건을 대상으로 단독 저자 문헌 건수, 공저자 문헌 건수를 각 연도별로 분석하면 데이터논문의 경우, 단독 저자 데이터논문이 11건, 2인 이상 공저 데이터논문은 247건으로 발표된 전체 데이터논문 258건에서 공저로 출판된 데이터가 96%를 차지하고 있

<표 1> Biodiversity Data Journal의 연도별 JIF, 데이터논문 건수, 학술논문 건수, 전체 문헌 건수 대비 데이터논문 비율, 공저 문헌 건수, 공저 문헌 비율
(조사 리소스: WoS / 조사 시점: 2022.06.23. / JIF 업데이트: 2022.07.08.)

구분 \ 연도		2015년	2016년	2017년	2018년	2019년	2020년	2021년	2015년-2021년
JIF		-	-	-	1.029	1.331	1.225	1.540	
데이터논문 건수		21	23	17	9	32	68	88	258
학술논문 건수		88	109	55	61	67	130	179	689
전체 문헌 건수		109	132	72	70	99	198	267	947
데이터논문 비율		19%	17%	24%	13%	32%	34%	33%	27%
데이터 논문	단독 저자 문헌 건수	0	2	1	0	2	1	5	11
	공저 문헌 건수	21	21	16	9	30	67	83	247
	공저 문헌 비율	100%	91%	94%	100%	94%	99%	94%	96%
학술 논문	단독 저자 문헌 건수	10	15	3	5	4	4	10	51
	공저 문헌 건수	78	94	52	56	63	126	169	638
	공저 문헌 비율	89%	86%	95%	92%	94%	97%	94%	93%

다. 학술논문의 경우, 단독 저자 학술논문이 51건, 2인 이상 공저 학술논문은 638건으로 발표된 전체 학술논문 689건에서 공저연구가 93%를 차지하며, 해당 기간 BDJ에 게재된 데이터 논문이나 학술논문 모두 단독 저자보다 공저로 발표되었다. 이는 생물다양성 분야의 특성상 데이터 출판이나 학술연구 대부분이 공동연구로 수행되고 있음을 볼 수 있다.

3.2.2 데이터 분석 방법

지금까지 살펴본 결과를 토대로 본 연구에서는 2015년부터 2021년까지 BDJ에 게재된 전체 문헌 947건 중에서 2인 이상 연구자가 참여한 공저 문헌 885건(데이터논문 247건+학술논문 638건)을 최종 분석 대상으로 하고 WoS에서 서지레코드를 추출하여 분석을 수행했다.¹⁾ 수집된 서지레코드에서 BibExcel²⁾을 활용하여 저자레코드를 추출하였고, 이형을 전거하여 마스터테이블을 만들고 데이터논문과 학술논문의 데이터세트를 각 연도별로 생성하였다.

진화적 측면의 동적 공저 네트워크 분석을 위해서 이재운(2013)이 개발한 COOC 프로그램을 활용하여 인접행렬을 작성하였고, NodeXL 프로그램³⁾으로 산출된 결과를 네트워크 수준, 노드 수준, 하위집단 수준으로 구분하고 이에 대한 비교 및 분석을 수행하였다. 네트워크에 내재된 속성을 확인하기 위해 시계열 접근에 의한 평균 군집화 계수와 Sci2 프로그램⁴⁾을 활

용하여 산출된 무작위 네트워크에 대한 군집화 계수를 비교하였다. 또한, 시간적 공저 네트워크 중심성 분석을 위해 본 논문에서 개발한 FCSW 프로그램과 TTBC 프로그램을 활용하여 FCSW 행렬을 생성하고 각 데이터세트의 대상 연도별 TTBC 값을 산출하였다(서선경, 2022; 2023). 그리고 동적 네트워크의 시기별 정적 네트워크 스냅샷 방식으로 시간적 공저 네트워크를 시각화했으며, 이는 Gephi 프로그램⁵⁾을 활용했다. 분석된 결과를 바탕으로 그 의미를 해석하고, 연구 문제에 대한 결론을 도출했으며, 이처럼 본 연구에서 수행한 시간적 공저 네트워크 분석 절차를 도식화하면 <그림 2>와 같다.

4. 생물다양성 분야의 시간적 공저 네트워크 분석

4.1 동적 공저 네트워크 분석 결과

본 절에서는 생물다양성 분야의 시간 변화에 따른 동적 공저 네트워크의 진화를 살펴보기 위해 BDJ를 대상으로 데이터논문과 학술논문을 구분하고 2015년부터 2021년까지 1년 단위 누적적 분할 방식으로 시계열 접근에 의한 동적 네트워크 분석을 수행하였다. 누적적 분할은 시간이 지남에 따른 네트워크의 성장을 파악하는 것에 유용하다(Weingart et al., 2010).

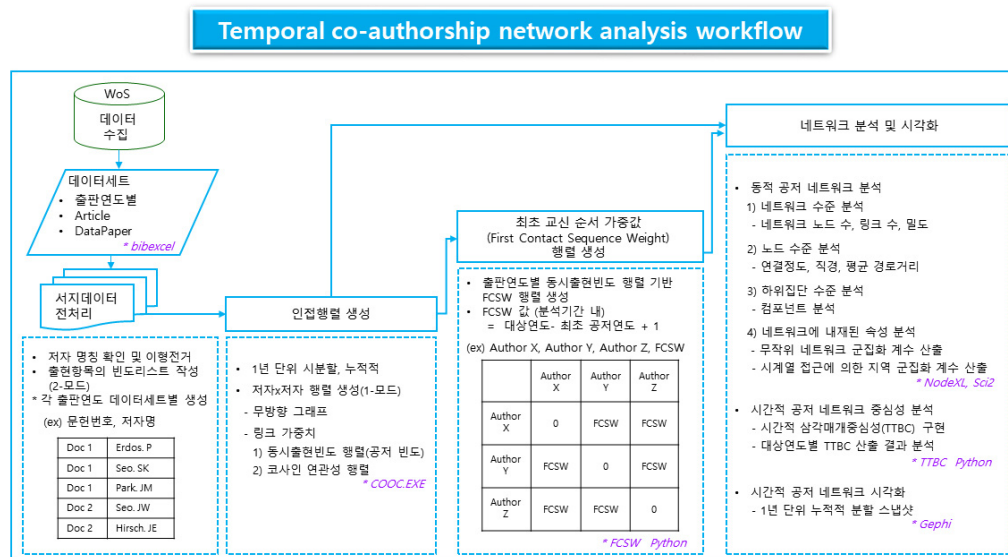
1) 본 논문에서는 분석 도구에 데이터 반입을 위해 저자명에 콤마나 대쉬와 같은 구분자를 삭제하여 전 처리하였기 때문에 제시한 표와 그림에서 저자명의 이름과 성이 구분되지 않는다.

2) <https://homepage.univie.ac.at/juan.gorraiz/bibexcel/>

3) <https://www.smrfoundation.org/nodexl/>

4) <https://sci2.cns.iu.edu/user/index.php>

5) <https://gephi.org/>



〈그림 2〉 시간적 공저 네트워크 분석을 위한 연구 절차도

분석된 결과는 데이터세트의 기간을 1년 단위 누적적으로 다음 2015년, 2015-2016년, 2015-2017년, 2015-2018년, 2015-2019년, 2015-2020년, 2015-2021년과 같이 시기별로 구분하여 제시하였다.

4.1.1 네트워크 수준 분석 결과

시기별 생성된 네트워크의 크기, 밀도와 같이 네트워크 수준에서 구조적 속성을 분석한

결과는 〈표 2〉와 같다.

네트워크의 크기는 네트워크를 구성하는 노드 수와 링크 수에 의해 파악할 수 있다. 〈표 2〉와 같이 시간 변화에 따른 생물다양성 분야 데이터논문과 학술논문의 동적 공저 네트워크 크기는 각 데이터세트 모두 그 규모가 점차 커지고 있는 것이 확인됐다. 또한, 2015년 이래로 두 문헌 유형의 동적 공저 네트워크는 모두, 밀도 값이 점차 0값에 가까워지는 것으로 나타났

〈표 2〉 Biodiversity Data Journal의 동적 공저 네트워크에 대한 네트워크 수준 분석

구분 \ 연도		2015년	2015년 -2016년	2015년 -2017년	2015년 -2018년	2015년 -2019년	2015년 -2020년	2015년 -2021년
데이터 논문	노드 수	198	314	406	462	686	1,282	1,712
	링크 수	3,301	4,071	4,423	4,665	6,286	19,781	23,366
	네트워크 밀도	0.169	0.083	0.054	0.044	0.027	0.024	0.016
학술 논문	노드 수	461	794	1,023	1,221	1,441	1,929	2,633
	링크 수	13,066	14,429	15,798	16,504	17,398	18,997	21,806
	네트워크 밀도	0.123	0.046	0.030	0.022	0.017	0.010	0.006

으며, 이를 통해 생물다양성 분야의 동적 공저 네트워크가 시간이 흐름에 따라 점점 더 성가게(sparsely) 연결되는 네트워크임을 알 수 있다. 이와 같은 네트워크에서는 시간이 지날수록 대부분의 노드 쌍이 서로 직접 연결되지 않고 있는 것으로 파악할 수 있다.

