



## 한국 언론의 동조화된 파편화

뉴스 정보원 순위별 파편화 지수 분석

박대민 | 신문대학교 미디어커뮤니케이션학부 조교수·커뮤니케이션연구소 소장

### Analyzing Synchronized Fragmentation in Korean Media\*

Using the Ranked Fragmentation Index of News Sources

Daemin Park\*\*

(Assistant Professor, School of Media and Communication, Sunmoon University; Director, Communications Research Institute)

This study explores the dual perspectives of media synchronization and fragmentation, where some scholars criticize media for aligning with political factions, while others suggest that fragmentation intensifies social conflicts. The study provides a more nuanced view of the structural dynamics of media ecosystems by proposing the idea of "synchronized fragmentation," which shows that media can simultaneously display both synchronization and fragmentation. To examine this phenomenon, trends in synchronization and fragmentation in the utilization of news sources were analyzed across 19 South Korean media outlets, including 11 national and 8 economic newspapers, from 2000 to 2020. The analysis focuses on sources cited in political, economic, social, cultural, and international news sections. News sources network analysis was conducted to rank the sources based on their degree centrality, and a fragmentation index was proposed to assess the redundancy in the use of sources across different newspapers, taking their rank into account. The fragmentation index assesses the degree of synchronization and fragmentation in source utilization across media outlets by calculating the ratio of the total number of duplicated and unique sources by rank. The Korea Press Foundation's BigKinds database was used to gather the data, and 16,538,668 articles with 66,867,349 citations were found. The findings reveal a "synchronized fragmentation" phenomenon, where both synchronization and fragmentation coexist in the use of news sources. Top-ranked sources show synchronization across outlets, while mid- to low-ranked sources are more fragmented, highlighting an overall dominance of fragmentation. The fragmentation index showed five distinct

---

\* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임)(NRF-2021S1A5A8069213).

\*\* [dmpark@sunmoon.ac.kr](mailto:dmpark@sunmoon.ac.kr)

phases: the first phase had strong synchronization among top-ranked sources, while the second phase saw intensified synchronization with the fragmentation index reaching its minimum. In the third phase, fragmentation increased, leading to a gradual relaxation of synchronization. The fourth phase exhibited continued fragmentation among lower-ranked sources, resulting in a peak fragmentation index. In the fifth phase, fragmentation slightly decreased. This framework of synchronized fragmentation was observed across different news sections in general. The politics section showed the strongest synchronization, while the culture section had the most fragmentation. Both the economy and society sections exhibited strong synchronization in top-ranked sources but experienced rapid fragmentation in mid- to low-ranked sources. The international section had moderate fragmentation overall. The study suggests that while synchronization is evident among top-ranked sources in South Korean media, the media ecosystem is largely pluralistic, with mid- to low-ranked sources significantly influencing its structure. The fragmentation index introduced in this study offers a valuable measure for understanding the non-linear relationship between synchronization and fragmentation in media ecosystems.

**Keywords:** synchronized fragmentation, fragmentation index, mean fragmentation index, news source network, Korean media ecosystem

## 국문초록

사회적으로 정파적 견해에 따라 언론계가 반대 진영에 동조화돼 있다고 비난하거나 반대로 언론계가 파편화돼 사회 갈등을 조장한다고 지적하는 견해가 있다. 이론적으로도 언론의 동조화와 파편화에 대한 연구가 양립한다. 본 연구는 언론이 동조화되어 있으면서도 파편화돼 있는 동조화된 파편화된 상태를 보여 준다.

이 연구는 2000년부터 2020년까지 한국의 전국지 11개와 경제지 8개 등 총 19개 매체의 정치면, 경제면, 사회면, 문화면, 국제면 기사에서 인용된 정보원을 대상으로 정보원 활용의 동조화와 파편화 경향을 분석하였다. 뉴스 정보원 연결망 분석을 활용해 정보원들의 연결 정도 중앙성을 기준으로 순위를 매긴 뒤, 순위를 고려한 정보원 활용 중복 정도를 파악할 수 있는 파편화 지수(fragmentation index)를 제안했다. 파편화 지수는 정보원의 순위에 따라 가능한 중복된 정보원 수와 누적된 중복되지 않은 정보원 수의 비율로 측정하며, 매체 간 정보원 활용의 동조화 정도와 파편화 정도를 평가한다. 분석 대상으로 한국언론진흥재단 빅데이터를 통해 총 16,538,668건의 기사에서 66,867,349개의 인용문을 수집했다.

분석 결과는 매체 간 정보원 활용에서 동조화와 파편화가 동시에 나타나는 '동조화된 파편화' 현상을 보여 준다. 즉, 상위권 정보원들은 매체 간 동조화되어 있지만, 중하위권에서는 파편화가 더 두드러지게 나타났으며 전체적으로는 파편화가 우세하다. 구체적으로는 파편화 지수 추이를 5개 구간으로 나눠 관찰할 수 있었다. 1구간은 최상위권 구간으로 1위 정보원 간의 동조화가 강하게 나타나는 구간이다. 2구간은 파편화 지수가 최소값에 도달하는 구간으로 동조화가 심화된다. 3구간에서는 파편화 지수가 증가하며 동조화가 완화되는 경향이 뚜렷하게 나타난다. 4구간은 하위 정보원까지 지속적으로 파편화가 진행됨에 따라 파

편화 지수가 최대값에 도달한다. 마지막 5구간에서는 파편화가 다소 완화되는 경향이 관찰됐다. 이는 동조화된 파편화의 일반 모델로 제시했다. 지면별로는 정치면의 동조화가 전반적으로 강한 반면 문화면은 파편화가 가장 강했다. 경제면과 사회면은 상위권에서는 동조화되지만 중하위권에서 파편화가 빠르게 진행됐다. 국제면은 파편화 정도가 중간 수준이었다.

동조화된 파편화 현상은 한국 언론이 상위권 정보원에서는 동조화됐지만 전체적으로는 다원적 언론 생태계를 형성하고 있으며 여기에는 중하위권 정보원의 역할이 중요함을 시사한다. 이 연구에서 제안한 파편화 지수는 매체 간 차이를 효과적으로 측정하는 지표로 활용될 수 있으며 동조화와 파편화의 비선형적 관계를 이해하는 데 기여할 것으로 기대된다.

핵심어 : 동조화된 파편화, 파편화 지수, 평균 파편화 지수, 뉴스 정보원 연결망, 한국 언론 생태계

# 1. 문제제기

언론 보도는 동조화되어 있는가 파편화되어 있는가? 예컨대 정치인들이 언론이 진보에 치우친 ‘가짜뉴스’를 생산한다는 주장이나, 언론이 보수 진영의 ‘애완견’이라고 주장하는 것도 언론이 정파적으로 한쪽 진영에 쏠려 있는 동조화(synchronization) 현상을 비판하는 것으로 볼 수 있다. 반면 언론이 정파적으로 분열돼 있으며 사회적 갈등을 야기하고 있다고 본다면 이는 언론의 파편화(fragmentation)를 비판하는 것이라고 볼 수 있다.

동조화나 파편화와 같은 용어는 복잡계 물리학(complex physics)에서 흔히 사용되는 용어이다. 특히 사회물리학의 하나인 의견동역학(opinion dynamics)을 통해 커뮤니케이션 연구와도 연결된다. 본 연구가 관심을 갖는 언론의 동조화는 간단히 말해 여러 언론사가 유사한 의제를 유사한 관점과 논조로 보도하는 현상이라고 볼 수 있다. 반대로 언론의 파편화는 언론사들이 서로 다른 이야기를 하는 상황을 말한다. 그러나 언론의 동조화와 파편화를 판단하는 것은 생각보다 복잡하다. 우선 언론의 동조화나 파편화는 개별 언론 수준이 아니라 전체 언론계 수준에서 파악해야 하는 현상이다. 그렇다면 소수 언론이 모두 정부의 통제를 받는 경우를 제외하면, 수많은 매체가 존재하는 상황에서 완벽한 언론 동조화는 가능하지 않을 것이다. 국내만 해도 2022년 기준 언론 사업체는 5,774개에 달하는데(한국언론진흥재단, 2023), 이들 언론사가 완전히 동조화될 것으로 보기는 어렵다. 게다가 언론사들이 모든 시기에, 모든 사안에 동조화되리라고는 더 더욱 생각하기 어렵다. 그럼에도 불구하고 언론의 동조화 경향은 여전히 지속적이고 광범위하게 일어날 수 있다. 예컨대 언론이 다른 언론사의 기사나 출입처의 보도자료를 베껴 쓰는 처널리즘(churnalism) 등은 언론 동조화를 야기하는 강력한 힘이다(Davies, 2008).

언론의 동조화와 파편화 문제는 언론사 간 의제나 프레임의 일치와 불일치와 같은 문제로 볼 수 있다. 그러나 의제나 프레임 못지않게 중요한 것이 정보원이다. 비록 정보원의 불일치가 의제나 프레임의 불일치를 의미하는 것은 아닐 수 있다. 그러나 미디어 편향 연구에서 정보원 편향이 중요한 것은 정보원이 기사 작성의 핵심 지표이기 때문이다(Bennett, 1990; Brown et al., 1987). 의제가 동조화돼 있다고 해서 반드시 유사한 정보원을 쓸 필요는 없지만, 꾸준하게 유사한 정보원을 활용하고 있는 언론사들은 내용적으로도 동조화돼 있을 가능성이 크다. 반대로 언론사들이 내용적으로 파편화돼 있다면 서로 다른 정보원을 발굴하지 않을 수 없을 것이다.

본 연구는 소수의 주요 언론이 아니라 국내 언론계를 주도하는 주요 전국지 11개와 경제지 8개 등 총 19개 매체를 대상으로 개별 언론이 아닌 언론계 수준에서 언론의 동조화/파편화 정도를 파악하고자 한다. 또한 언론의 동조화/파편화를 특정 시기나 특정 의제에 한정해 분석하는 대

신, 전체 기사 수준에서 거시적으로 파악하고자 한다. 즉 본 연구에서는 수집 대상 매체의 2000년부터 2020년까지 21년치 정치면, 경제면, 사회면, 문화면, 국제면의 기사 속 인용문을 한국언론진흥재단 빅키인즈(BigKinds)를 통해 전수 수집한 뒤, 전체 기사를 연도별, 지면별로 나누고, 전체 기사 수준, 지면별 수준, 그리고 연도 및 지면별 수준으로 분석할 것이다. 특히 본 연구에는 정보원 측면에서 언론의 동조화/파편화를 파악한다. 이에 따라 분석 대상은 총 16,538,668건의 기사 내 66,867,349개의 인용문에서 추출한 정보원이다.

이러한 동조화/파편화 정도는 매체 간 정보원의 일치 여부로 분석한다. 특히 정보원을 순위화하여 상위권, 중위권, 하위권 정보원들의 중복 정도와 이들이 동조화/파편화에 미치는 영향을 살펴보고자 한다. 이를 위해 뉴스 정보원 연결망 분석을 활용해 정보원의 인용 패턴을 위상학적으로 표현하고, 정보원들의 연결정도 중앙성(degree centrality)을 구하여 이를 기준으로 순위를 매긴 뒤, 순위별로 정보원 활용 중복 정도를 파악하는 파편화 지수(fragmentation index)를 제안할 것이다. 파편화 지수는 각 순위에 따라 완전히 중복되지 않았을 때 가능한 정보원 수와 누적된 중복되지 않은 정보원 수의 비율로 측정한다. 파편화 지수를 통해 전체 언론 체계 수준에서 매체 간 정보원 활용의 동조화 또는 파편화 정도를 평가한다.

동조화와 파편화 문제는 방법론적, 이론적, 실천적 관점에서 중요한 문제다. 여기서는 방법론적 측면과 이론적 측면을 간단히 서술하고, 실천적 측면은 결론에서 다루도록 한다. 우선 방법론적 문제를 살펴보자. 흔히 언론학에서는 내용분석을 통한 매체 간 비교 연구를 수행하는 경우가 많다. 통상 매체 간 비교 연구는 비슷한 시기의 특정 의제에 대한 정보원이나 주제어 등을 대조해 매체 간 차이를 밝힌다. 이때 모든 정보원이나 주제를 분석할 수 없으므로, 중요도 순으로 상위권에 속한 것끼리 대조한다. 그런데 시기가 비슷하고 의제가 동일할 경우, 정보원이나 주제어에 어느 정도 매체 간 동조화가 나타날 수밖에 없다. 그렇다면 매체 간 비교 연구에서 중요한 것은 유의미한 차이를 발견할 수 있는 순위, 즉 동조화가 아닌 파편화가 두드러지는 순위 구간이 어디인가를 파악하는 것이 된다.

다음으로 이론적 문제를 살펴보자. 어떤 연구는 언론의 동조화를 강조하는가 하면 다른 연구는 파편화를 우려한다. 이렇게 동조화와 파편화가 혼재해 있다면, 그 양상은 크게 두 가지 경우가 있을 수 있다. 첫째, 보편성이 없는 경우다. 즉 시기별, 지면별, 사안별로 동조화가 강한 경우와 그렇지 않은 경우가 나뉜다는 것이다. 그렇다면 언론의 동조화와 파편화에 대한 보편 이론은 성립되지 않으며, 사례 연구만 가능할 것이다. 둘째, 보편성이 있는 경우다. 즉 시기와 지면을 막론하고 거시적으로는 항상 동조화와 파편화가 동시에 나타난다는 상태를 의미한다. 이러한 상태는 파편화가 강한 동조화된 파편화, 또는 동조화가 강한 파편화된 동조화로 개념화할 수도 있

을 것이다.