4.1.2 노드 수준 분석 결과

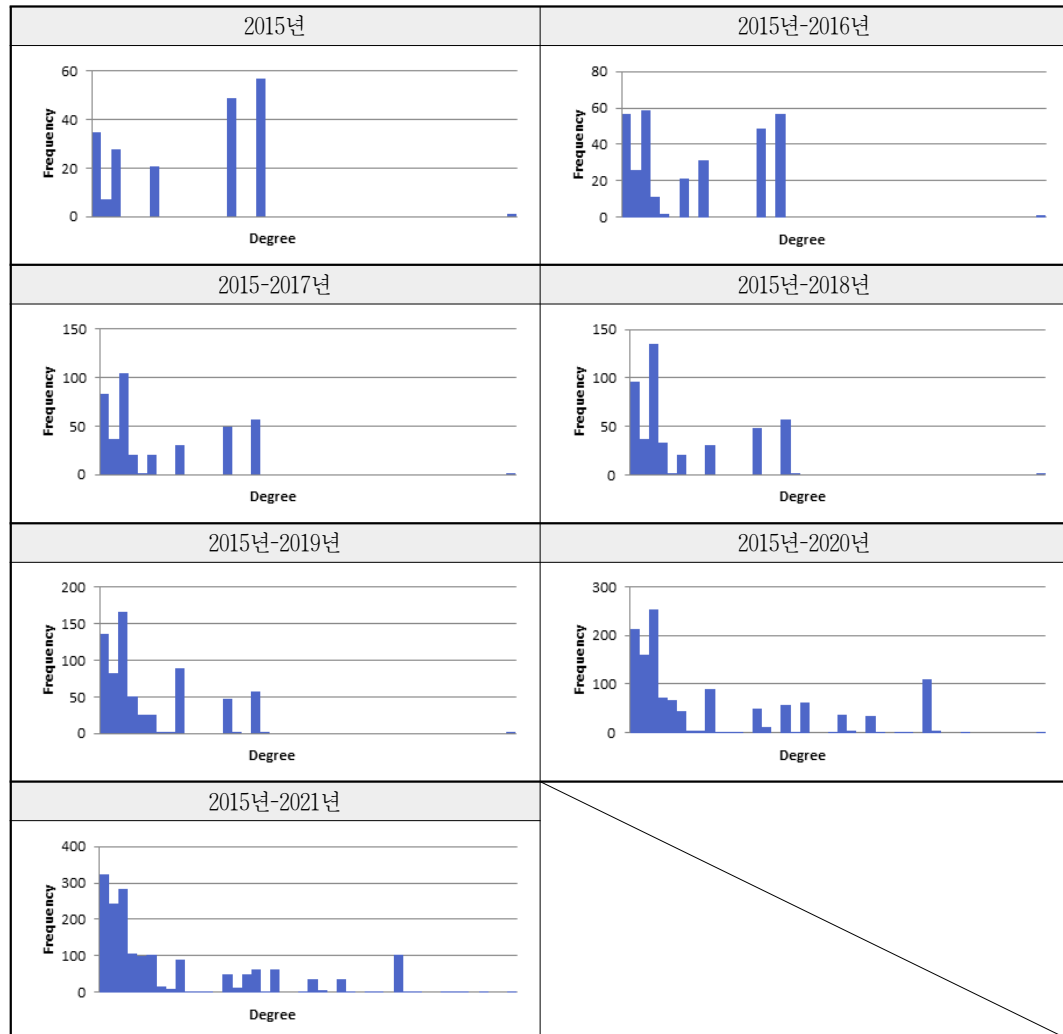
네트워크의 노드 수준 구조적 속성에 해당하는 지표 중 연결정도, 직경, 평균 연결거리 등을 분석한 결과는 <표 3>과 같다. 이를 토대로 직경과 평균 경로거리의 시계열 분석 결과를 살펴보면, 데이터논문은 평균 경로거리가 1,650(2015년)에서 2,940(2015-2021년), 직경이 2(2015년)에서 8(2015-2021년)이며, 학술논문은 평균 경로거리가 1,572(2015년)에서 4,295(2015-2021년), 직경이 4(2015년)에서 13(2015-2021년)으로 두 문헌 유형의 동적 공저 네트워크는 모두, 네트워크가 성장할수록 평균 경로거리와 직경이 증가하는 추세를 보였다. 다음으로

연결정도를 살펴보면, 2015년에 데이터논문의 최대 연결정도가 139로 이에 해당하는 저자는 de Jong Yde였고, 학술논문의 최대 연결정도가 234로 이에 해당하는 저자는 Penev Lyubomir였다. 실제 데이터를 살펴보았을 때, 두 문헌 유형의 동적 공저 네트워크는 모두, 2015년 이래 최대 연결정도에 해당하는 저자들이 부동의 1위를 유지하고 있음을 확인했다.

생물다양성 분야 동적 공저 네트워크의 시간 경과에 따른 연결정도 분포 추이를 파악하기 위해 1년 단위 누적적 분할 측면에서의 연결정도 분포 그래프를 작성하였다. 이를 데이터논문과 학술논문으로 구분하여 시기별로 정리하면 <그림 3>, <그림 4>와 같다. 데이터논문과 학술논문의 동적 공저 네트워크는 둘 다, 2015년부터 많은 링크를 갖고 있는 소수의 큰 허브들이 존재하고 희소한 허브에서부터 소수의 링크를 갖는 대다수의 노드들(다수의 작은 노드들)이 공존하고 있는 것을 확인할 수 있다. 현실 세계의 많은 네트워크들이 링크 수가 역함

<표 3> Biodiversity Data Journal의 동적 공저 네트워크에 대한 노드 수준 분석

구분 \ 연도		2015년	2015년 -2016년	2015년 -2017년	2015년 -2018년	2015년 -2019년	2015년 -2020년	2015년 -2021년
데이터 논문	최대 연결정도	139	143	149	149	149	149	149
	평균 연결정도	33.34	25.93	21.79	20.19	18.33	30.86	27.30
	중앙값 연결정도	49	21	9	9	10	12	12
	최소 연결정도	1	1	1	1	1	1	1
	직경	2	4	4	4	5	7	8
	평균 경로거리	1,650	1,636	1,639	1,740	2,546	2,407	2,940
학술 논문	최대 연결정도	234	234	234	241	241	241	244
	평균 연결정도	56.69	36.35	30.89	27.03	24.15	19.70	16.56
	중앙값 연결정도	10	5	6	6	6	6	6
	최소 연결정도	1	1	1	1	1	1	1
	직경	4	6	8	9	13	12	13
	평균 경로거리	1,572	2,035	2,324	2,912	3,610	3,959	4,295