결론부터 말하면 본 연구는 언론계가 동조화되어 있는 동시에 파편화된 상태, 특히 동조화된 파편화 상태임을 보여준다. 분석 결과 구체적인 구간별 순위 범위는 연도나 지면에 따라 달라질 수 있지만, 전반적으로는 파편화 경향이 강한 동조화된 파편화 상태를 보인다. 상위권의 소수 정보원 간에는 동조화가 일어나지만, 압도적 다수를 차지하는 중하위권 정보원에 의해 파편화가 진행된 결과다. 동조화나 파편화 중 어느 한쪽만 발견하는 것은 전체 정보원이 아니라 상위권 정보원만 분석했기 때문이다. 방법론 측면에서는 대다수의 매체 간 비교 연구가 25위권 이내를 비교하지만, 차이를 보려면 오히려 그 이후 구간을 봐야 한다는 것을 의미한다. 25위권 이내는 동조화가 두드러지는 구간이기 때문이다. 25위권 이내를 분석할 경우는 개별 매체 수준의 비교 연구가 아닌, 여러 매체를 종합한 언론 체계 수준의 분석을 해야 한다.

## 2. 기존문헌 검토

본 연구는 언론 보도의 동조화와 파편화 현상을 정보원 인용의 동조화와 파편화라는 측면에서 살펴본다. 정보원 인용의 동조화와 파편화 현상은 연결망 분석을 통해 위상학적으로 살펴보고자 한다. 이 절에서는 우선 언론 보도의 동조화와 파편화 현상과 정보원 인용의 동조화와 파편화 현상에 대한 기존 문헌을 검토한다. 이어 정보원 인용의 동조화와 파편화를 실증적인 방식으로 분석하는 틀로서 의견동역학의 동조화와 파편화 논의를 논의한다.

### 1) 언론 보도의 동조화와 파편화

언론 보도가 동조화되어 있는지 파편화되어 있는지에 대해서는 여러 이론이 대립하고 있다. 먼저 동조화 관점으로는 프랑크푸르트학파(Frankfurter Schule)의 문화산업론이나 프로파간다 모델(propaganda model)까지 거슬러 올라갈 수도 있다. 이에 따르면 언론은 정부나 기업의 영향 아래 이데올로기적으로 동조화된(Herman & Chomsky, 1988; Horkheimer & Adorno, 1944/2001).

이후 데이터에 의한 실증적 연구가 제시됐다. 언론이 정부나 대기업 등 공식적 정보원에 편향돼 있다는 연구도 동조화 경향을 강조한다(Brown et al., 1987). 언론사가 다른 언론사나 통신사의 기사 등을 거의 베껴 쓰는 처널리즘이나 보도자료에 의존하는 현상, 출입처 관행 등도 언론계의 동조화 현상을 시사한다(Boczowski, 2010; Davies, 2008; Lewis et al., 2008).

이는 넓은 의미의 정보원 수준의 분석이라 할 수 있다. 의제 수준의 분석도 있다. 매체 간 의제 설정 이론은 주요 매체의 보도가 다른 언론사에게 확산되면서 언론사 간 동조화 역할을 설명한다(Reese & Danielian, 1989). 인지 수준의 분석도 있다. 침묵의 나선 이론은 언론 보도가 의견 기후를 형성하고, 대중이 의견 기후를 감지해 침묵하고, 이것이 언론 보도에 영향을 줘서 동조화를 야기하는 피드백 과정을 경고한다(Noelle-Neumann, 1993). 프레임 연구에서도 에피소드 프레임(episodic frame)과 같은 전형적인 프레임이 언론에 통용되고 있음을 비판한다(Iyengar, 1991). 동조화는 명확한 정보 전달 및 사회적 공감대를 형성하는데 도움을 줄 수 있지만, 언론의 편향을 야기하고 민주적 토론 및 권력 비판 기능을 약화시킬 우려도 있다.

과편화 관점도 만만치 않다. 우선 대부분의 매체 간 비교 연구는 매체들 간의 차이를 강조한다. 우선 정보원 활용이 매체별로 다를 수 있다(이종혁, 2022). 상당수의 뉴스 프레임 연구는 보수지와 진보지가 각각 정치적 색깔에 맞는 프레임을 사용하는 프레임 편향(frame bias)을 지적한다(Entman, 2003). 언론에 대한 선택적 노출(selective exposure), 양극화(polarization), 반향실 효과(echo chamber effect), 필터 버블(filter bubble) 등에 대한 연구는 과편화된 언론 및 미디어 지형과 그에 따른 인지적 과편화를 시사한다(Baum & Groeling, 2008; Iyengar & Hahn, 2009; Prior, 2007; Stroud, 2011; Pariser, 2011; Sunstein, 2001). 이러한 과편화는 사용자의 확증 편향을 강화하고, 의견 양극화를 초래하며, 정보 과부하나 사회 분열을 초래할 수 있다. 다만 최근에는 선택적 노출, 양극화, 반향실 효과, 필터 버블 효과 등이 알려진 것만큼 크지 않다는 연구들도 나오고 있다(Dubois & Blank, 2018; Levendusky & Malhotra, 2016; Ross Arguedas et al., 2022; Stroud, 2010). 즉 언론을 비롯한 미디어 지형의 과편화와 미디어 플랫폼의 알고리즘은 때로는 미디어 다양성을 증진시키며 다수 사용자가 다양한 미디어를 접함으로써 양극화나 반향실 효과, 필터 버블 효과 등을 완화할 수 있게 한다. 이러한 반론은 두 가지 사실을 함축한다. 첫째, 사용자에게 양극화가 나타나든 나타나지 않든, 최소한 현재 미디어 생태계는 상당 수준으로 과편화되어 있다. 둘째, 이러한 미디어 과편화는 반드시 나쁜 결과만 초래하는 것은 아니며, 공론장의 다원성을 증진하고 사용자의 정보 선택권을 확대하며 다양한 관점을 가질 수 있는 기회를 제공한다(Napoli, 2011).

종합하면 언론의 동조화 또는 과편화는 여러 수준에서 이뤄진다고 볼 수 있다. 첫째, 정보원 수준이다. 정보원 연구가 이에 해당한다. 둘째, 의제 수준이다. 의제 설정 이론이 대표적이다. 셋째, 인지적 수준이다. 예컨대 프레임은 언론사가 구성해 사용자의 인지에 영향을 준다. 각 수준의 동조화와 과편화 경향은 꼭 일치하는 것은 아니다. 예컨대 언론사 간 정보원이 동조화된 경

우, 의제 수준에서도 동조화될 가능성이 크겠지만, 언론사 간 정보원이 파편화돼 있더라도 정보원들이 유사한 입장을 가진 경우 의제 수준에서 동조화될 수도 있다. 동조화와 파편화 경향이 정보원 수준과 의제 수준에서 다를 수 있다는 것은 두 수준의 분석 수준이 다름을 반증한다. 그러나 큰 의제는 동조화되더라도 정보원이 다르다면 인용문 내용은 달라질 가능성이 크다. 즉 큰 의제는 유사하지만, 각론으로 들어가면 들어갈수록 세부 인용문 주제어는 달라질 수도 있다. 이는 의제 수준과 다른 주제어 수준이 있을 수 있음을 의미한다. 중요한 점은 정보원 수준과 의제 수준, 그리고 주제어 수준은 다른 수준이며, 따라서 동조화와 파편화 경향은 각 수준에서 각각 연구되어야 한다는 것이다. 본 연구는 우선 정보원 수준의 동조화와 파편화를 중심으로 살펴본다. 다른 수준의 분석은 후속 연구를 통해 수행한다. 기존 연구는 각 수준에서 언론사의 동조화 경향과 파편화 경향을 모두 시사하며 그 효과는 긍정적 측면과 부정적 측면을 모두 갖는다.

## 2) 정보원의 동조화와 파편화

정보원의 인용은 르포와 같은 직접 관찰, 데이터 등 수치 제시와 더불어 주요 사실성 기제이다 (van Dijk, 1988). 이중 인용은 가장 널리 활용된다. 인용은 실명과 익명 모두 가능하지만 실명 정보원 인용이 원칙이다. 익명 인용은 정보원 보호 등 제한적으로 사용될 것이 권장된다 (Culbertson, 1980; Swain & Robertson, 1995). 또한 언론에서 인용은 직접 인용으로 표기되는 것이 원칙으로 큰 따옴표로 인용 부분이 명시된다. 즉 언론에서 인용된 정보원은 최소한 기관명이 제시된 정보원으로 중요할 뿐만 아니라 다수이며 직접 인용된다. 이에 본 연구에서는 직접 인용됐으며 개인 또는 기관 이름이 등장하는 정보원을 분석한다.

다양한 정보원 연구들은 매체 유형을 막론하고 정부 관료나 기업 임원 등 공식 정보원에 편향됐다고 지적한다. 지표 가설(indexing hypothesis)에 따르면 언론은 공식 정보원 의견을 지표(index)로 기사를 작성한다(Bennett, 1990). 물론 언론은 공식 정보원에 대항하는 대안 정보원을 발굴하기도 한다(Entman, 2003; Harp et al., 2010). 정보원 인용은 매체 성향에 따라 다르게 나타나기도 한다. 예컨대 보수 매체는 보수적 정보원에, 진보 매체는 진보적 정보원을 중시한다(이종혁, 2022). 다만 이러한 편향은 일방적인 형태로 나타나지는 않는다. 예컨대 언론사는 성향에 따라 편향된 정보원 외에 중립적인 정보원도 주변적으로나마 배치한다(한동섭·임종수, 2001). 또다른 연구에서는 언론사가 성향에 무관하게 공식 정보원에 크게 의존하지만, 진보지는 시민이나 사회, 종교 단체 정보원을 적극 발굴하기도 한다(박대민, 2015; 박주현, 2020). 정리하면 대체로 보수 매체나 진보 매체 모두 공통된 공식 정보원에 의존하는 동조화가 나타나지만, 주변 정보원은 매체의 성향을 반영하는 파편화 경향이 나타나는 것으로 보인다. 다만 기존



연구는 특정 시기, 소수 매체의 특정 주제에 관련된 비교적 적은 수의 기사를 대상으로 분석돼 일반화 가능성 측면에서 한계가 있다.

### 3) 언론 보도에 대한 위상학적 분석

위상학(topology)은 어떤 도형의 위상을 연구하는 학문이다. 위상학은 척도를 고려하지 않는 경우 즉 도형이 늘이기, 비틀기, 구부리기 등과 같은 연속적 변형(continuous deformations)을 가한 경우에도 도형의 불변하는 특징을 분석한다. 그래프 위상학(graph topology)에서 연결망은 꼭지점에 해당하는 노드(node), 변에 해당하는 엣지(edge)로 구성되는 도형으로 표현된다. 서로 다른 연결망은 위상학의 관점에서 비교된다. 예컨대 두 연결망의 노드 수가 같고, 노드별 엣지 수도 같으면 두 그래프는 동형사상(isomorphism)이다(Wilson, 1979). 다른 예로 서로 다른 연결망의 분포는 연결정도 지수(degree exponent) 값에 따라 완전 연결망(complete network)이나 정규 격자(regular lattice), 척도 없는 연결망(scale free network), 두터운 꼬리 분포(fat tailed distribution) 등으로 파악할 수 있다(강병남, 2010; 박대민, 2023).

커뮤니케이션학에서도 위상학적 분석이 활용될 수 있다. 예컨대 여론 확산 등에 대한 위상학적 분석으로는 사회물리학(sociophysics)의 의견동역학 연구를 활용한 사례를 들 수 있다. 의견동역학은 온도에 따른 양자의 자성 변화를 설명하는 이징 모형(ising model)을 의견 확산에 적용한 것이다(Castellano et al., 2009). 이징 모형은 자성체의 성질을 설명하는 물리학 모델이다. 자성체는 작은 입자들로 이뤄져 있으며 각 입자는 스핀(spin) 값을 갖는다. 스핀 값은 +1(up) 또는 -1(down)이다. 입자들은 통상 주변 입자와 같은 스핀 값을 가지려는 경향이 있다. 그러나 온도가 높아지면 입자 움직임이 활발해지면서 주변 입자와 스핀 값이 달라질 수 있고, 반대로 온도가 낮아지면 모든 입자가 같은 방향으로 정렬돼 자성을 띄게 된다.

의견동역학은 이징 모형을 의견 확산 모델에 적용한다. 이징 모형의 양자는 행위자(agent)에, +/-의 스핀 값은 찬/반의 의견 값에 대응되며 온도나 반응 조건 등은 여론 기후나 동의 조건과 같은 초기 조건이 된다. 초기 조건에 따라 의견이 동일해지는 동조화, 의견이 양분되는 양극화(polarization), 수많은 의견이 난립하는 파편화 등이 일어난다. 이러한 방식으로 의견 확산을 ABM(agent based model)로 시뮬레이션할 수 있다. 의견 동역학의 접근은 의제 설정 이론이나 침묵의 나선 이론 등에 적용된 바 있다(Neuman et al., 2014; Sohn, 2022).

### 4) 정보원의 동조화와 파편화에 대한 위상학적 분석

동조화와 파편화는 자연과 사회의 다양한 현상을 수학적으로 모델링하고자 하는 복잡계 물리학

의 주요 주제 중 하나이다. 복잡계 물리학에서 동조화란 개별 구성 요소들이 상호작용을 통해 동일한 리듬, 상태, 또는 의견으로 수렴하는 과정을 뜻한다. 서로 다른 두 진자 시계의 동조화, 수 천 마리 반딧불이의 동시 발광, 동거 여성의 월경 주기 일치 등 물리학이나 생물학적 현상에서부터 군중의 발걸음이나 박수가 동조화되거나, 인종 분리, SNS에서 의견 확산 등 사회 현상에 이르기까지 다양하게 나타난다(Schelling, 1971; Strogatz, 2003). 파편화는 여러 개의 하위 집단으로 나뉘어 서로 다른 의견, 행동, 또는 상태를 유지하는 현상이다. 파편화는 상호작용이 이뤄질수록 전체 상태가 무질서하게 발산한다. 양극화는 전체 상태가 서로 다른 2개의 극성을 띠는 현상으로 볼 수 있다. 예컨대 의견동역학에서 동조화는 행위자가 + 또는 - 중 어느 한쪽의 스펙트럼 값만 갖는 현상을, 양극화는 + 와 - 를 모두 가지면서 2개의 거대 군집으로 수렴하는 현상을, 파편화는 +와 - 를 모두 가지면서 거대 군집이 출현하지 않은 상태로 상호작용이 중단되어 전체 상태를 결정하기 어려운 현상을 뜻한다(Castellano et al., 2009).