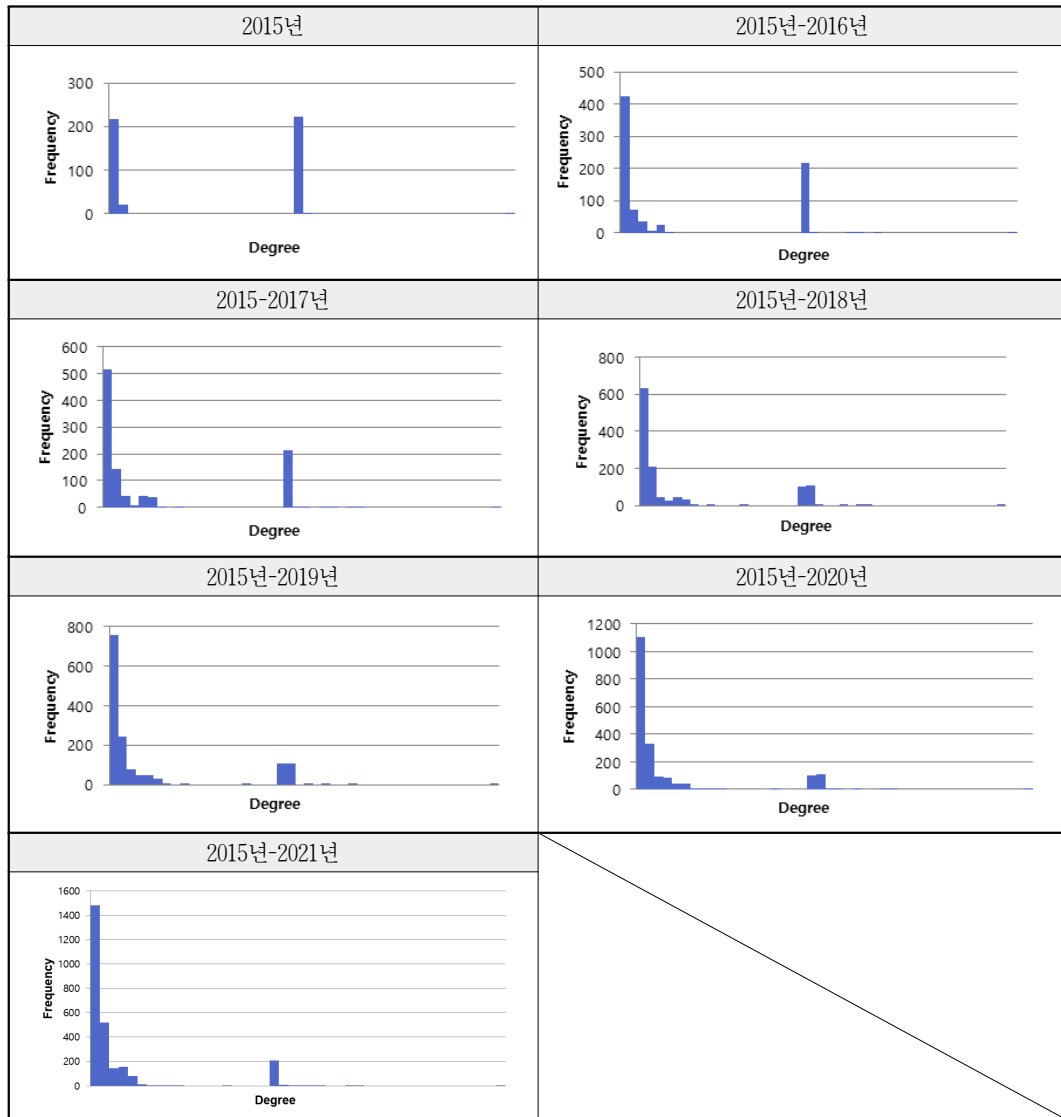
〈그림 3〉 Biodiversity Data Journal의 데이터논문 동적 공저 네트워크에 대한 연결정도 분포

수 분포를 따르는 무척도 네트워크로 나타나는 복잡계 네트워크 형태를 나타내고 있는데 이는 생물다양성 분야의 성장하는 동적 공저 네트워크에 대한 연결정도 분포 추이에서도 확인됨을 알 수 있다.

4.1.3 하위집단 수준 분석 결과

네트워크를 집단 수준에서 분석하기 위해 전

체 네트워크를 하위집단으로 구분해내고, 이를 대상으로 다양한 특성을 분석할 수 있다. 네트워크 하위집단은 노드의 유사성과 응집력, 관계의 완전성을 기준으로 분류 가능하며, 정적 네트워크에서는 컴포넌트(component) 분석, 파당(clique) 분석, 구조적 등위성 분석 등을 사용한다. 여기서 컴포넌트 분석과 파당 분석은 노드들의 군집 기준을 응집도로 하여 하위



〈그림 4〉 Biodiversity Data Journal의 학술논문 동적 공저 네트워크에 대한 연결정도 분포

집단을 구분한다. 컴포넌트 분석은 노드들의 연결된 형태에 따라 구성집단을 식별하고, 파당 분석은 모든 노드들이 연결 상태를 유지하는 파당의 성격에 따라 결속집단을 식별한다. 그리고 구조적 등위성 분석은 네트워크에서 유사한 지위와 역할을 하는 노드들의 집단을 클

러스터링 한다(이수상, 2012).

본 논문에서는 1년 단위 누적적 분할 방식으로 컴포넌트 분석을 수행했다. 컴포넌트는 네트워크 안에서 최대로 연결된 하위 네트워크를 뜻하며, 컴포넌트의 모든 노드들은 직접적인 관계가 없어도 상호 연결되어있고, 다른 컴포넌트들

과는 연결되지 않는다. 즉, 데이터논문과 학술 논문 공저 네트워크에서 컴포넌트를 식별하면 직접 또는 몇 단계를 거치든 간접적으로 연결된 저자 집단이 식별된다. 여기서 네트워크 내에 연결 관계가 전혀 없는 노드들이 존재할 수 있는데, 본 논문의 분석 데이터세트는 공저 문헌만을 대상으로 하였기 때문에 하나의 노드로 구성된 컴포넌트(고립노드)는 0이다.

시기별 연결 컴포넌트 수, 가장 큰 컴포넌트의 노드 수, 링크 수, 밀도, 두 번째로 큰 컴포넌

트의 노드 수, 링크 수, 밀도를 분석한 결과를 정리하면 <표 4>와 같다.

두 문헌 유형의 동적 공저 네트워크는 모두, 네트워크가 성장할수록 연결 컴포넌트 수가 증가했다. 이는 데이터논문과 학술논문에 모두, 매년 새로운 저자들이 등장하여 누적적으로 연결 컴포넌트의 수가 증가하는 것으로 볼 수 있다. 그리고 2015년에 가장 큰 컴포넌트의 노드 수(데이터논문 노드 수 140, 학술논문 노드 수 241)와 링크 수(데이터논문 링크 수 3,145, 학

<표 4> Biodiversity Data Journal의 동적 공저 네트워크에 대한 컴포넌트 분석

구분 \ 연도		2015년	2015년 -2016년	2015년 -2017년	2015년 -2018년	2015년 -2019년	2015년 -2020년	2015년 -2021년
데이터 논문	연결 컴포넌트 수	12	22	34	41	59	90	116
	가장 큰 컴포넌트의 노드 수(%)	140 (71%)	144 (46%)	150 (37%)	161 (35%)	217 (32%)	282 (22%)	347 (20%)
	가장 큰 컴포넌트의 링크 수(%)	3,145 (95%)	3,155 (77%)	3,176 (72%)	3,242 (69%)	3,846 (61%)	4,285 (22%)	4,793 (21%)
	가장 큰 컴포넌트의 네트워크 밀도	0.323	0.306	0.284	0.252	0.164	0.108	0.080
	두 번째로 큰 컴포넌트의 노드 수(%)	10 (5%)	31 (10%)	31 (8%)	31 (7%)	31 (5%)	137 (11%)	252 (15%)
	두 번째로 큰 컴포넌트의 링크 수(%)	45 (1%)	465 (11%)	465 (11%)	465 (10%)	465 (7%)	6,326 (32%)	8,322 (36%)
	두 번째로 큰 컴포넌트의 네트워크 밀도	1	1	1	1	1	0.679	0.263
학술 논문	연결 컴포넌트 수	60	112	133	153	174	233	312
	가장 큰 컴포넌트의 노드 수(%)	241 (52%)	326 (41%)	388 (38%)	493 (40%)	622 (43%)	710 (37%)	822 (31%)
	가장 큰 컴포넌트의 링크 수(%)	12,728 (97%)	13,504 (94%)	14,030 (89%)	14,893 (90%)	15,591 (90%)	16,056 (85%)	16,562 (76%)
	가장 큰 컴포넌트의 네트워크 밀도	0.440	0.255	0.187	0.123	0.081	0.064	0.049
	두 번째로 큰 컴포넌트의 노드 수(%)	15 (3%)	18 (2%)	30 (3%)	22 (2%)	22 (2%)	36 (2%)	65 (2%)
	두 번째로 큰 컴포넌트의 링크 수(%)	34 (0.1%)	114 (1%)	435 (3%)	113 (1%)	113 (1%)	254 (1%)	487 (2%)
	두 번째로 큰 컴포넌트의 네트워크 밀도	0.324	0.745	1	0.489	0.489	0.403	0.234

술논문 링크 수 12,728)를 살펴보면, 2015년에 공저 네트워크에서 데이터논문은 전체 저자의 71%가, 학술논문은 전체 저자의 52%가 직접 또는 간접적으로 연결된 컴포넌트에 속하는 것을 의미한다. 데이터논문의 경우, 가장 큰 컴포넌트의 노드 수는 시간이 지날수록 점차 증가하나 해당 시기의 전체 네트워크 노드 수에서 가장 큰 컴포넌트의 노드 수가 차지하는 비율과 그 컴포넌트의 밀도가 점차 감소하는 추세였다. 학술논문의 경우에는 가장 큰 컴포넌트의 노드 수가 시간이 지날수록 점차 증가하나, 해당 시기의 전체 네트워크 노드 수에서 가장 큰 컴포넌트의 노드 수가 차지하는 비율 추이에는 일정 패턴이 확인되지 않고 그 밀도는 감소했다.

여기서 주목할만한 부분은 컴포넌트의 링크 수와 밀도에 따르면 데이터논문에서 2015년부터 2019년까지 두 번째로 큰 컴포넌트와 학술논문에서 2015년-2017년에 두 번째로 큰 컴포넌트의 네트워크 밀도가 1인 것이 확인됐다는 점이다. 이처럼 네트워크에서 모든 노드 쌍이 연결된 완전 그래프 형태를 갖는 하위집단을 파당이라고 한다. 이러한 파당은 하나의 문헌에 다수의 공저자들이 참여하여 게재된 문헌

형태만으로 공저 네트워크가 형성된 것이다. 현실 세계에서는 완전히 연결된 최대 하위 네트워크인 파당을 쉽게 발견하기 어렵다. 그러나 BDJ 데이터논문과 학술논문 동적 공저 네트워크에서는 파당이 식별됐다. 이는 생물다양성 분야의 공저자들이 서로 잘 알고 있는 사이로 강한 연결 관계를 맺고 있다고 해석할 수 있다. 이에 파당의 수를 1년 단위 누적적 시기별 전체 컴포넌트 수에서 차지하는 비율로 살펴보면 <표 5>와 같으며, 네트워크가 커질수록 파당 수가 증가하는 것으로 나타났다. 하지만 해당 시기별 전체 연결 컴포넌트에서 차지하는 파당의 비율은 데이터논문은 70% 이상으로 일정한 패턴은 나타나지 않았으며, 학술논문의 경우 해당 시기별 전체 연결 컴포넌트에서 차지하는 파당의 비율은 75% 이상으로 네트워크가 성장하면서 그 비율이 점차 감소하는 추세로 나타났다. 이와 같이 BDJ 데이터논문과 학술논문의 동적 공저 네트워크에서 전체 연결 컴포넌트 중 파당의 비율이 70% 이상 유지되는 것은 생물다양성 분야의 동적 공저 네트워크에서 국지적으로 응집된 저자들의 관계 구조가 시간이 경과되어도 지속되고 있는 것으로 볼 수 있다.

<표 5> Biodiversity Data Journal의 동적 공저 네트워크에 대한 파당 분석

구분 \ 연도		2015년	2015년 -2016년	2015년 -2017년	2015년 -2018년	2015년 -2019년	2015년 -2020년	2015년 -2021년
데이터 논문	연결 컴포넌트 수	12	22	34	41	59	90	116
	파당 수	9 (75%)	16 (73%)	26 (76%)	32 (78%)	47 (80%)	68 (76%)	90 (78%)
학술 논문	연결 컴포넌트 수	60	112	133	153	174	233	312
	파당 수	52 (87%)	93 (83%)	109 (82%)	124 (81%)	140 (80%)	182 (78%)	243 (78%)

4.1.4 네트워크에 내재된 속성 분석 결과

네트워크의 내재된 속성 중의 하나인 군집화 계수의 추이를 파악하여 생물다양성 분야 데이터논문과 학술논문 동적 공저 네트워크의 구조적 특징을 확인하였다. 평균 군집화 계수를 1년 단위 누적적 분할 방식으로 분석한 결과는 <표 6>과 같다.

Watts와 Strogatz(1998)는 네트워크에서의 평균 경로거리가 무작위 네트워크에서의 평균 경로거리와 비슷하고, 군집화 계수가 무작위 네트워크보다 훨씬 클 경우, 그 네트워크는 좁은 세상 네트워크의 특성을 갖는다고 했다. Watts와 Strogatz(1998)가 제시한 좁은 세상 네트워크의 특성을 파악하기 위해 각 데이터세트 별로 무작위 네트워크에서의 평균 경로거리와 Watts와 Strogatz(1998)의 군집화 계수를 산출하였다. 그 결과, 데이터논문과 학술논문 모두, 무작위 네트워크에 대한 평균 경로거리와 군집화 계수는 시간 경과에 따른 변화 없이 그 값이 각각 3.262(무작위 네트워크의 평균 경로거리), 0.542(무작위 네트워크의 평균 군집화 계수)였다.

이를 앞서 분석된 <표 3>의 평균 경로거리 결과를 포함하여 비교하면, BDJ 데이터논문 동적 공저 네트워크의 평균 경로거리는 최대 2,940(2015년-2021년)으로 무작위 네트워크 평균

경로거리인 3.262보다 작지만, 학술논문 동적 공저 네트워크의 평균 경로거리는 3.610(2015년-2019년), 3.959(2015년-2020년), 4.295(2015년-2021년)로 대상연도 2019년부터 무작위 네트워크의 평균 경로거리 3.262보다 조금 큰 수치를 지닌 것으로 확인됐다. 그리고 <표 6>과 같이 분석된 평균 군집화 계수 결과 값들은 모두 무작위 네트워크의 군집화 계수 0.542보다 더 큰 것으로 확인됐다. 이는 BDJ 데이터논문과 학술논문 동적 공저 네트워크에서 제3의 공유저자를 갖는 두 저자가 공저 관계를 형성할 확률이 무작위 기회만으로 공저 관계가 형성될 예상보다 훨씬 높다는 것을 의미한다. 그러나 두 문헌 유형의 동적 공저 네트워크의 평균 군집화 계수 수치는 시간이 흐르면서 네트워크 크기가 커지는 것에 따른 일정한 패턴은 확인되지 않았다.

분석된 결과를 바탕으로 정리하면 학술논문 동적 공저 네트워크의 경우, 대상연도 2021년에 갖는 평균 경로거리 최댓값 4.295는 무작위 네트워크 평균 경로거리 3.262보다 1.32배 컸다. 그리고 평균 군집화 계수는 데이터논문과 학술논문 동적 공저 네트워크 모두, 무작위 네트워크의 평균 군집화 계수보다 1.5배 정도 크게 나타났다.

이를 종합하면 BDJ 데이터논문과 학술논문

<표 6> Biodiversity Data Journal의 동적 공저 네트워크에 대한 평균 군집화 계수

구분 \ 연도		2015년	2015년-2016년	2015년-2017년	2015년-2018년	2015년-2019년	2015년-2020년	2015년-2021년
데이터 논문	평균 군집화 계수	0.961	0.949	0.936	0.938	0.941	0.947	0.944
학술 논문	평균 군집화 계수	0.905	0.886	0.882	0.873	0.878	0.885	0.885

동적 공저 네트워크는 서로 잘 아는, 끼리끼리 모이는 성향의 강한 결속집단들이 존재함에도 불구하고 시간의 흐름 속에서 좁은 세상 현상이 나타났고 이 구조가 유지되고 있음으로 해석할 수 있다.

4.2 시간적 공저 네트워크 중심성 분석 결과

공저 네트워크는 시점별 등장하는 저자와 형성되는 링크가 변하면서 성장하는 동적 네트워크이다. 누적적 접근해서 각 대상연도별로 TTBC 값을 산출하면 분석 데이터세트의 시작연도부터 대상연도별로 매개가 되는 저자(노드) 식별이 가능하고, 그 변화와 추세를 파악할 수 있다. 구현된 TTBC 프로그램으로 1년 단위 누적적 분할별 생물다양성 분야 데이터논문 및 학술논문에서 분석된 TTBC 값이 높은 저자 상위 20위까지를 정리하면 데이터논문은 <표 7>, 학술논문은 <표 8>과 같다.