의견동역학은 미시적 접근을 취한다. 개별 행위자들의 초기 조건(예컨대 +, -), 개별 행위자들 간의 상호작용 규칙(예컨대 다수 의견에 따름), 규칙이 작동하는 임계값(threshold) 등이 주어지면, 상호작용이 무한대 또는 상당히 진행됨에 따라, 전체 체계의 거시적인 최종 상태가 동조화나 파편화, 또는 양극화 등으로 수렴한다고 본다. 즉 행위자의 초기 상태와 행위자의 행동 규칙 등 미시적인 초기 조건을 알면, 거시적인 최종 상태를 예측할 수 있다고 가정한다. 의견동역학은 주로 수학적 모델을 시뮬레이션 기법을 통해 검증하는 방식으로 수행됐다. 의견동역학의 여러 연구에 따르면 초기 조건과 상호작용 규칙이 달라짐에 따라 동조화, 양극화, 파편화가 다양하게 일어날 수 있다(Castellano et al., 2009).

의견동역학은 위상학적 접근을 통해 커뮤니케이션 연구를 수학적으로 이론화하는데 기여할 수 있지만 몇 가지 한계를 갖고 있다. 가장 큰 문제는 지나치게 이론적이라는 점이다(Flache et al., 2017). 예컨대 의견은 찬(+), 반(-)으로만 나뉘지 않는다. 같은 찬반 내에서도 의견이 다양할 수 있다. 찬반의 강도도 다르다. 현실에 맞는 초기 조건을 정확히 이론화하는 것도 쉽지 않다. 이는 초기 조건에 따라 최종 거시 상태가 완전히 달라질 수 있음을 감안하면 중요한 제한점이다. 상호작용이 최종 상태에 이르기까지 초기 조건이 장기적으로 유지되는 경우도 이론적으로만 가능할 수 있다. 시간을 극한으로 보내는 것은 시뮬레이션 상에서만 가능한 것이다. 동조화나 파편화가 동시에 존재하는 하이브리드 구조도 존재할 수 있다. 예컨대 글로벌화와 지역적 파편화의 공존, 합의와 다양성의 공존이 가능하다(Axelrod, 1997; Flache et al., 2017). 의견동역학은 주로 행위자에 대한 분석을 주로 하며 텍스트 등 의미 수준의 분석도 제한돼 있다(박대민, 2023).

의견동역학의 한계를 극복하기 위해서는 수학적 모델을 시뮬레이션 기법으로 검증하는 것이 아니라, 실제 빅데이터를 활용해 모델을 도출하는 것이 필요할 수 있다. 또한 커뮤니케이션 연구의 관점에서 보면 행위자만이 아니라 의미 수준의 분석도 필요할 수 있다. 즉 뉴스 빅데이터 분석처럼 기사에서 자연어처리 등으로 얻은 단어들의 의미론적 관계를 분석할 필요가 있다. 예컨대 언론의 정보원 인용 패턴은 뉴스 정보원 연결망 분석을 통해 위상학적으로 표현할 수 있다. 뉴스 정보원 연결망은 정보원을 노드로, 기사 공동출현 여부를 엣지로 하는 무방향(undirected) 1원 연결망(1 mode network)을 말한다(박대민, 2013). 이를 이용해 정보원 인용으로 나타는 미디어 편향(media bias)을 위상학적으로 분석한 사례도 있다. 이에 따르면 정보원 인용 패턴을 뉴스 정보원 연결망으로 위상학적으로 표현했을 때, 연도, 매체, 지면, 의제에 상관없이 기사가 충분히 많기만 하면, 어떤 기사든 그 기사로부터 도출된 뉴스 정보원 연결망들은 그 분포가 두터운 꼬리 분포(fat-tailed distribution)로 수렴하는 동일한 위상학적 특징을 갖는다(박대민, 2014, 2023). 이는 시기나 언론사, 지면, 의제를 막론하고 소수의 주요 정보원에 극도로 편향된 동시에, 중요도는 낮지만 압도적 다수를 차지하는 정보원들이 대거 공존한다는 뜻이다.

다만 해당 연구는 개별 매체 수준의 정보원 인용에 대한 위상학적 분포를 밝혀냈지만, 전체 매체 체계 수준에서 매체 간 정보원 인용의 위상학적 특징은 제시하지 못했다. 개별 매체가 소수의 허브(hub) 정보원에 편향되어 있더라도 매체 간에 비교하면 그 허브 정보원이 서로 다를 수 있다. 즉 매체 내 편향은 강하더라도, 매체 간 편향 내용은 다르게 나타나면서 매체 전체적으로는 다양성을 확보할 수 있다. 사실 이는 매체 간 비교 연구가 가정하는 전제이기도 하다. 예컨대 보수지는 보수 정보원에, 진보지는 진보 정보원에 편향될 수 있다(이종혁, 2022). 하지만 매체 간 동조화도 강하게 나타날 수도 있다. 예컨대 정치면에서는 대부분의 매체에서 대통령을, 사회면에서는 교육부를 가장 중요한 정보원으로 다룬다(박대민, 2016). 더 나아가 상위권에서는 매체 간 정보원 인용이 동조화되더라도 중하위권에서는 다르게 나타나는 식으로 혼합될 수도 있다. 예컨대 북한 관련 사실에서 매체 간 정보원 차이는 중위권 이하에서 두드러진다(김경화·노기영, 2011). 즉 동조화와 과편화가 동시에 나타나는 하이브리드 구조가 정보원 인용에서 나타날 수 있다. 이러한 하이브리드 구조도 동조화가 강한 경우와 과편화가 강한 경우로 나뉘 볼 수 있다. 본 연구에서는 전자를 과편화된 동조화, 후자를 과편화된 동조화로 명명한다.

본 연구는 우선 정보원 인용 패턴을 실제 데이터에 따라 뉴스 정보원 연결망으로 위상학적으로 표현한다. 뉴스 빅데이터는 인용 및 기사 작성과 같은 의미론적 상호작용이 충분히 진행된 상태를 보여주는 실증 데이터로 간주할 수 있다. 뉴스 빅데이터에서 정보원 인용의 위상학적 특

정이 동조화, 파편화, 또는 하이브리드 구조(동조화된 파편화, 또는 파편화된 동조화) 중 어떤 상태로 수렴하는지를 파악할 수 있다. 또한 이러한 수렴 상태가 매체별, 시기별, 지면별로 차별적으로 나타나는지, 대동소이하게 나타나는지도 살펴보고자 한다.

### 3. 이론적 논의

본 연구에서는 여러 매체를 비교하기 위한 방안으로 순위를 고려한 파편화 지수를 제안한다. 파편화 지수는 간단한 수식을 통해 공통된 정보원이나 주제의 수에 따라 매체 간의 차이, 즉 동조화 정도와 파편화 정도를 총체적으로 보여준다.

파편화 지수의 아이디어는 다음과 같다. 예컨대 A, B, C 등 3개의 매체가 각각 a, b, c 등 3명의 정보원을 인용했다고 하자. 3개 매체가 활용한 정보원 수는 중복을 포함하면 9명이다. 만일 3개 매체가 완전히 같은 정보원을 썼다면 각각 활용한 3명의 정보원이 모두 같으므로 중복을 제거한 정보원 수는 3명이 된다. 이 경우 차이가 가장 적게 된다. 반면 3개 매체가 완전히 다른 정보원을 활용했다면 3개의 매체는 중복을 제거해도 9명의 정보원을 인용했을 것이다. 매체 간 차이 정도는 중복을 포함한 전체 정보원 수에서 중복을 제거한 정보원 수로 나눔으로써 파악할 수 있다. 앞의 예에서 정보원이 완전히 같은 경우 그 값은  $3/9$ 이 된다. 반대로 완전히 다르다면  $9/9$ 가 된다.

다만 3개 매체가 완전히 동일한 정보원을 인용했다고 해도 서로 다른 중요도로 인용했을 수도 있다. 예컨대 매체 A는 정보원 a를 가장 중요하게 인용했고, 정보원 c를 가장 덜 중요하게 다뤘지만, 매체 B는 정보원 c를 가장 중요하게 인용했고 정보원 a를 가장 중요하지 않게 취급했다면, 매체 A와 매체 B는 인용 방식이 다르다 할 수 있다. 따라서 파편화 지수는 순위를 고려해 계산해야 한다.

그러나 예컨대 순위가 1위, 10위 등으로 차이가 나더라도, 10위권 순위 내에서 동일한 정보원을 결국 인용한다면, 순위차는 나지만 결국 동조화됐다고 할 수 있다. 즉 상위 순위에서는 파편화됐지만, 하위 순위에서 동조화됐다고 할 수 있다. 반대로 순위 내에서는 동조화됐지만, 중복되지 않는 정보원 수가 늘어나면서 파편화가 심화될 수도 있다. 사실 언론계가 인용하는 정보원이 연도별, 지면별로 많게는 1만명도 넘는다는 점을 감안할 때, 약간의 순위차로 정보원 인용이 다르다고 볼 수 없다. 그보다는 어떤 정보원을 인용했는지 안 했는지 여부 자체가 언론사 간의 차이를 더 명확히 드러낸다.

본 연구에서 정보원의 중요성과 순위는 뉴스 정보원 연결망 분석을 통해 나온 정보원의 연결정도 중앙성으로 구한다. 연결망에서 노드의 순위는 연결정도 중앙성 외에도 위세중앙성(eigenvector centrality), 사이중앙성(betweenness centrality), 인접중앙성(closeness centrality) 등으로 구할 수 있다. 이중 커뮤니케이션학은 물론 사회학 등 여러 분야에서 가장 널리 쓰이는 것이 연결정도 중앙성이다(이완수 최명일, 2020). 연결정도 중앙성은 두터운 꼬리 분포 등 연결망의 기본적인 분포를 파악하는 연결정도 지수 값을 계산할 때 사용한다(강병남, 2010). 또한 연결정도 중앙성은 빅데이터 분석에 비교적 적합하다. 다른 중앙성 값들이 연결망 내 모든 노드값을 고려하는 전체 중앙성(global centrality)인 반면 연결정도 중앙성은 이웃한 연관이 수만 간단히 계산하면 되는 지역 중앙성(local centrality)이기 때문이다. 본 연구에서 분석할 뉴스 정보원 연결망은 많게는 수만 개의 노드가 등장하는 빅데이터에 의한 연결망으로 연결정도 중앙성과 같은 지역 중앙성이 계산에 용이하다.

뉴스 정보원 연결망 분석으로 구한 연결정도 중앙성은 그 해석도 비교적 직관적이다. 우선 한 정보원의 연결정도 중앙성 값은 그 정보원과 같은 기사들에 등장한 정보원 수를 뜻한다. 어떤 정보원의 연관 정보원 수가 많다는 것은 그 정보원이 다양한 기사에서, 다양한 주제에 대해, 다양한 정보원과 함께 인용됐다는 뜻으로, 일종의 논객임을 보여준다. 이러한 논객이 중요하다고 간주하는 것은 저널리즘 관행 측면에서도 타당하다. 따라서 각 정보원의 순위는 연결정도 중앙성 값이 클수록, 즉 중복 인용을 제외하고 공동 인용된 연관 정보원 수가 많은 정보원이 더 높은 것으로 간주한다.

본 연구는 매체 간 정보원의 중첩 정도를 순위를 고려해 파악하기 위한 지표로 파편화 지수를 제안한다. 파편화 지수는 특정 순위에서 누적된 정보원이 얼마나 다른지를 계산한다. 특정 순위  $r$ 에서 중복 제거된 누적 정보원 수를  $s_d$ , 중복 포함된 누적 정보원 수를  $s_i$ 이라고 할 때, 파편화 지수  $f_r$ 은 아래와 같이 간단히 구할 수 있다.

$$f_r = \frac{s_d}{s_i}$$

$s_i$ 는 인용할 수 있는 최대 슬롯,  $s_d$ 는 인용되어 채워진 슬롯이며  $f_r$ 는 슬롯이 채워진 비율로 간주할 수도 있다. 완전히 파편화된 경우  $f_r$ 은 1이 된다. 모든 정보원을 다르게 인용했을 경우 중복된 정보원 수가 0이므로 중복 제거된 정보원 수  $s_d$ 과 중복 포함 정보원 수  $s_i$ 는 같기 때문이다.

반대로 완전 동조화된 경우, 즉 완전 중복되어 있다면 다음과 같다. 매체 수가  $m$ 일 때, 동 순위가 없다면 모든 매체가 매 순위에서 동일 정보원 1명만 인용하는 것이므로 중복 제거된 정보

원 수는 순위  $r$ 과 같은  $r$ 이 된다. 중복 포함 정보원 수는 ' $r \times m$ '이다. 따라서 완전히 동조화된 경우  $f_r$ 은 ' $1/m$ '이 된다.

한편 동순위 정보원의 수가  $a$ 라면 순위  $r$ 까지 중복 제거된 정보원 수는 ' $r+a$ ', 중복 포함된 정보원 수는 ' $r \times m + a$ '이다. 따라서  $f_r$ 은 ' $(r+a)/(r \times m + a)$ '이 된다. 계산해보면 이 값은 ' $1/m$ '보다 크다. 따라서  $f_r$ 의 최대값은 완전 파편화된 상태인 1, 최소값은 완전 동조화된 상태인  $1/m$ 이 된다.

언론사들이 특정 순위까지 모든 정보원을 공동의 정보원 풀 안에서 선택한다면,  $f_r$ 은 최소값인  $1/m$ 에 가까워질 것이다. 모든 순위, 즉 최하위 순위까지 고려할 때, 정보원 풀을 완전히 공유한다면, 완전 동조화돼  $f_r$ 은 최소값인  $1/m$ 로 수렴할 것이다. 반대로 언론사들이 다른 언론사가 인용하지 않은 정보원을 최대한 추가로 인용한다면,  $f_r$ 은 완전 파편화를 의미하는 최대값인 1의 값에 가까워질 것이다. 최하위 순위로 갈수록 중복되지 않은 정보원이 늘어난다면, 즉 파편화가 심화된다면  $f_r$ 은 1에 점점 가까워질 것이다.