BDJ 데이터논문 기반 시간적 공저 네트워크에서 대상연도 2019년까지 TTBC 값이 가장 높은 저자는 de Jong Yde로 확인됐다. de Jong Yde는 2015년에 공저로 7편의 문헌을 생산한 저자이다. 2015년, 2016년, 2017년까지 공저로 데이터논문을 게재하여 공저 네트워크에 등장했는데 시간이 흐르면서 새로운 공저 관계 형성에 영향을 주는 공유 저자 역할로 대상연도 2017년에 TTBC 수치가 141.4로 그 영향력이 증대된 것이 확인됐다. 이후 2019년까지 TTBC 값이 141.4로 유지되며 1위를 차지하는데, TTBC 값이 누적적으로 2017년부터 2019년까지 동일하게 유지된 것은 de Jong Yde가 2018년과

2019년에 공저 네트워크에 등장하지 않았었다는 것을 의미한다. 그러나 2020년에 de Jong Yde의 TTBC 값이 143.4로 변화된 것을 확인할 수 있다. 이는 2020년에 de Jong Yde가 새로운 데이터논문을 공저로 게재하지 않았었음에도 TTBC 값이 부여된 것으로 2점이 높아졌다는 것은 de Jong Yde를 공유 저자로 했던 두 저자가 실제로 공저 관계를 형성하여 2번의 폐쇄형 삼각관계가 이루어진 시점임을 파악할 수 있다. 실제 데이터를 살펴보면 2015년에 저자 de Jong Yde와 특정 문헌에서 함께 공저했던 저자 Forshage Mattias와 Ronquist Fredrik가 또 다른 특정 문헌에서 2015년에 저자 de Jong Yde와 함께 공저했던 저자 Holston Kevin하고 2020년에 공저 문헌을 게재하여 폐쇄형 삼각관계가 나타났다.

다음 대상연도 2020년을 변곡점으로 저자 Fattorini Simone의 TTBC 값이 149.9로 1위의 순위가 변화된 것을 확인할 수 있다. Fattorini Simone는 2015년, 2018년, 2020년에 데이터논문을 공저로 각 1편씩 생산한 저자이다. 대상연도 2018년에 TTBC 값이 53.9로 2순위로 나타났고, 대상연도 2019년까지 그 순위를 유지하다가 대상연도 2020년에 매개 영향력이 가장 높은 저자로 측정됐다. 이는 데이터논문 기반 시간적 공저 네트워크에서 Fattorini Simone가 그동안 누적적으로 지니고 있었던 TTBC 수치와 더불어 아마도 2020년에 새로운 공저 관계 형성으로 인해 새로운 매개자로 부상하게 된 것으로 해석된다. 실제로 데이터를 살펴보면, Fattorini Simone는 2015년에 de Jong Yde와 같은 컴포넌트 그룹에 속해 있었고, 1년 단위 시분할 한 2018년 공저 네트워크에서의 컴포넌

〈표 7〉 Biodiversity Data Journal 데이터문 기반 시간적 공저 네트워크의 시간적 삼각매개중심성 TTBC 값 상위 20위 저자

연도	대상연도 2016년 (2015-2016년)		대상연도 2017년 (2015-2017년)		대상연도 2018년 (2015-2018년)		대상연도 2019년 (2015-2019년)		대상연도 2020년 (2015-2020년)		대상연도 2021년 (2015-2021년)				
	구분	자차량	TTBC	자차량	TTBC	자차량	TTBC	자차량	TTBC	자차량	TTBC				
테이터논문 T T B C 값 상위 20위 위자	1	de Jong Yde	55.6	1	de Jong Yde	141.4	1	de Jong Yde	141.4	1	Fattorini Simone	149.9	1	Senzator Stepan A	210.9
		Gastauer Markus	0.3	2	Fattorini Simone	53.9	2	Fattorini Simone	53.9	2	Cardoso Pedro	148.5	2	Filippova Nina	206.8
		Mera Neto Joao A A	0.3	3	Gastauer Markus	4.3	3	Borges Paulo A V	14.5	3	de Jong Yde	143.4		Glazunov Valerii A	194.4
	2	Miazaki Angela S	0.3	4	Diniz Ecio Souza	3.2	4	Gabriel Rosalina	11.3	4	Forza Rafaela Campostrini	121.2	3	Knyazev Mikhail S	194.4
		Villet Martin H	0.3	4	Miazaki Angela S	0.3	5	Groom Quentin	8.8		Lanna Joao Monnerat	106.8		Shilov Denis S	194.4
	6	Dowling Ashley P G	0.2	5	Tahata Ryoichi	1.4	6	Fonseca Crespo Luis Carlos	8.1	5	Oliveira Felipe Alves	106.8	6	Filippov Ilya V	187
		Fisher Danielle M	0.2		Yamasaki Takeshi	1.4		Brito Mariana R	6.4	7	Borges Paulo A V	59.3	7	Cardoso Pedro	170
		Skvarla Michael Joseph	0.2	7	Mera Neto Joao A A	0.3		Kano Yuichi	6.4	8	Dalcin Eduardo C	34	8	Fattorini Simone	149.9
				8	Miazaki Angela S	0.3	7	Pimentel Cesar M M	6.4		Morim Mari P	34	9	de Jong Yde	143.4
					Viller Martin H	0.3		Yamasaki Takeshi	6.4	10	Mammola Stefano	30.8	10	Forza Rafaela Campostrini	132.1
								Forza Rafaela Campostrini	4.8	11	Gabriel Rosalina	20.9	11	Borges Paulo A V	120.8
					11	Fisher Danielle M	0.2	11	Lanna Joao Monnerat	4.8	12	Kitching Ian J	18	12	Lanna Joao Monnerat
								Skvarla Michael Joseph	0.2		Oliveira Felipe Alves	16.9	13	Oliveira Felipe Alves	106.8
										14	Fonseca Crespo Luis Carlos	15.3	14	Shiryaev Anton	60.5
								Gastauer Markus	4.3	15	Moldovan Oana Teodora	14.5	15	Nesterkova Dina V	55.5
								Barbosa Renaldo Imbrozio	3.6	16	Philippova Nina	12.2	16	Arkhipov Vladimir Yu	44.4
								Diniz Ecio Souza	3.2	17	Groom Quentin	8.8	17	Dalcin Eduardo C	34
								Nogueira Elisabete C	3.2	18	Grattarola Florencia	8.4		Morim Mari P	34
								Rumel Veljo	2.2	19	Pincheira Denoso Daniel	8.4	19	Artavev Oleg	33.6
								Pertiz Ricardo de Oliveira	1.8	20	Malmubres Olarte Jagoba	7.8	20	Mammola Stefano	30.8
							Tabata Ryoichi	1.4	20		

트 그룹 저자들과 2020년 공저 네트워크에서의 컴포넌트 저자들은 모두 각각 상이한 저자들이었으며, 그 중심에 Fattorini Simone가 공유 저자로 포함되어 있었음을 확인했다.

또 다른 변화로 대상연도 2020년에는 저자 Cardoso Pedro가 TTBC 148.5로 2015-2019년까지 1위를 차지했던 de Jong Yde의 TTBC 값인 141.4 이상으로 2순위로 식별됐다. Cardoso Pedro는 데이터논문 네트워크 영역에는 2019년에 등장했는데, 2020년에 공저로 데이터논문을 5편 생산한 것이 확인됐다. 2021년에 순위는 4위로 떨어졌지만 TTBC 값은 170으로 증가하였다.

그리고 2021년에 TTBC 값 210.9로 1위를 차지한 저자 Senator Stepan A와 TTBC 값 206.8로 2위를 차지한 Filippova Nina, 공동 3위를 차지한 저자 Glazunov Valerii A, Knyazev Mikhail S, Shilov Denis S, 다음 저자 6위 Filippov Ilya V의 데이터논문 기반 시간적 공저 네트워크 등장연도를 확인하면 Filippova Nina만 2019년 등장했고, 모두 2020년에 등장했던 저자들이다. 이들은 대상연도 2021년으로부터 가장 최근에 공저 관계를 형성한 저자들임에도 TTBC 값이 매우 높은 것을 볼 수 있다. 이는 이들이 2020년에 최대 공저자 수 112명으로 게재된 데이터논문 1편의 참여 저자들로 공저 네트워크가 형성되었기 때문이며, 2021년에도 재등장한 이 저자들이 새로운 공저 관계를 형성하여 산출된 TTBC 값으로 대상연도 2021년에 매개 영향력이 높은 저자들로 식별되었음을 확인할 수 있다. 이 결과 및 해석에 대한 유용성을 확인하기 위해 정적 네트워크 방식으로 NodeXL을 활용하여 대상연도 2021년을 대

으로 매개중심성 값을 산출하였다.

대상연도 2021년에 데이터논문의 정적 네트워크 매개중심성 값 상위 1위는 저자 Fattorini Simone와 2위는 저자 de Jong Yde 였으며, 그 뒤 순위로 3위 저자 Borges Paulo A V, 4위 저자 Cardoso Pedro, 5위 저자 Filippova Nina가 확인됐다. 그러나 대상연도 2021년의 TTBC 값 상위 1위 저자 Senator Stepan A와 공동 3위 저자 Glazunov Valerii A, 저자 Knyazev Mikhail S, 저자 Shilov Denis S는 정적 네트워크 방식으로 매개중심성 순위가 대상연도 2021년에서 각각 14위와 공동 24위를 차지하고 있다. 저자 Senator Stepan A는 대상연도 2020년에 연결정도 113에서 대상연도 2021년에 연결정도가 132로 증가하였고 개별 노드 평균 군집화 계수가 대상연도 2020년에 0.965에서 대상연도 2021년에 0.730으로 감소하였다. 그리고 저자 Glazunov Valerii A, 저자 Knyazev Mikhail S, 저자 Shilov Denis S는 대상연도 2020년에 연결정도 111에서 대상연도 2021년에 연결정도가 129로 증가했으며, 개별 노드 평균 군집화 계수가 대상연도 2020년에 1에서 2021년에 0.765로 감소했다. 이를 해석하면 이 시기에 등장한 연구 집단들의 브릿지로서 그 영향력이 증가한 것을 시간성이 추가된 TTBC 값을 산출함으로써 해당 시기에 중요성을 가진 저자(노드)가 더욱 부각 되었다.

BDJ 학술논문 기반 시간적 공저 네트워크에서는 2015년, 2016년, 2017년에 등장했던 저자 Arvanitidis Christos가 대상연도 2016년에 TTBC 값 452, 대상연도 2017년에 TTBC 값 485.8이 대상연도 2019년까지 유지되어 시간적 삼각매개중심성이 가장 높은 저자로 식별됐다.

다음으로 저자 Borges Paulo A V는 2015년부터 등장해서 2021년까지 꾸준히 학술논문을 BDJ에 게재한 저자이며, TTBC 값이 대상연도 2016년 289.5로 시작하여 2017년에 372.9, 2018년에 401.5, 2019년에 419.4로 수치가 높아지면서 2020년에 509.8, 2021년에 573으로 2020년 이후에는 학술논문 기반 시간적 공저 네트워크에서 시기별 전역중심성이 가장 높은 저자로 나타났다.

또한, 저자 Cardoso Pedro도 2016년, 2017년, 2018년, 2019년에 학술논문 기반 시간적 공저 네트워크에 등장했는데, TTBC 값도 대상연도 2017년 42.6부터 대상연도 2018년 157.9, 대상연도 2019년 254.3, 대상연도 2020년 255.2까지 수치가 증가하였다. 저자 Cardoso Pedro는 BDJ에 데이터논문과 학술논문을 여러 편 공저로 게재한 저자로 생물다양성 분야의 시간적 공저 네트워크에서 이 시기에 영향력이 높았던 것을 알 수 있다.

다음으로 저자 Penev Lyubomir은 2015년에 학술논문 기반 시간적 공저 네트워크에서 최대 연결정도 값을 갖는 저자로 2015년 학술논문 3편을 게재하였으며, 이 3편 중 2편이 학술논문 1편당 최대 공저자 수인 114명과 그 후순위인 112명으로 게재된 논문들이었다. 이후 저자 Penev Lyubomir은 2018년과 2021년에 재등장했는데, 대상연도 2016년과 2017년까지 TTBC 값이 0이었다가 대상연도 2018년에 163.8로 대상연도 2020년까지 유지되었고 대상연도 2021년에 그 수치가 235.8로 높아졌다. 데이터를 살펴보면, 2015년에 연결정도 중심이 가장 높았었고, 시간 지연이 없는 경우인 해당 시점 2015년에는 저자 Penev Lyubomir가 매개중심성


측면의 영향력이 높았던 저자임을 알 수 있다. 여기서 시간적 삼각매개중심성 산출 논리에 따르면 이 영역에서 2015년에 저자 Penev Lyubomir와 공저 관계를 형성했었던 다수의 저자들이 시간이 흐른 후, 2016년이나 2017년에 다른 공저 관계를 형성하지 않았다는 것을 추측할 수 있다. 이는 2015년에 저자 Penev Lyubomir와 공저를 했었던 다수의 저자들이 2015년 이후로 BDJ에 새로운 논문을 게재하지 않았기 때문에 네트워크에 등장하지 않았을 수도 있다. 이후 저자 Penev Lyubomir가 2018년과 2021년에 재등장하면서 형성된 공저 관계에서는 시간 경과에 따른 제3의 공유 저자 역할로서 TTBC의 값이 증대되었음을 확인할 수 있다.

다음 대상연도 2021년에 TTBC 값 180.8로 순위 10위에 나타난 저자 Telenius Anders를 살펴보면, 2015년에 학술논문 1편을 최대 공저자 수인 114명과 게재하면서 공저 관계를 형성했었지만 2020년까지 BDJ에 다른 공저 관계가 형성되지 않았다가 2021년에 공저로 새로운 학술논문 1편을 게재한 것이 확인됐다. 저자 Telenius Anders도 2021년에 재등장하면서 2015년에 공저 관계를 형성했던 저자들과 2021년에 새로운 공저 관계를 형성한 저자들 사이에 브릿지로서 그 영향력이 특정 시기에 발휘되는 것을 확인할 수 있다. 이같이 특정 시점에 공저자 수가 많은 논문 1편에 참여했던 것으로 시간 지연이 수반되어 높은 TTBC 값을 갖는 저자로 식별되는 것은 생물다양성 분야처럼 논문 1편에 매우 많은 저자들이 참여하는 특정 주제 분야에서만 나타날 수 있는 특수한 현상으로 볼 수 있다.