참고로 매체가 1개뿐이라면  $f_r$ 은 최대값과 최소값이 1로 일치하게 된다. 즉 완전 동조화와 완전 파편화는 동일한 값이 된다. 매체가 늘어날수록 완전 동조화됐을 때  $f_r$ 값이 작아진다. 즉 매체가 늘어났음에도 불구하고 여전히 동조화됐다면, 파편화 지수는 더욱 낮게 평가된다.

〈Figure 1〉은 매체 수가 10개라고 할 때, 동순위 정보원이 없는 경우, 동순위 정보원이 1명인 경우와 10명인 경우로 10위까지 시각화한 예시이다. 최소값은 매체 수의 역수인 0.1, 최대값은 1이며 동순위가 있는 경우 그 사이에 위치한다.

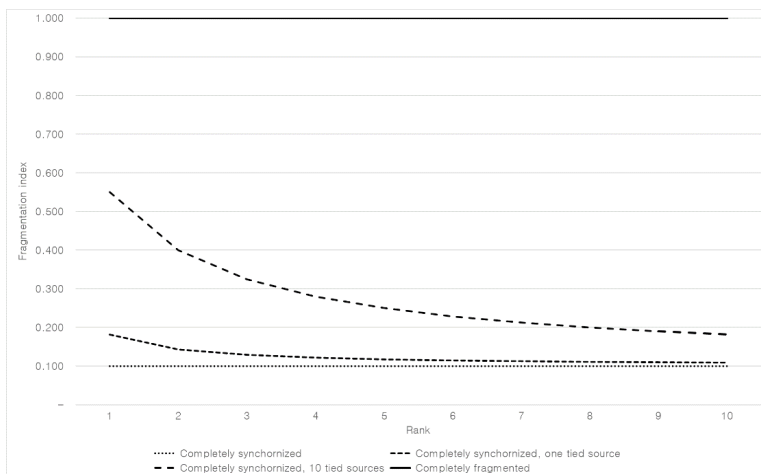


Figure 1. Examples of fragmentation index

Note. Fragmentation Index(파편화 지수, MFI), Rank(순위)

본 연구의 분석 대상 매체는 최대 19개이다. 동순위가 없을 때 이론적으로 얻을 수 있는 파편화 지수의 최소값은 1/19, 즉 반올림하면 0.053이다.

## 4. 연구문제

본 연구는 한국 언론 보도 내 정보원의 중요도를 뉴스 정보원 연결망 분석을 통해 산출하고, 이를 바탕으로 파편화 지수 또는 평균 파편화 지수(Mean Fragmentation Index, MFI)를 산출함으로써 한국 언론의 동조화와 파편화 경향은 어떠한지를 살펴본다, 이러한 경향은 거시적인 전체적 경향, 지면별 경향, 그리고 세부적인 지면 및 연도별 경향으로 나누어 살펴본다. 이에 따른 연구문제는 다음과 같다.

**연구 문제 1.** 전체 매체의 평균 파편화 지수를 통해 본 한국 언론의 동조화와 파편화 경향은 어떻게 나타나는가?

**연구 문제 2.** 지면별 평균 파편화 지수를 통해 본 한국 언론의 동조화와 파편화 경향은 어떻게 나타나는가?

**연구 문제 3.** 지면별 연도별 파편화 지수를 통해 본 한국 언론의 동조화와 파편화 경향은 어떻게 나타나는가?

## 5. 연구대상, 연구방법

### 1) 연구대상

본 연구는 정보원 수준의 동조화와 파편화 현상을 분석한다. 이를 위해 2000년부터 2020년까지 21년간 전국지 11개와 경제지 8개 등 총 19개 매체에 실린 5개 지면 기사의 직접 인용 정보원을 수집했다. 구체적으로 분석 대상 매체는 전국지 11개(경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보), 경제지 8개(매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제) 등 총 19개 매체다. 분석 지면은 정치, 경제, 사회, 문화, 국제 등 5개 지면이다. 분석 기간은 2000년 1월 1일부터 2020년 12월 31일까지이다.

분석 매체: 전국지 11개(경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보), 경제지 8개(매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 헤럴드경제)

분석 지면: 정치, 경제, 사회, 문화, 국제

분석 기간: 2000년 1월 1일부터 2020년 12월 31일까지

인용문 데이터는 한국언론진흥재단의 뉴스 빅데이터 분석 시스템인 빅카인즈를 통해 수집됐다. 데이터는 수집 데이터의 속성 및 자료형 등을 정의해 재단을 통해 빅카인즈의 개발사로부터 직접 받았다. 연도별, 매체별, 지면별 수집 데이터 세트는 1,700개(21개 연도×19개 매체×5개 지면=1,995개 세트이나 매체별 창간 연도 등의 차이로 1,700개)이다. 수집 대상 기사 수는 총 16,538,668건이지만 인용문이 없는 기사는 분석에서 제외된다. 수집 기사 수에서 데이터세트 수를 나누어 연도별, 매체별, 지면별 평균 분석 기사 수를 구하면 9,729개다. 수집 인용문 수는 66,867,349개다.

개체명 인식(named entity recognition)으로 얻은 정보원 수는 총 10,779,412명이다. 한국언론진흥재단 빅카인즈 매뉴얼에 따르면 빅카인즈의 개체명 인식은 F1 점수 기준 평균 0.915의 성능을 보인다. 여기에는 개인 실명 정보원, 기관 정보원, 익명 정보원이 모두 포함된다. 연도별, 매체별 평균 정보원 수는 정치 5,775명, 경제 9,564명, 사회 6,271명, 문화 5,025명, 국제 5,160명이다.

정보원 전처리는 최소화됐다. 'unknownactor'으로 표기된 익명 정보원은 일괄적으로 삭제했다. 추가적인 정제는 이루어지지 않았다. 예컨대 동명이인, 또는 이명동인(예: 이 회장, 이재용 회장, 이재용 삼성전자 회장)과 같은 개체명을 수정하는 단어 의미중의성 해소(word sense disambiguation)는 수행하지 않았다. 그 이유는 다음과 같다. 첫째, 자연어처리에서의 의미중의성 해소의 성능은 높지 않다. 챗GPT와 같은 최신 AI조차도 개체명의 의미중의성 해소에 실패해 환각을 일으킬 정도다(박대민·이한중, 2024). 둘째, 뉴스 빅데이터에서 수작업으로 의미중의성 해소를 하는 것은 비현실적이다. 수작업으로 의미중의성 해소를 하려면 해당 정보원이 있는 기사를 일일이 들여다보고 수정해야 한다. 셋째, 연결정도 중앙성의 특성이 의미중의성에 의한 오측이 순위나 분포에 미치는 영향을 최소화한다. 우선 중요도가 낮게 나온 오측 정보원의 경우 의미중의성을 고려해도 영향이 크지 않다. 어차피 순위가 낮을 뿐만 아니라, 순위가 낮은 수많은 정보원 중 하나이기 때문에 분포에 큰 영향을 주지 않는다. 중요도가 높은 오측 정보원의 경우 다른 정보원의 연결정도 중앙성 값을 높일 수 있지만, 그 값은 1만 올라간다. 공동 출현 횟



수는 연결강도에만 반영될 뿐이며, 연결정도 중앙성 값에는 중복 제거되어 반영되기 때문이다. 또한 뉴스 정보원 연결망에서는 노드가 추가될수록 중요한 노드가 더 중요해지는 선호적 연결(preferential attachment)이 발생한다(박대민, 2023). 이는 오측 정보원의 영향을 받는 정보원은 중요할수록 그 연결정도 중앙성 값이 더 증가하게 된다는 뜻이다. 즉 상위권 정보원 간의 순위는 변동하지 않으며, 오히려 각 순위간 연결정도 중앙성 값의 격차는 더 벌어지고 공고해진다. 예컨대 1위 정보원의 연결정도 중앙성 값이 1,000이고 2위 정보원의 값이 700, 3위 정보원 값이 500이라면, 오측 정보원이 10개 늘어날 경우 1위 정보원은 1,010으로 연결정도 중앙성 값이 가장 많이 증가하고, 2위 정보원은 706으로 그 다음, 3위 정보원은 501으로 가장 적게 증가하는 경향이 있다는 것이다. 종합하면, 기사가 충분히 많아서 노드인 정보원 수가 일정 수를 넘어가면 연결정도 중앙성으로 구한 순위는 상위권일수록 불변하며, 전체적 분포도 특정 연결정도 지수값을 갖는 두터운 꼬리 분포로 수렴하는 경향이 있다(박대민, 2023). 따라서 본 연구의 분석 결과는 전처리를 하지 않아 생기는 오측에 강건(robust)하다. 넷째, 의미중의성을 해소하기 위해 오측 데이터를 아예 삭제할 수도 있다. 그러나 모든 의미중의성 사례를 파악해 정확히 삭제하지 않고 자의적으로 편향되게 삭제하게 되면, 분포와 순위를 심각하게 왜곡할 가능성이 있다. 데이터가 충분히 많아 오측의 영향력이 축소된다는 점을 감안했을 때, 임의 정제에 의한 왜곡 위험을 감수할 필요가 없다. 다섯째, 본 연구는 일종의 복잡계(complex system)에 해당하는 복잡한 정보원 인용 패턴에 대한 근사적 결과(approximate results)를 얻고자 한다. 이는 물리학에서도 계산의 실용성을 확보할 뿐만 아니라, 오차를 인정하되 강건함을 얻는 모델로 보완하고, 단순화를 통해 보편성을 얻기 위해 널리 사용되는 접근이다.

## 2) 연구방법

본 연구에서는 뉴스 빅데이터 분석을 수행했다. 뉴스 빅데이터 분석은 자연어처리와 의미연결망 분석을 결합한 방법이다(박대민, 2023). 먼저 자연어처리 데이터는 빅카인즈에서 얻었다. 빅카인즈는 기사에서 정보원과 인용문을 추출해 매핑한 인용문 데이터를 제공한다. 정보원은 이름, 소속, 직함에 대해 각각 개체명 인식을 진행해 결합한 'INFOSRC'로, 기사 식별자는 매체명, 발행일, 일련번호를 조합한 'ART\_ID'로 제공한다. 의미연결망 분석은 뉴스 정보원 연결망 분석을 실시했다. 뉴스 정보원 연결망의 노드 정보는 'INFOSRC'에, 엣지 정보는 'ART\_ID'에 있으며 이를 통해 링크드 리스트(linked list) 형태의 연결망 데이터를 구할 수 있다. 이를 활용해 연도별, 매체별, 지면별로 모든 정보원에 대한 연결정도 중앙성 값을 구했다.

UCINET이나 NETMINER와 같은 연결망 분석 프로그램이나 Python이나 R의 연결망 분

석 라이브러리를 통해 각각 수만 개의 노드로 이뤄진 수천 개의 데이터세트를 연결망으로 분석하기에는 난관이 많다. 이에 본 연구에서는 JAVA 기반으로 개발된 SNAlyzer (<https://bit.ly/47esex6>)를 사용했다(박대민, 2023). SNAlyzer는 파일 병합 및 분할 모듈인 projector, 정제 모듈인 Cleanser, 파일명 표준화 모듈인 FileNameStandardizer, 연결정도 중앙성 계산 모듈인 DegreeAnalyzer, 기술통계 및 순위화 모듈인 StatisticsAnalyzer 등으로 구성된다. SNAlyzer는 파일당 1,048,575개 인스턴스를 한 번에 처리한다. 또한 입력 폴더의 엑셀 파일을 순차적으로 분석한 뒤 출력 폴더에 저장해 수천 개의 데이터세트를 한 번에 분석할 수 있다.

순위화와 파편화 지수의 계산은 ChatGPT-4o를 사용해 Python 코드를 짜서 수행했다. 뉴스 정보원 연결망은 두터운 꼬리 분포를 갖기 때문에 중하위권에서 연결정도 중앙성 값이 같은 동순위 정보원이 대거 발생한다. 동순위 처리는 동순위 정보원은 모두 같은 순위를 부여한 뒤, 비교를 위해 순차적으로 순위 ID를 재할당했다. 예컨대 10명의 정보원이 있고 5위에 동순위 정보원이 5명이 있다면, 이들을 모두 5위로 본 뒤 10위 정보원을 6위로 할당했다.

본 연구에서는 파편화 지수의 변화를 전체 경향, 지면별 경향, 연도별 지면별 경향 등 세 단계로 순차적으로 살펴본다. 전체 경향과 지면별 경향은 동일한 순위 ID에서 파편화 지수의 평균, 즉 MFI를 내어 전체, 지면별로 계산해 분석한다.

그런데 연도별, 지면별로 파편화 지수의 최대값과 최소값, 파편화 지수의 추세가 달라지는 각 세부 순위 ID 구간들의 시작점 등이 다를 수 있다. 특히 데이터세트별로 파편화 지수의 전체 순위 ID 구간(1위부터 최하위까지 구간)이 다를 수 있다. 이는 특정 데이터세트의 최종 순위 ID 이후에 평균할 합산 값 하나가 아예 없어진다는 것을 의미한다. 이 경우 <Figure 2>처럼 특정 데이터세트의 최종 순위 ID 이후 MFI에 급격한 계단식 변화가 발생하게 된다. 따라서 이러한 MFI 추세 왜곡을 피하기 위해 MFI 계산시에는 평균을 내는 데이터세트 중 최종 순위 ID가 가장 낮은 데이터세트를 기준으로 전체 데이터세트의 해당 순위 ID 이후의 MFI값을 계산에서 제외한다. 예컨대 10개 데이터세트 중 한 데이터세트의 순위 ID가 1위부터 100위까지 있고, 다른 데이터세트의 순위 ID는 1위부터 1,000위까지 있다면, MFI는 10개 데이터세트의 1위부터 100위까지만 구해 제시하고 100위를 초과하는 데이터는 MFI를 구하지 않는다. 다만 이러한 절삭으로 전체, 또는 지면별 MFI를 구할 때는 정보 손실이 발생한다. 하지만 이는 지면별 연도별로 나눠진 개별 데이터세트의 파편화 지수를 평균 내지 않은 상태에서 분석함으로써 보완할 수 있다. 연도별, 지면별 최종 순위 ID는 아래 표와 같다. 문화면 2000년도 데이터세트의 최종 순위 ID가 114위로 가장 작다. 이에 따라 전체 데이터세트의 MFI는 114위까지 구했다. 지면별 MFI는 각 지면별 최소값으로 절삭했다.