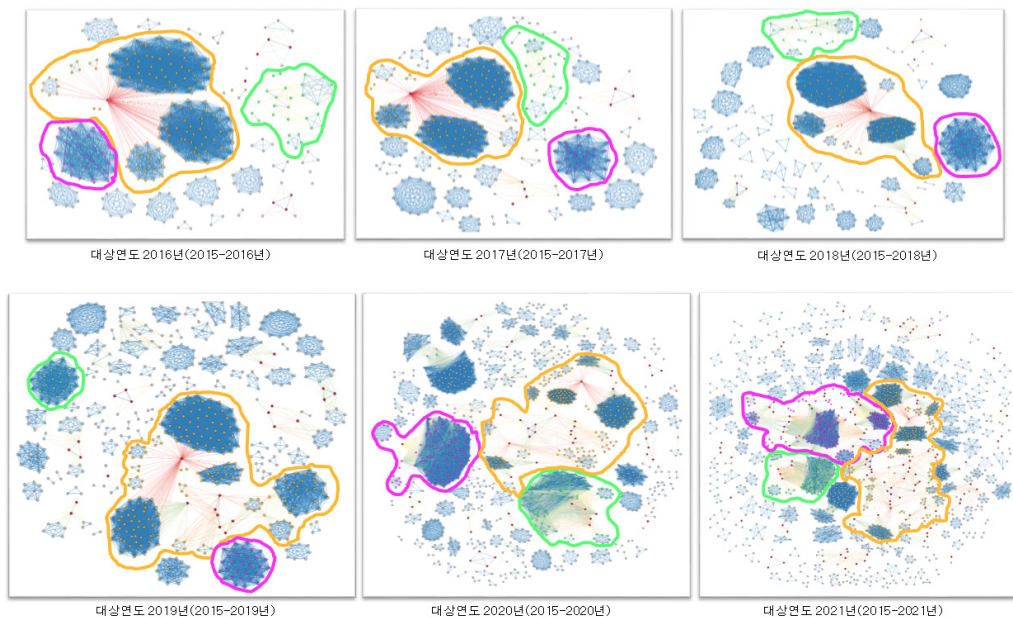
4.3 시간적 공저 네트워크 시각화 결과



4.3.1 데이터논문과 학술논문 기반 시간적 공저 네트워크의 구조적 차이

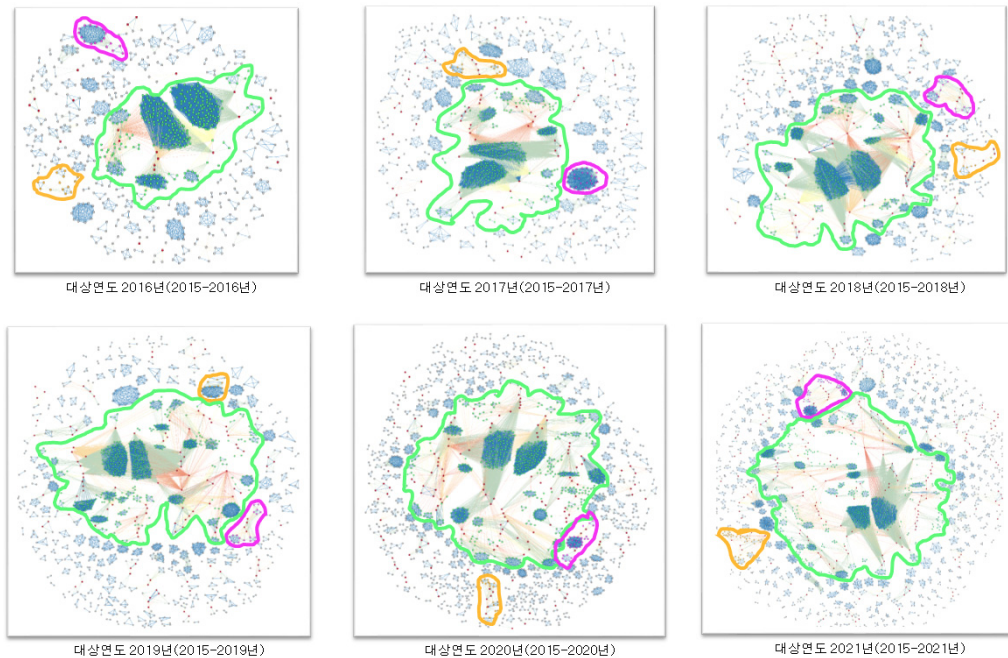
네트워크 시각화는 1년 단위 누적적 분할 방식의 동적 네트워크를 시기별 정적 네트워크 스냅샷으로 표현했다. BDJ 데이터논문과 학술논문에 기반한 시간적 공저 네트워크 지도를 대상연도별 스냅샷으로 작성했으며, 노드 색상에 컴포넌트를 반영하여 전체 컴포넌트에서 가장 큰 컴포넌트, 두 번째, 세 번째로 큰 컴포넌트까지 표시하고 나머지는 회색으로 나타냈다. 데이터논문과 학술논문을 구분하기 위해 데이터논문의 가장 큰 컴포넌트는 주황색, 두 번째-분홍색, 세 번째-연두색으로 표시하고 학술논문은 가장 큰 컴포넌트-연두색, 두 번째-분홍

색, 세 번째-주황색으로 표시했으며, TTBC 값을 지닌 노드만 빨간색으로 나타났다. 그리고 링크 가중치인 코사인 연관성 값은 링크 색상으로 반영했다. 파란색일수록 코사인 연관성이 높은 수치이고, 상아색은 중간, 빨간색은 코사인 연관성이 낮은 수치를 표현한다().

데이터논문과 학술논문의 시간적 공저 네트워크를 비교하기 위해 <그림 5>, <그림 6>과 같이 파노라마 뷰로 나타냈으며, 시간이 지남에 따라 가장 큰 컴포넌트와 두 번째로 큰 컴포넌트의 규모가 서로 다르게 나타나는 것과 같이 시각적으로 확연하게 드러나는 각 시간적 공저 네트워크의 구조적 차이를 확인할 수 있다. 정리하면, BDJ 데이터논문과 학술논문의 시간적 공저 네트워크는 모두 분절 구조를 지닌 네트워크이지만, 데이터논문 기반 시간적 공저 네



<그림 5> Biodiversity Data Journal 데이터논문 기반 시간적 공저 네트워크의 시각화
- 1년 단위 누적적 분할 파노라마 뷰(by Gephi/코사인 연관성 low  high )



〈그림 6〉 Biodiversity Data Journal 학술논문 기반 시간적 공저 네트워크의 시각화
- 1년 단위 누적적 분할 파노라마 뷰(by Gephi/코사인 연관성 low■high)

트위크는 서로 떨어져 있는 저자 집단들이 시간이 지남에 따라 각 집단별로 그 세(勢)를 넓혀나가며 비교적 크기가 큰 저자 집단들이 여러 개 나타나는 것이 확인됐다. 이에 반해, 학술논문 기반 시간적 공저 네트워크가 데이터논문 기반 시간적 공저 네트워크보다는 하나의 가장 큰 컴포넌트와 여러 작은 컴포넌트들로 구성된 구조에서 시간이 경과 하면서 점점 거대한 연결 컴포넌트가 부상하는 것으로 나타났다.

5. 결 론

네트워크 분석 영역에서 시간에 따라 변하는 패턴을 정량적으로 설명하기 위한 접근 방식이

확산되고 있다. 본 연구는 데이터와 논문을 포함하여 저자들 간 연결 패턴의 역동적인 변화를 식별하기 위해 생물다양성 분야 혼합형 데이터학술지인 BDJ를 분석 대상으로 시간적 공저 네트워크의 구조적 특성을 규명하였다. 시간적 공저 네트워크 분석을 위해 동적 공저 네트워크 분석과 시간적 공저 네트워크 중심성 분석 그리고 시간적 공저 네트워크 시각화로 구분하여 연구를 수행한 결과, 본 논문의 연구 문제에 대해 도출된 결론을 정리하면 다음과 같다.

먼저, 동적 공저 네트워크 분석을 수행한 결과 다음과 같은 구조적, 진화적 특성을 확인했다. 첫 번째로 네트워크 수준에서 분석한 결과, 생물다양성 분야의 데이터논문과 학술논문 동

적 공저 네트워크는 시간이 경과 되어도 네트워크 밀도가 증가하는 조밀한 네트워크 구조가 아닌 분절적 네트워크 구조를 유지하는 것으로 확인됐다. 두 번째로 노드 수준에서 분석한 결과, 두 문헌 유형의 동적 공저 네트워크는 모두, 시간이 지남에 따라 네트워크가 성장할수록 평균 경로거리와 직경이 증가하는 경향임을 확인했다. 또한, 생물다양성 분야의 데이터논문과 학술논문 동적 공저 네트워크는 먹힘수 분포를 따르는 무척도 네트워크 속성을 보유하고 있음을 확인했다. 세 번째로 하위집단 수준에서 분석한 결과, 두 문헌 유형의 동적 공저 네트워크는 모두, 네트워크가 성장할수록 연결 컴포넌트 수가 증가함을 확인했으며, 이는 BDJ에 매년 새로운 저자들이 등장하여 누적적으로 연결 컴포넌트 수가 증가한 것으로 해석할 수 있다. 또한, 생물다양성 분야의 데이터논문과 학술논문 동적 공저 네트워크는 하위집단이 대부분 네트워크 폐쇄 형태를 갖는 과당으로 나타났으며, 네트워크가 커질수록 과당 수가 증가했고, 전체 연결 컴포넌트에서 차지하는 과당의 비율이 70% 이상 유지됨을 확인했다. 이를 통해 강한 결속력을 지닌 저자 집단의 공저 활동이 주로 이뤄지는 것이 파악됐으며, 국지적으로 응집된 저자들의 관계 구조가 시간이 경과 되어도 지속되고 있음을 확인했다. 네 번째로 네트워크에 내재된 속성을 분석한 결과, 생물다양성 분야의 데이터논문과 학술논문 동적 공저 네트워크는 저자들의 유유상종 현상이 나타나고 좁은 세상 네트워크의 특성을 보유하며, 네트워크가 진화하는 동안에 이 구조적 특성이 유지되고 있음을 확인했다.

다음으로 본 연구에서 구현한 TTBC를 활용

하여 시간적 공저 네트워크 중심성 분석을 수행한 결과, 산출된 TTBC 값을 통해 시점별 등장하는 저자들과 새로운 공저 관계를 형성해서 그들 사이에 구조적 및 시간적 매개가 되는 저자들이 식별됐다. 그리고 생물다양성 분야의 데이터논문과 학술논문 기반 시간적 공저 네트워크에서 시간에 따른 중요한 저자들의 변화 추이가 파악됐다. 또한, 저자별 TTBC 값의 변화를 살펴보면 시간 지연이 수반되어 실질적으로 공저 관계 형성에 영향을 미쳤을 때 그 값이 상승했고 그 시점이 파악됐다. 이러한 분석 결과는 TTBC가 시간이 지남에 따라 상황이 변화하고 그 중요성과 영향력을 지닌 노드가 달라지는 것을 확인할 수 있게 하고, TTBC 값의 수치적 변화를 통해 노드의 실질적 매개 영향력이 발휘되는 시점을 확인할 수 있게 한다는 것을 보여준다. 즉, TTBC는 공저 네트워크뿐만 아니라 시간이 경과 하면서 진화하는 네트워크에서 최초 연결 시점을 파악하여 분석 대상, 단위를 설정해서 바로 적용할 수 있는 시간 순서 영향이 고려된 새로운 측면의 전역중심성이다. 이는 결과적으로 관계 형성 시간 순서가 반영되었기 때문에 흐름 역학과 위상(토폴로지) 역학의 영향력 측면에서 식별되는 전역중심성 척도라 할 수 있다.