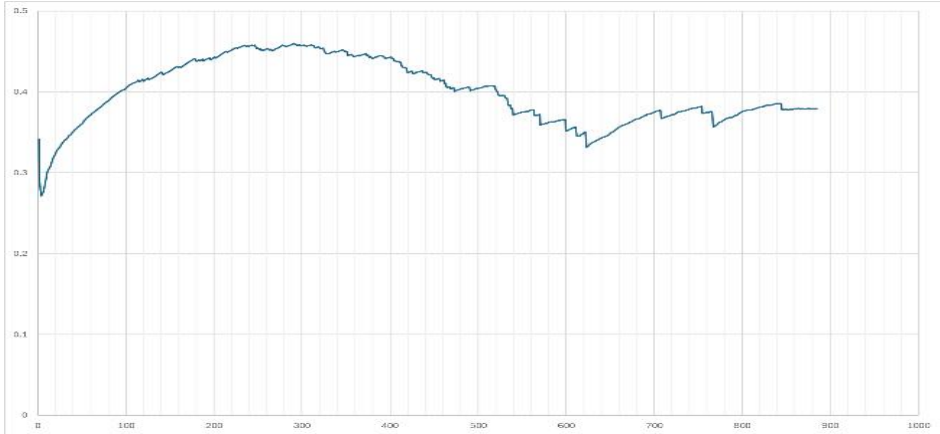


Figure 2. Skewing MFI trends for overall rank IDs across the entire datasets

Table 1. Final Rank ID Values by Year and by Section

Year	Politics	Economy	Social	Culture	International
2000	302	238	158	114	140
2001	346	253	201	119	195
2002	349	314	195	125	183
2003	407	325	251	141	225
2004	468	373	262	161	246
2005	394	378	278	179	255
2006	410	375	265	178	243
2007	446	351	293	187	247
2008	450	418	297	214	280
2009	464	404	313	235	291
2010	520	425	320	247	324
2011	533	461	351	261	347
2012	563	442	326	250	304
2013	530	414	357	327	324
2014	536	412	364	318	337
2015	611	437	393	356	391
2016	707	419	412	311	401
2017	766	447	463	319	456
2018	754	490	519	358	472
2019	844	522	599	406	570
2020	885	533	622	406	539

Note: Year(연도), Politics(정치), Economy(경제), Social(사회), Culture(문화), International(국제)

## 6. 결과

### 1) 전체적 경향

전체 데이터셋을 대상으로 한 순위 ID별 MFI 추이는 <Figure 3>과 같다. 전체 데이터셋에 서는 MFI를 114위까지 구했다.

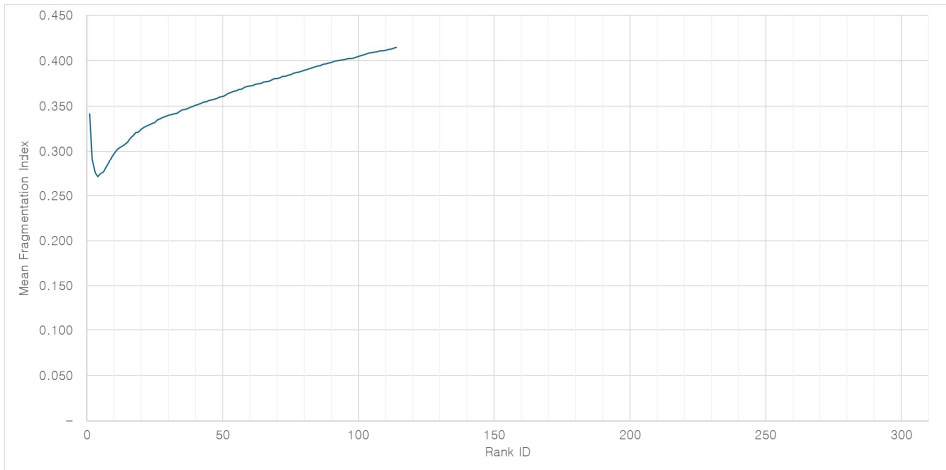


Figure 3. MFI trends of the entire datasets

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

<Figure 3>의 시각화 결과는 어느 정도 직관적인 구간 구분을 보여준다. 그러나 좀 더 엄밀하게 구간 구분하기 위해 MFI의 증분, 다시 말해 각 순위에서의 미분값을 살펴보고 기울기 값이 상대적으로 급변하는 구간, 즉 변곡점을 찾았다. 기울기 급변은 1) 기울기가 음에서 양으로 전환하거나, 2) 그 밖에 기울기 값의 변화 추이가 달라진다는 것을 의미한다.

먼저 1구간은 순위 ID가 1위인 구간이다. MFI가 0.342로 작은 편이다. 즉 상당히 동조화된 상태다. 평균 16.1개 매체에서 평균 16.5명(중복 포함)을 인용했는데, 중복 제거하면 평균 5.6명의 정보원을 활용했다. 즉 최소한 3개 매체에서 같은 정보원을 썼다는 뜻이다. 이는 정치면의 대통령처럼 가장 중요한 정보원은 매체를 막론하고 대동소이하기 때문이다.

2구간은 MFI가 최소값까지 떨어지는 구간이다. 여기서는 4위권까지다. MFI가 0.271까지 감소한다. 즉 동조화가 심화됐다. 4위권까지 평균 66.7명의 정보원(중복 포함)을 인용했는데, 중복을 제거하면 평균 정보원 수는 18.1명으로 대폭 감소한다. 이는 4위권까지는 각 언론사들이 새로운 정보원을 인용하더라도 다른 언론사에서 인용한 정보원을 똑같이 인용하기 때문이

다. 즉 1위 정보원은 물론 최상위권 정보원은 대다수의 언론사가 중요하게 생각하는 정보원이다.

3구간은 MFI 증분이 비교적 큰 구간이다. 전체 데이터세트에서는 25위권까지로 이후 구간의 증분은 대부분 0.002 이하로 떨어진다. 25위권 이전에는 MFI 값이 비교적 가파르게 증가한다. 즉 동조화가 상대적으로 빠르게 완화된다. 그러나 절대적인 파편화 정도는 여전히 낮은 수준이다. 즉 MFI 값이 1구간 이하일 정도이다.

4구간은 MFI의 증분이 다소 작은 구간이다. 즉 동조화가 완화되고 파편화가 점진적으로 진행된다. 이 데이터세트에서는 25위 이후 구간이다. 컷오프한 최종 순위에서 MFI 값은 0.415로 가장 크다. 즉 4구간은 파편화가 가장 많이 진행되는 구간이다.

이는 뉴스 정보원 연결망에서 두터운 꼬리에 해당하는 하위권 정보원이 파편화를 촉진한다는 것을 의미한다. 이러한 정보원은 각 언론사가 발굴한 고유한 개인 실명 정보원이 대다수를 차지한다. 연결정도 중앙성 값이 3 이하로 중요도 순위가 낮은만큼 소수의 기사에서 자신들끼리만 인용됐을 가능성이 크지만, 언론사간 인용 패턴을 차별화하고 언론사 고유의 목소리를 내는데 일조한다. 종합하면 상위권 구간에서는 동조화가 우세한 반면 중하위권 구간에서 완만하게 파편화가 이뤄진다.

## 2) 지면별 경향

지면별 추이는 전체적인 경향과 일맥상통한다. 즉 1구간에서는 동조화된 상태로 시작한다(MFI 낮음). 2구간에서는 동조화가 심화된다(MFI 하락). 3구간에서 동조화가 비교적 빠르게 완화된다(MFI 상승). 4구간에서 파편화가 점진적으로 진행된다. 다만 1구간의 MFI 절대값이나 4구간에서 최종적인 파편화 수준 등이 차이가 난다.

이 절에서는 지면별 구간의 특성을 통해 파편화 정도를 살펴보고 그 의미를 해석해본다.

### (1) 정치면

정치면의 MFI 추세는 <Figure 4>와 같다. 정치면의 MFI는 1구간에서 0.323으로 약간 높다. 즉 상대적으로 동조화가 덜된 상태이다. 그러나 2구간의 MFI는 전 구간 최소값인 0.187이다. 구간 범위는 6위까지이다. 3구간은 26위까지로 MFI는 0.222이다. 이후 4구간은 302위까지 컷오프를 했다. MFI는 0.338로 낮은 편이다. 즉 파편화가 덜 돼 있다.

앞에서 살펴본 전체 지면의 구간별 MFI와 비교할 때, 가장 낮은 순위까지 포괄함에도 불구하고 전반적으로 MFI 값이 상당히 낮다. 즉 정치면의 동조화 경향은 구간을 막론하고 5개 지면 중 가장 강하다고 말할 수 있다.

이는 정치면이 공식적 정보원에 대한 편향이 강하기 때문으로 보인다. 정치 분야는 출입처 제도가 엄격하게 운영되는 경우가 많다. 즉 대통령, 청와대 또는 대통령실, 국회, 각 부처와 공공 기관 등은 통상 출입기지만 취재 가능하다. 대통령 해외 취재 시 소위 풀(pool)을 운영하는 경우도 적지 않다. 기자들에게는 동일한 보도자료가 배포되며, 주요 정보원도 대동소이하다. 그러나 정치면이라고 해서 공식 정보원 편향과 동조화를 방치할 이유는 없다. 출입처를 뛰어 넘어 국회나 정부의 발언을 정치 이외의 각계 관련자나 시민의 목소리를 통해 검증하고 의견을 반영할 필요가 있다.

## (2) 경제면

경제면의 MFI 추세는 <Figure 5>와 같다. 경제면의 MFI는 1구간에서 0.248로 시작한다. 2구간은 2위까지로 2위권의 MFI는 0.221이다. 3구간은 24위까지로 MFI는 0.338이다. 이후 4구간은 238위까지 컷오프를 했으며, MFI는 0.479이다.

경제면은 1구간에서는 정치면보다 동조화가 심했고 2구간에서도 동조화 경향이 강했다. 상위권 정보원은 경제 부처 및 대기업과 은행권 소속 정보원이다. 이들은 어느 정도 출입처 제도도 강하며 경제에서 차지하는 비중도 높은 곳이다.

그러나 3구간에서 과편화 지수가 빠르게 상승했으며 최종적으로 정치면보다는 동조화 경향이 크게 완화됐다. 즉 중요도 상위권 주요 정보원은 공유하지만, 중하위권 정보원은 매체별로 다르게 인용한다. 기자들이 정보원을 발굴하는 측면 외에도, 중소기업 등 저명하지 않은 정보원들도 정치면에 비해 상대적으로 언론사에 접근하기 쉽다는 점을 반영하는 것으로 보인다.

## (3) 사회면

사회면의 MFI 추세는 <Figure 6>과 같다. 사회면의 MFI는 1구간에서 0.365로 높은 편이다. 즉 동조화가 덜 되어 있다. 그러나 2구간은 7위권까지로 동조화가 급격히 진행된다. 해당 순위의 MFI는 0.175로 전체 지면 중 가장 작다. 즉 상위권 정보원 간의 동조화가 5개 지면 중 가장 심하다.

반면 3구간과 4구간의 중하위권 정보원 인용에서는 동조화 정도가 상당히 완화된다. 3구간은 28위까지로 MFI는 0.285이다. 이후 4구간은 158위까지 컷오프를 했으며, 해당 순위의 MFI는 0.452이다.

사회면은 경제면과 유사하다. 상위권은 정치면보다 동조화가 심하다. 상위권에 속한 정보원은 교육부, 경찰, 서울시, 검찰 등으로 정치면처럼 엄격한 출입처 제도가 존재하는 곳들이다.

반면, 하위권에서 과편화가 상당 수준 진행되는 양면적인 모습을 보여준다. 사회면 기지는 출입처에만 의존하지 않으며 학계, 종교계, 시민 등 다각도로 고유한 정보원을 발굴하기 때문인 것으로 보인다.

#### (4) 문화면

문화면의 MFI 추세는 <Figure 7>과 같다. 문화면의 MFI는 1구간에서 0.424로 5개 지면 중 가장 높다. 2구간은 2위까지로 2위권의 MFI는 0.355로 최소값 역시 5개 지면 중 가장 크다. 3구간은 25위까지로 MFI는 0.503이다. 이후 4구간은 114위까지 컷오프를 했으며, 해당 순위의 MFI는 0.613으로 가장 높다.

과편화 정도로 볼 때 문화면은 정보원 다양성이 가장 높다. 이는 문화면이 출입처의 의존도가 상대적으로 낮고, 국내외 정보원을 모두 아우르며, 고급문화계와 대중문화계, 종교계 등 다양한 분야와 음악, 미술, 연극, 무용 등 다양한 장르를 다루기 때문으로 보인다. 또한 다른 분야에 비해 문화부 등 공식적 정보원에 대한 독자 관심도가 낮으며 집중 인용되는 소수 주요 정보원이 있기보다는 한두 번 인용할 수밖에 없는 스타나 비주류 정보원을 주로 다루기 때문으로 보인다.

#### (5) 국제면

국제면의 MFI 추세는 <Figure 8>와 같다. 국제면의 MFI는 1구간에서 0.349로 전체 평균과 비슷하다. 즉 평균적인 수준에서 동조화돼 있다. 2구간은 14위권까지이다. 최소값인 MFI는 0.319로 높은 편이다. 즉 과편화가 어느 정도 진행된다. 3구간은 27위까지로 MFI는 0.323이다. 4구간은 140위까지로 해당 순위의 MFI는 0.437로 높은 편이다. 즉 과편화가 상당 수준 진행된다 있다.

국제면은 평균적인 수준의 과편화된 동조화를 보여준다. 상위권에서는 정치면과 경제면 사이의 동조화가 돼 있으며 미국, 중국, 일본의 정상급 정치인들이나 외교 관련 인사들이 주요 정보원이 된다. 중하위권에서는 다양한 정보원이 제시되는 편이다. 그러나 이는 발로 뛰는 취재 결과라기 보다는 취재 범위가 전 세계, 전 영역으로 넓기 때문인 것으로 보인다. 즉 특파원이나 세계 각국 매체가 소개한 내용을 포괄적으로 다루면서 나타난 결과이다.