마지막으로 시기별 정적 네트워크 스냅샷으로 동적 네트워크를 시각화 한 결과, 생물다양성 분야 데이터논문과 학술논문의 시간적 공저 네트워크에서 드러나는 구조적 차이가 확인됐다. 두 문헌 유형의 시간적 공저 네트워크는 모두, 분절적인 네트워크를 구성했지만, 학술논문 기반 시간적 공저 네트워크가 데이터논문 기반 시간적 공저 네트워크보다는 가장 큰 컴포넌트

에 연결의 연결을 통해 이어져 있는 형태를 보였다. 그러나 BDI 데이터논문 기반 시간적 공저 네트워크는 진화하면서 여러 개의 비교적 크기가 큰 컴포넌트들이 존재하는 것으로 나타났다. 이는 특색있는 사항으로 데이터를 다루는 연구자 집단의 협업 구조에 대한 특성으로 볼 수 있다.

시간적 네트워크에 관한 연구들은 네트워크 과학, 물리학, 컴퓨터 과학, 신경과학, 통계학, 생태학, 응용 수학 등 여러 연구 영역에서 다각적 측면으로 수행되고 있다. 본 연구에서 개발한 시간적 삼각매개중심성 TTBC는 계량정보학 영역에서 시간 역학을 모델링하기 위한 초기 진입점으로 서지 데이터세트에 대한 시간적 네트워크 분석의 새로운 현대적 기술이 될 수 있을 것으로 기대한다. 또한, 본 연구 결과에서

는 분석 대상을 데이터학술지로 하여 데이터논문과 학술논문 저자들 사이에 동적인 연구 협업 패턴을 시간적 공저 네트워크의 구조적 차이로 발견했다. 이는 각 문헌 유형과 주제 분야가 갖는 특성이 네트워크 분석 결과에 반영되어 뚜렷하게 구별되는 협업 양상을 보였지만, 본 연구는 분석 대상을 생물다양성 분야의 데이터학술지 1종으로 하여 그 현상을 파악하였다는 제한점이 있다. 향후, 본 연구의 후속연구로 분석 범주와 대상을 확장하여 시간적 공저 네트워크 분석을 수행한다면 주제 분야별 시간적 공저 네트워크의 구조적 특성을 규명하고, 그 결과를 비교할 수 있을 것이다. 그리고 거시적 측면에서 분석 대상별, 분석 단위별로 연구를 수행한다면 과학계 동적 협업 네트워크의 구조적 특성을 파악할 수 있을 것이다.

참 고 문 헌

- 김지현, 서선경, 이정연, 안채영, 이승은 (2019). 데이터 저널 현황 조사 및 발간정책 수립. 한국과학기술정보연구원.
- 서선경 (2022). FCSW 프로그램 (버전 0.1) [컴퓨터 소프트웨어].
- 서선경 (2023). TTBC 프로그램 (버전 0.1) [컴퓨터 소프트웨어].
- 이수상 (2010). 공저 네트워크 분석에 관한 기초연구: 문헌정보학 분야 4개 학술지를 중심으로. 한국도서관·정보학회지, 41(2), 297-315. <https://doi.org/10.16981/KLISS.41.2.201006.297>
- 이수상 (2012). 네트워크 분석 방법론. 서울: 논형.
- 이재운 (2006). 계량서지적 네트워크 분석을 위한 중심성 척도에 관한 연구. 한국문헌정보학회지, 40(3), 191-214. <https://doi.org/10.4275/KSLIS.2006.40.3.191>
- 이재운 (2013). COOC 프로그램 (버전 0.4) [컴퓨터 소프트웨어].
- 정은경 (2019). Scientific Data 학술지 분석을 통한 데이터 논문 현황에 관한 연구. 정보관리학회지, 36(1), 117-135. <https://doi.org/10.3743/KOSIM.2019.36.1.117>
- Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution

- of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, 311(3), 590-614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)
- Blonder, B., Wey, T. W., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, 3(6), 958-972. <https://doi.org/10.1111/j.2041-210X.2012.00236.x>
- Chen, Y. (2016). An analysis of characteristics and structures embedded in data papers: A preliminary study. *Libellarium*, 9(2), 145-156. <https://doi.org/10.15291/libellarium.v9i2.266>
- Chen, Y., Börner, K., & Fang, S. (2013). Evolving collaboration networks in scientometrics in 1978-2010: A micro-macro analysis. *Scientometrics*, 95(3), 1051-1070. <https://doi.org/10.1007/s11192-012-0895-2>
- Farine, D. R. (2018). When to choose dynamic vs. static social network analysis. *Journal of Animal Ecology*, 87(1), 128-138. <https://doi.org/10.1111/1365-2656.12764>
- Holme, P. & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97-125. <https://doi.org/10.1016/j.physrep.2012.03.001>
- Holme, P. & Saramäki, J. (2019). A map of approaches to temporal networks. *Temporal Network Theory*, 1-24. <https://doi.org/10.48550/arXiv.2103.13615>
- Kempe, D., Kleinberg, J., & Kumar, A. (2002). Connectivity and inference problems for temporal networks. *Journal of Computer and System Sciences*, 64(4), 820-842. <https://doi.org/10.1006/jcss.2002.1829>
- Kim, J., Tao, L., Lee, S., & Diesner, J. (2016). Evolution and structure of scientific co-publishing network in korea between 1948-2011. *Scientometrics*, 107(1), 27-41. <https://doi.org/10.1007/s11192-016-1878-5>
- Masuda, N. & Lambiotte, R. (2016). *A Guide to Temporal Networks Complexity Science Vol. 4*. London: World Scientific Publishing. <https://doi.org/10.1142/q0033>
- Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G., & Latora, V. (2013). Graph Metrics for Temporal Networks. In *Temporal Networks*. Berlin: Springer, 15-40. https://doi.org/10.1007/978-3-642-36461-7_2
- Savić, M., Ivanović, M., & Jain, L. C. (2019). Co-authorship Networks: An Introduction. In *Complex Networks in Software, Knowledge, and Social Systems*. Intelligent Systems Reference Library. New York: Springer International Publishing, 148. https://doi.org/10.1007/978-3-319-91196-0_5
- Seo, S. & Kim, J. (2020). Data journals: Types of peer review, review criteria, and editorial committee members' positions. *Science Editing*, 7(2), 130-135. <https://doi.org/10.6087/kcse.207>

- Sizemore, A. E. & Bassett, D. S. (2018). Dynamic graph metrics: Tutorial, toolbox, and tale. *NeuroImage*, 180, 417-427. <https://doi.org/10.1016/j.neuroimage.2017.06.081>
- Tabassum, S., Pereira, F. S. F., Fernandes, S., & Gama, J. (2018). Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, 8(5), e1256. <https://doi.org/10.1002/widm.1256>
- Thompson, W. H., Brantefors, P., & Fransson, P. (2017). From static to temporal network theory: Applications to functional brain connectivity. *Network Neuroscience*, 1(2), 69-99. https://doi.org/10.1162/NETN_a_00011
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442. <https://doi.org/10.1038/30918>
- Weingart, S., Guo, H., Börner, K., Boyack, K. W., Linnemeier, M. W., Duhon, R. J., Phillips, P. A., Tank, C., & Biberstine, J. (2010). *Science of Science (Sci²). Tool User Manual, Version Alpha 3*. Indiana: Indiana University, 2011. [Retrieved January, 15.]

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Chung, EunKyung (2019). An investigation on scientific data for data journal and data paper. *Journal of the Korean Society for Information Management*, 36(1), 117-135. <https://doi.org/10.3743/KOSIM.2019.36.1.117>
- Kim, JiHyun, Seo, SunKyung, Lee, JungYeoun, Ahn, ChaeYoung, & Lee, SeungEun (2019). A Study on the Current Status of Data Journals and the Establishment of Data Journal Policies. Korea Institute of Science and Technology Information.
- Lee, JaeYun (2006). Centrality measures for bibliometric network analysis. *Journal of the Korean Society for Library and Information Science*, 40(3), 191-214. <https://doi.org/10.4275/KSLIS.2006.40.3.191>
- Lee, JaeYun (2013). COOC program (ver 0.4) [Computer Software].
- Lee, SooSang (2010). A preliminary study on the co-author network analysis of Korean library & information science research community. *Journal of Korean Library and Information Science Society*, 41(2), 297-315. <https://doi.org/10.16981/KLISS.41.2.201006.297>
- Lee, SooSang (2012). *Network Analysis Methods*. Seoul: Nonhyung.
- Seo, SunKyung (2022). FCSW program (ver 0.1) [Computer Software].
- Seo, SunKyung (2023). TTBC program (ver 0.1) [Computer Software].