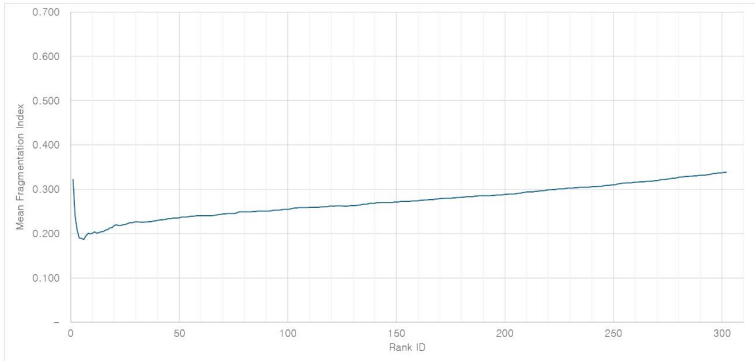


Figure 4. MFI trends by rank ID in political section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

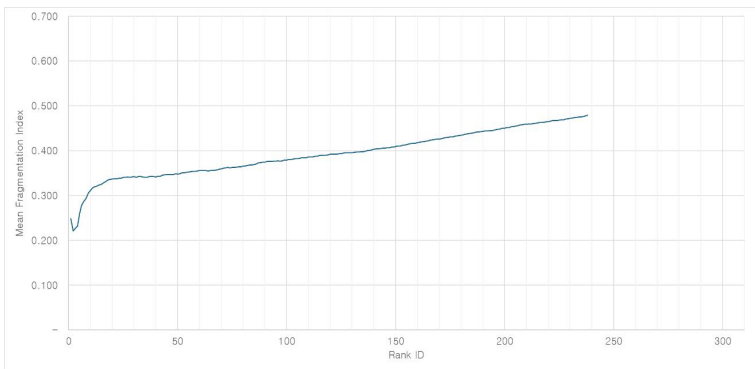


Figure 5. MFI trends by rank ID in economy section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

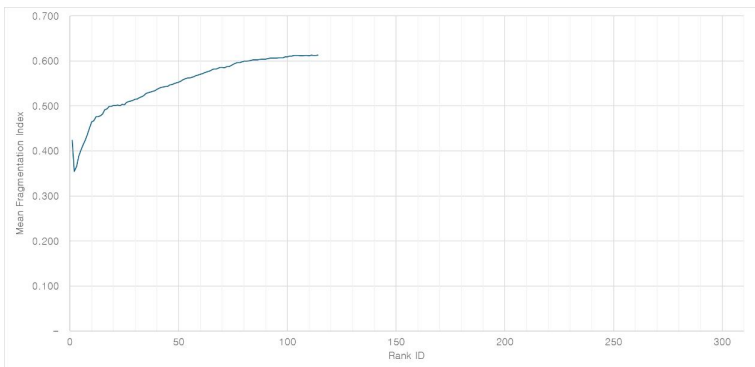


Figure 6. MFI trends by rank ID in social section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)



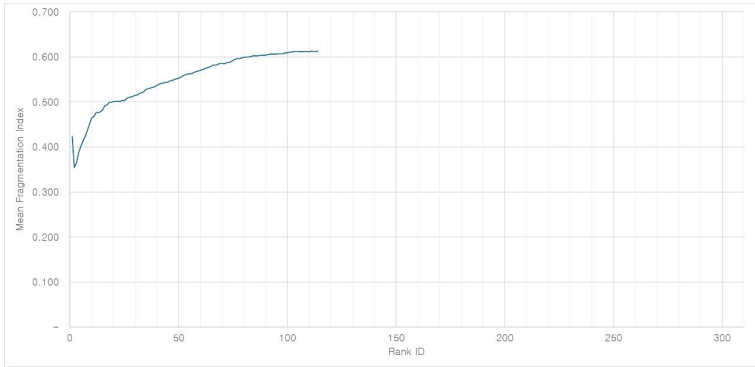


Figure 7. MFI trends by rank ID in culture section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

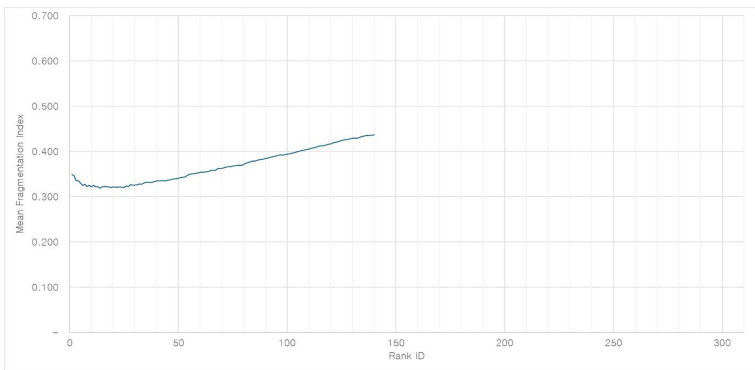


Figure 8. MFI trends by rank ID in international section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

### 3) 지면별, 연도별 경향

지면별 연도별로 데이터세트를 나누어 파편화 지수의 순위 ID별 추이는 <Figure 9>부터 <Figure 13>과 같다. 각 순위 구간의 범위나 구간별 파편화 지수의 최대최소 값은 다르다. 그러나 구분되는 각 구간이 존재한다는 점, 같은 구간에서의 움직임은 앞서 언급한 지면별 경향과 비슷하다. 즉 지면별, 연도별 차이에도 불구하고 전체적인 추이는 대동소이한 보편성을 보여준다. 앞선 지면별 분석과 비교해 추가적으로 발견되는 특이사항은 크게 두 가지이다.

첫째, 연도별 지면별 데이터세트에서는 파편화 지수의 최대값이 지면별 데이터세트보다 전반적으로 높게 나타난다는 것이다. 이는 데이터세트를 전체, 또는 지면별로 병합하고 절삭과 평균화를 수행하면 개별 데이터세트의 진짜 최대값이 누락되거나 상쇄되기 때문이다.

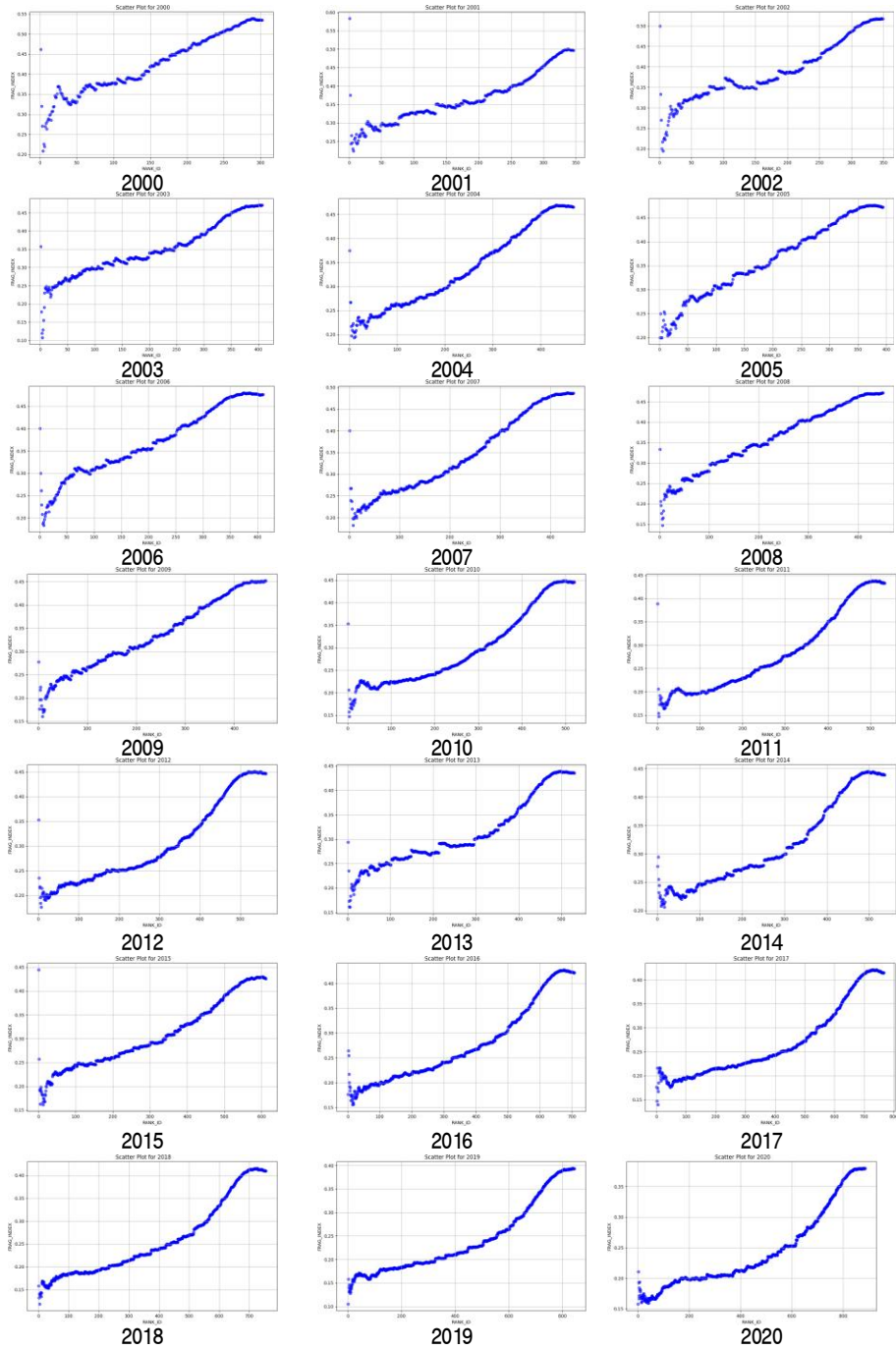


Figure 9. Fragmentation index by year in political section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

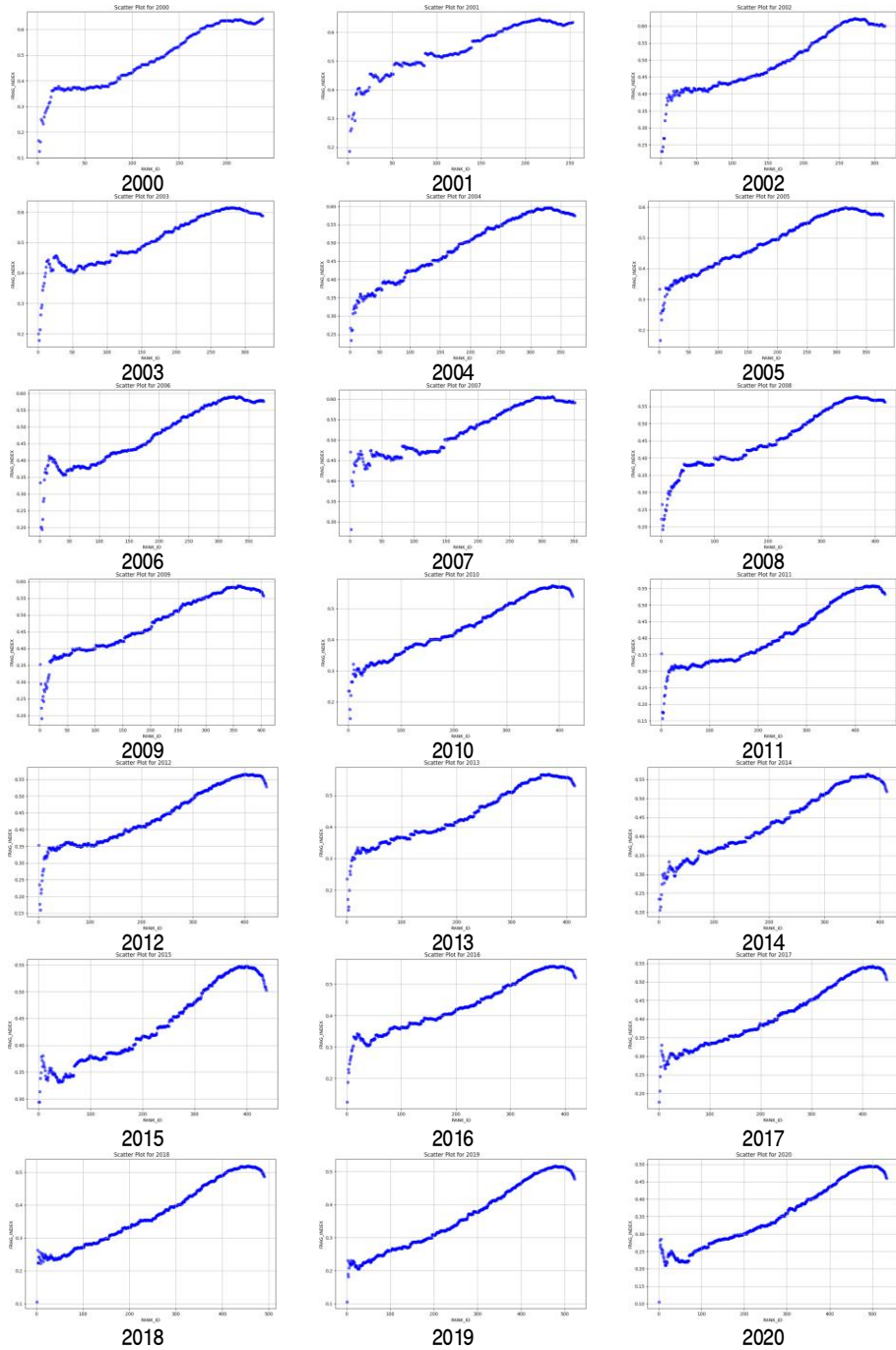


Figure 10. Fragmentation index by year in economic section

Note. Mean Fragmentation Index(평균 파편화 지수, MF), Rank ID(순위 ID)

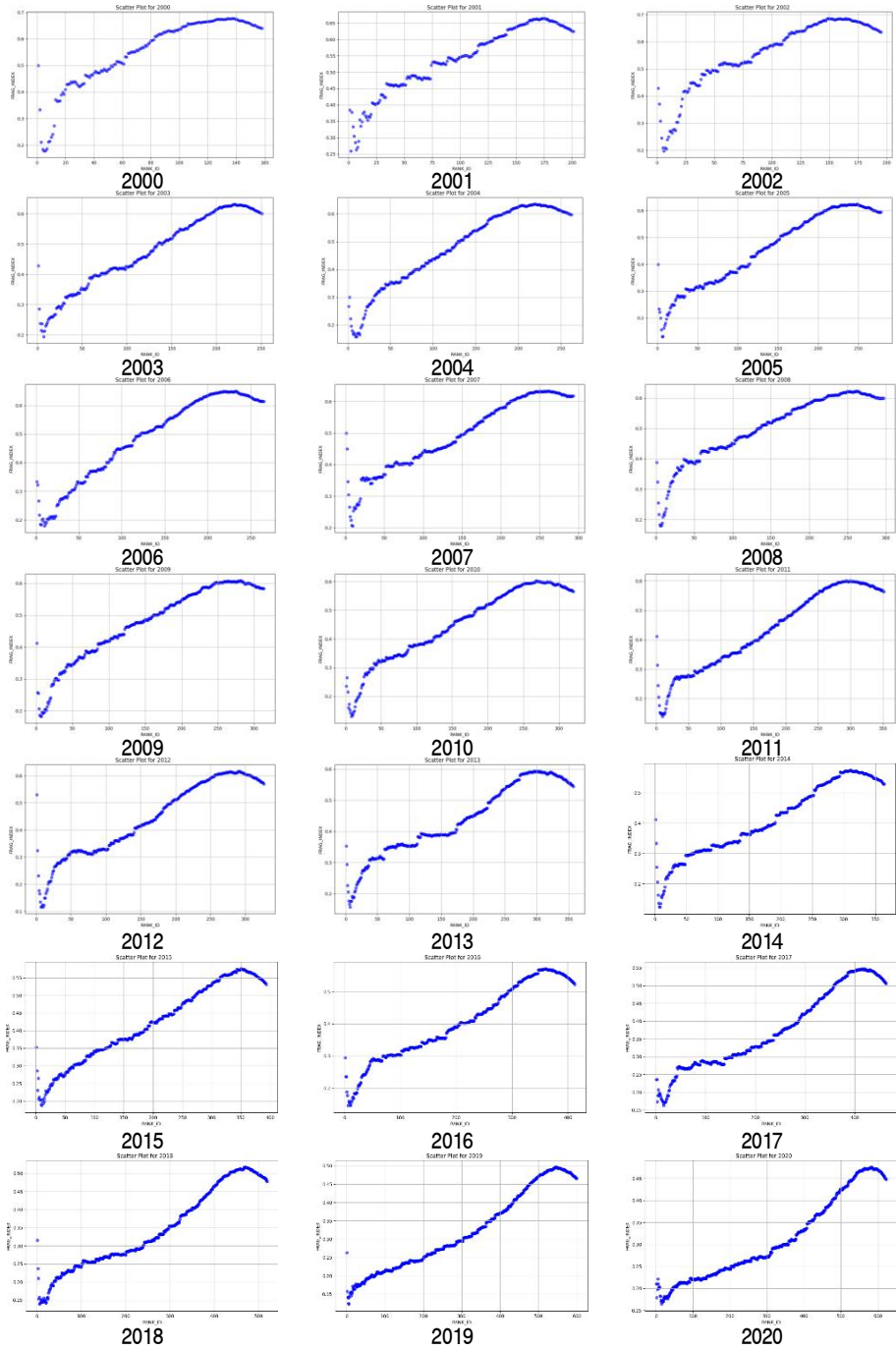


Figure 11. Fragmentation index by year in social section

Note. Mean Fragmentation Index(평균 파편화 지수, MF), Rank ID(순위 ID)

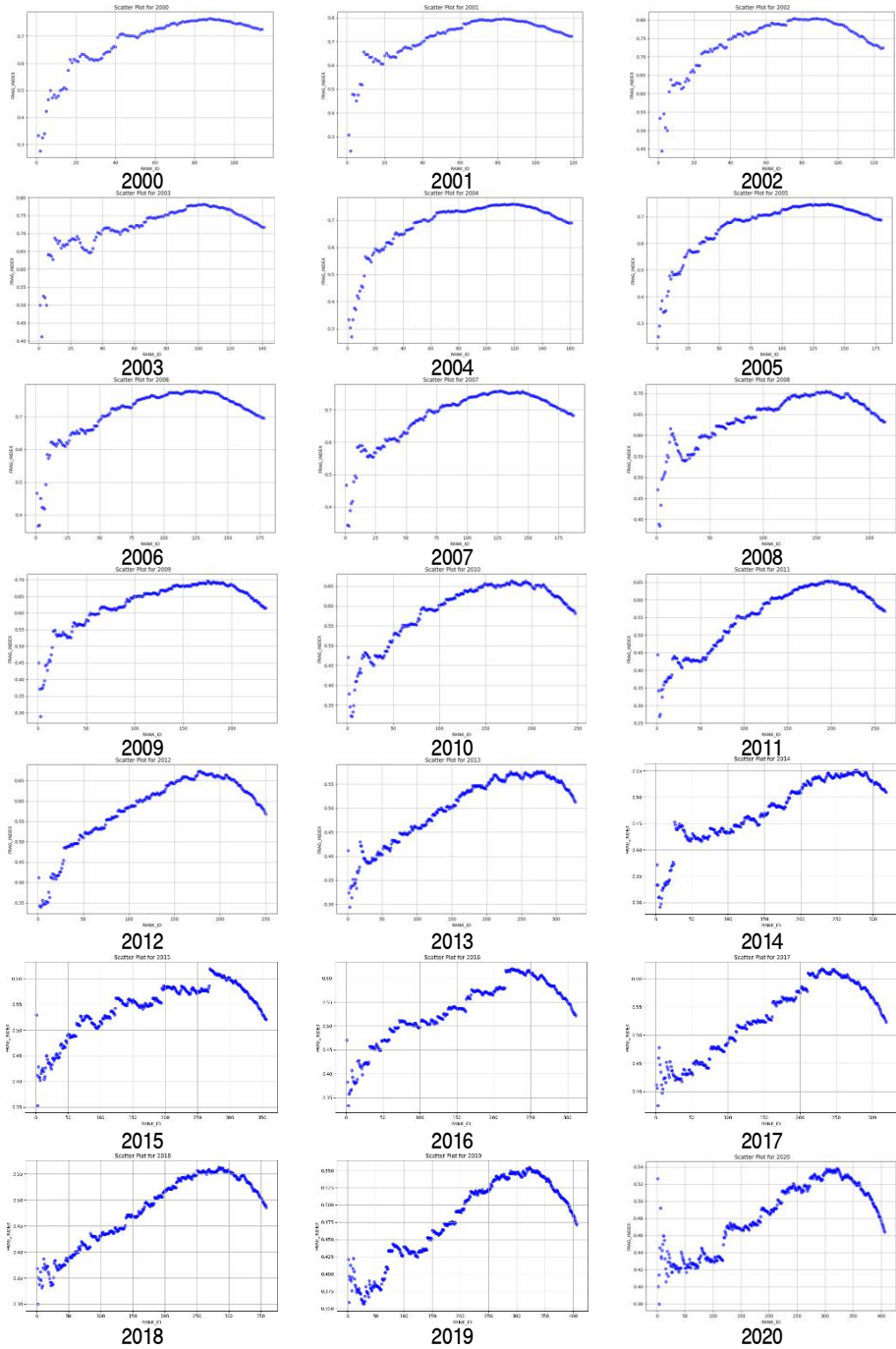


Figure 12. Fragmentation index by year in culture section

Note. Mean Fragmentation Index(평균 파편화 지수, MF), Rank ID(순위 ID)

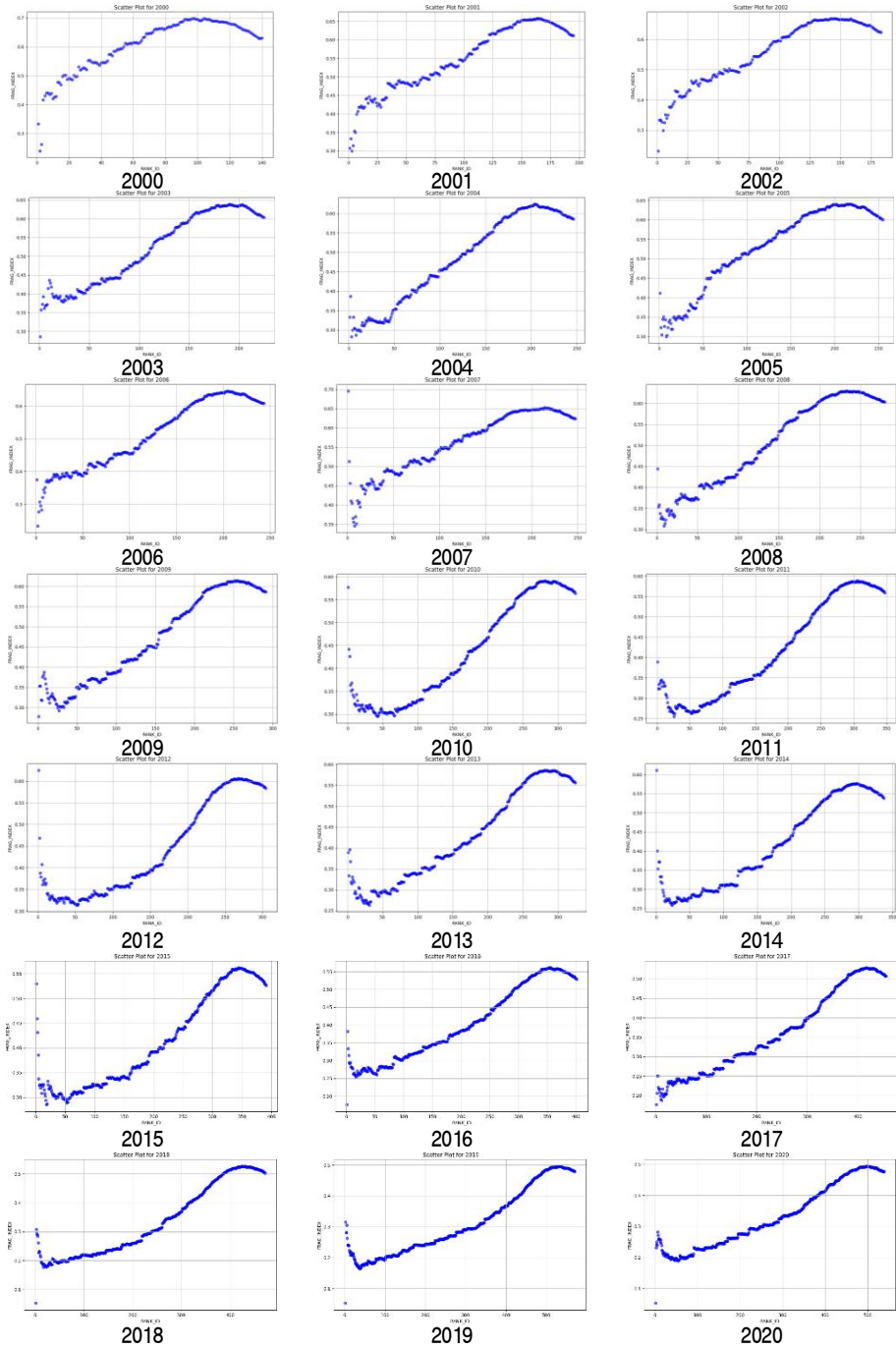


Figure 13. Fragmentation index by year in international section

Note. Mean Fragmentation Index(평균 파편화 지수, MFI), Rank ID(순위 ID)

개별 데이터세트는 동조화된 파편화 중 특히 하위권의 파편화 현상을 더욱 극적으로 보여 준다. 예컨대 2002년 문화면 데이터세트의 경우 파편화 지수 최대값은 0.804이나 된다. 완전 파편화의 경우가 1이라고 할 때 굉장히 큰 값이라고 볼 수 있다. 참고로 개별 데이터세트의 파편화 지수 최대값을 지면별로 살펴보면, 정치면은 2001년 0.583, 경제면은 2000년 0.643, 사회면은 2002년 0.686, 국제면은 2000년 0.698을 기록했다. 이러한 결과는 동조화와 파편화 중 전체적으로 파편화 경향이 우세하며, 이러한 우세가 하위권에서 두드러진다는 점을 명확히 해준다.

둘째, 최하위권까지 검토했을 때 파편화 지수의 상승세가 둔화 내지 하락하는 경향이 나타난다는 것이다. 즉 4구간 이후 5구간이 발견된다. 앞서 이러한 경향을 발견하지 못한 이유는 순위를 최저값으로 절삭하면서 개별 데이터의 데이터가 누락된 점, 누락되지 않은 데이터의 경우 평균화 과정에서 최하위권에서의 하락 추이가 상쇄된 점을 들 수 있다. 5구간에서 파편화 정도가 완화되고 동조화가 다시 심화되는 이유는 크게 두 가지로 보인다. 첫째, 굉장히 많은 정보원이 두터운 꼬리에 속해 있기 때문에 이러한 하위권 정보원들 간에도 최소한 소수의 매체들 간에는 중복 인용이 일어난다는 것을 시사한다. 둘째, 자연어처리상의 체계적 오류다. 동일한 자연어처리상의 오류가 반복적으로 나타나면서 동조화 경향이 좀 더 부각될 수 있다. 이러한 경향은 전체적인 분포에 결정적인 영향을 주지는 않지만, 두터운 꼬리가 실제보다 좀 더 강조되는 결과를 낼 수 있는 있다.

## 7. 요약 및 제언

동조화된 파편화의 일반 모델을 나타내는 그래프는 <Figure 14>와 같다. 동조화된 파편화의 일반 모델은 5개 구간으로 나누어진다. 1구간은 최상위권 구간으로 1위 정보원 간의 매체 파편화 정도를 보여준다. 상당 수준의 동조화 상태에서 시작된다. 동조화의 기준점으로 간주할 수 있다. 2구간은 파편화 지수의 최저값까지 도달하는 구간이다. 이 구간에서는 동조화가 빠르게 진행된다. 3구간은 그래프의 기울기가 가파르게 상승하는 구간이다. 이는 동조화가 비교적 빠르게 완화된 것을 의미한다. 4구간은 기울기가 완만하게 상승하는 구간이다. 파편화 지수 값은 최대값에 도달한다. 이 구간에서는 파편화가 느리지만 두터운 꼬리에 해당하는 하위 순위까지 지속적으로 진행된다. 5구간은 그래프의 기울기가 하락하는 구간이다. 이는 파편화의 완화, 동조화의 진행을 의미한다.

지면별 특성을 살펴보면, 우선 정치면은 매체 간 정보원 활용에서 동조화가 심하다. 경제면

은 상위권인 1구간과 2구간에서 강한 동조화를 보이지만 중하위권에서 파편화가 빠르게 진행된다. 사회면은 2구간에서 가장 심하게 동조화되지만, 중하위권에서 파편화가 크게 진행된다. 문화면은 전구간에서 MFI 값이 가장 크다. 즉 정보원의 다양성이 가장 높다. 국제면은 전 구간에서 중간 정도의 파편화 정도를 보여준다. 한편 지면별, 연도별로 개별 데이터세트의 파편화 지수 추이를 보면 두터운 꼬리에서 파편화 정도가 완화되는 5구간이 관찰된다.

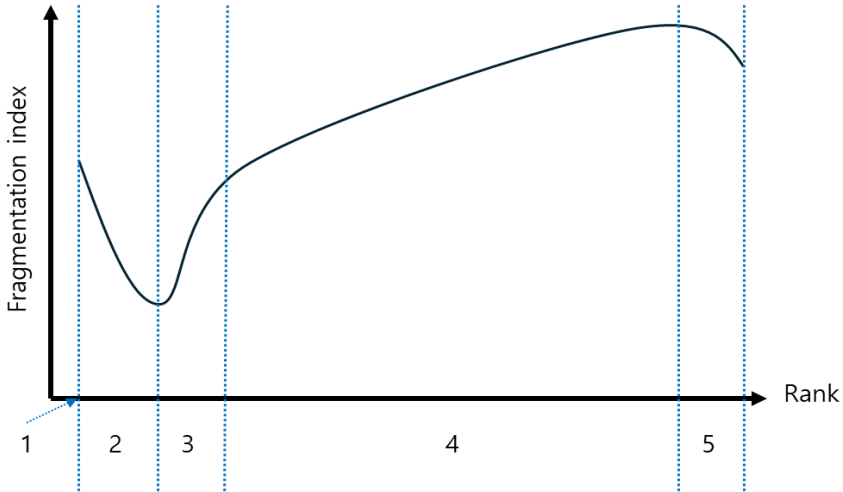


Figure 14. General model of synchronized fragmentation

Note. Fragmentation Index(파편화 지수, MFI), Rank(순위)

연구 결과는 정보원 인용에서 매체 간 동조화와 파편화가 동시에 일어난다는 것을 보여준다. 즉 상위권 정보원들 간에는 동조화가 일어나는 동시에, 중하위권 정보원들 간에는 파편화가 진행된다. 그렇다면 동조화된 파편화인가 파편화된 동조화인가? 상위권인 1, 2구간과 최하위권인 5구간에서 파편화가 완화되기는 하지만 전 구간, 전체 순위의 정보원을 볼 때 동조화보다는 파편화가 우세하다고 볼 수 있다. 즉 전체적인 관점에서 동조화된 파편화가 파편화된 동조화보다 현상에 더 적합한 용어다. 이는 뉴스 정보원 연결망이 두터운 꼬리 분포를 갖고 있기 때문이기도 하다(박대민, 2023). 이러한 분포는 중하위권 정보원 수가 상위권 정보원 수보다 훨씬 많다는 것을 의미한다. 파편화는 두터운 꼬리에 속하는 하위권 정보원과, 허브인 최상위권 정보원과 두터운 꼬리를 잇는 중위권 정보원에 의해 진행된다.

본 연구는 정보원 인용을 뉴스 정보원 연결망을 통해 위상학적으로 나타내고, 뉴스 정보원



연결망 분석으로 얻은 정보원의 중요도 순위를 활용해 총체적으로 비교했다. 특정 매체, 특정 주제, 특정 시기에 한정되지 않은 19개 매체, 5개 지면, 20년치 1,650만 건에 달하는 기사의 인용문을 전체, 지면별, 연도별로 분석해 하이브리드 구조, 정확히 말해 과편화 경향이 더 강한 동조화된 과편화 현상의 보편적 추이를 밝혀냈는데 이론적 의의가 있다. 특히 이론 추동적인 기존 의견동역학과 달리 실증 데이터에 입각해 동조화된 과편화의 일반 모델을 도출했다는 점에서 의의가 있다. 또한 행위자가 아닌 자연어처리로 추출된 기사 텍스트 내 정보원의 인용이라는 의미 구성 차원의 위상학적 분석을 수행했다는 점에서도 차별화된다. 더 나아가 본 연구는 ABM의 미시적 접근을 극복할 필요성을 제시한다. ABM은 행위자의 미시적인 초기 조건과 상호작용 규칙만 알면 거시적인 최종 상태를 파악할 수 있다고 간주한다. 그러나 본 연구가 시사하는 바는 개별 기사나 언론사가 보수적이든 진보적이든, 합리적이든 비합리적이든 상관없이, 거시 상태는 동조화된 과편화라는 강력한 보편성을 갖고 나타난다는 것을 시사한다. 쉽게 말해 충분한 기사량을 보도하기만 한다면 정치면에서 가장 중요한 정보원은 보수지나 진보지나 대통령일 수밖에 없고, 보수지가 보수 정당 대표를 조금 더 중시하더라도 진보 정당 대표는 그에 버금가게 중시할 수밖에 없으며, 진보지 역시 마찬가지로 지는 것이다.

이는 거시적인 언론학 연구의 필요성을 시사한다. 거시 언론학은 크게 두 가지 측면에 유의해야 한다. 첫째, 언론계 분석은 개별 언론이나 특정 주제 및 시기의 소수 기사를 분석해서 파악할 수 없으며, 충분히 많은 언론의 기사를 대규모로 분석해야 알 수 있다. 특정 기업이나 상품이 잘 팔린다고 해서 국가 경제성장률이 높다고 말할 수 없는 것처럼, 특정 언론사나 기사가 어떤 경향을 갖는다고 해서 전체 언론계가 그러하다는 결론을 내릴 수 없다. 둘째, 실제 언론계의 거시적 현상을 분석하기 위해서는 ABM의 미시적이고 이론 추동적인 접근이 아닌 거시적이고 데이터 추동적인 접근을 통해 연구할 수 있다. 개별 언론인이나 기사, 언론사가 불편부당하고 다양한 것과 관계없이 언론계의 거시적 현상은 특정 상태로 수렴할 수 있기 때문이다.

동조화된 과편화가 보편적이라면 그 실천적 의미는 적지 않다. 때로는 정파적인 견해에 따라 한 진영이 언론계를 반대 진영에 동조화돼 있다고 비난하면서, 자기 진영과 조응하는 소수 언론을 편향되게 지원하려는 사례가 있다. 반대로 언론계가 과편화되어 사회 갈등을 조장한다면서 언론 자유를 제한하고 자신의 입맛에 맞는 언론만 남겨 놓으려고 시도하는 경우도 있다. 그러나 동조화된 과편화 현상은 정파적 언론관에 따른 인위적 개입이 아니라 자유 언론이 직접 발로 뛰면서 발굴한 정보원들을 통해 자신들의 차이를 드러내고 다양성을 확보할 수 있음을 보여준다. 기자가 거리에서 만나 한 번 인용한 수많은 시민은 개개인의 중요도는 대통령보다 낮지만 언론사별 색깔을 만들고 언론계의 다양성을 보장한다.

지면별 분석 결과를 살펴보면 우선 정치 분야는 전체적으로 동조화가 강한 만큼 출입처 제도의 굴레를 벗어나 더 많은 각계 목소리를 담아야 할 필요성을 시사한다. 사회면과 경제면은 그러한 면에서는 비교적 균형이 잡혀 있다. 즉 출입처 정보원을 주요 정보원으로 인용하지만, 상대적으로 다양한 정보원에게 개방적이다. 문화 분야는 다양성 측면에서 상당히 뛰어나지만, 정보원 수 자체를 늘려야 할 필요가 있다. 국제면에서는 취재 범위를 고려할 때 파편화 정도가 가장 커야 함에도 불구하고 평균 수준에 그쳤다. 기사 작성 시 해외 매체에 대한 의존도를 낮출 필요가 있다.

무엇보다 동조화보다 파편화가 두드러지며, 두터운 꼬리를 구성하는 압도적 다수의 하위권 정보원에 의해 일어난다는 점은 한국 언론계가 그래도 민주적이고 다원주의적 성격을 갖고 있다는 것을 암시한다. 소수의 주요 공식적 정보원보다, 중요도 낮지만 압도적 다수가 다원적 여론 생태계를 주도하고 있다는 뜻이기 때문이다.

본 연구의 결과는 방법론적 함의도 갖고 있다. 매체간 비교 연구를 위해서는 상위권 정보원을 비교하는 것이 아니라 중하위권 정보원을 비교해야 한다. 왜냐하면 상위권 정보원 간에는 오히려 매체 간 동조화가 심화되며, 차별화, 즉 파편화는 중위권 이하에서 시작되어 하위권에서 두드러지기 때문이다. 기존의 매체 간 비교 연구는 흔히 특정 시기, 특정 주제의 소수 매체에 대해 통상 20위권 이내의 정보원을 비교하지만, 이 경우 진정한 매체 간 차이를 놓칠 수 있다. 20위권 이내만 보면, 전체적 관점에서는 동조화된 구간을 다소 자의적으로 세분화하여 연구자가 보고 싶은 차이만 분석하는 것으로 볼 수 있다. 20위권 이내의 정보원만 분석하고자 한다면, 개별 매체 간 비교 연구보다는 전체 시스템 차원의 연구, 즉 다양한 매체를 묶어서 언론계 전체에 대한 분석을 하는 것이 적합하다.

매체 간 차이는 전반적인 관점에서는 본 연구에서 제안한 파편화 지수로 그 분포를 파악함으로써 발견할 수 있다. 그러나 매체 간 핵심적인 차이를 내용 측면에서 보기 위해서는 중하위권 정보원, 특히 두터운 꼬리를 분석해야 한다. 문제는 이러한 정보원이 하나의 논문에서 기술하기 어려운 정도로 많다는 점이다. 이를 분석하려면 우선 다각도의 질적 연구를 수행해야 한다. 이는 마치 인류에 관한 연구를 위해 매우 특별한 마을 사례를 두텁게 쓰기로 연구하는 민속지학과 유사하다. 다양한 사례에 대한 차별화된 민속지학이 수없이 수행되면서 전체적인 그림이 그려지듯, 매체 간 차이는 수많은 중하위권 정보원에 관한 다양한 연구를 통해 밝혀질 수 있다. 즉 일종의 민속지학적 담론분석이 필요하다.

본 연구의 제한점 및 후속 연구를 위한 제언은 다음과 같다. 본 연구는 데이터 정제를 최소화한 상태에서 분석했다. 비록 분석 결과가 오측에 강건하기는 하지만, 보다 엄밀한 연구를 위해

서는 더 좋은 AI 기반 자연어처리 도구를 활용하고 조심스럽게 정제를 한 뒤 분석해볼 필요가 있다. 즉 단순한 특수문자 처리를 넘어서, 익명 정보원을 제거하고 사실상의 이명동인 표기를 수정하는 등의 방대한 작업을 수행한 뒤 다시 분석할 필요가 있다. 단 이러한 조치는 상위권(1~3구간)에서는 동조화를, 중하위권(4구간 이하)에서 과편화를 심화하는 식으로 본 연구의 분석 결과를 더욱 분명하게 뒷받침할 것으로 기대한다.

과편화 지수를 표준화함으로써 일반화 가능성을 높이는 방안도 모색할 수 있다. 즉 연도별 지면별로 정보원 수나 매체 수가 달라서 최종 순위도 달라지고 그에 따라 x축의 구간별 폭이나 y축의 과편화 지수 크기에 차이가 생겨난다. 이를 표준화할 경우, 구간의 폭이나 과편화 지수가 특정 상수값에 수렴할 가능성도 있다.

일반화 가능성을 더욱 높이기 위해 분석 매체와 지면을 확장할 필요도 있다. 매체의 경우 종합지나 경제지를 늘릴 수도 있으며, 지역지나 전문지, 인터넷신문, 방송, 잡지 등으로 유형을 확대할 수도 있다. 국내 언론사만이 아니라 영미권 등 해외 언론사의 보도도 분석할 필요도 있다. 국내 매체의 경우 빅카인즈의 수집 매체 수가 104개로 늘어났으며, 매체 유형도 다양화된 만큼 활용 가능성이 높아졌다.

동조화된 과편화 현상에 대한 구체적인 내용분석 내지 담론분석도 유용하다. 매체가 상위권에서 동조화돼 있다면 그들은 누구인가? 반대로 중하위권에서 누가 과편화를 주도하는가? 선행 연구를 고려하면 동조화는 정부나 대기업 등 공식적 정보원이나 일반가(generalist) 성격의 교수, 평론가, 컨설턴트, 애널리스트 등이, 과편화는 다양한 소속의 정보원이 주도할 것으로 보인다.

본 연구는 정보원 중심으로 동조화된 과편화 현상을 분석했다. 앞서 기존 문헌 검토에서 살펴본 것처럼, 동조화나 과편화는 주제어, 의제, 프레임이나 인지 수준에서도 일어날 수 있다. 예컨대 주제어 역시 과편화 지수를 사용해 분석 가능하다. 정보원의 동조화된 과편화 현상으로 미루어볼 때 주제어 역시 상위권 주제어에 대해서는 동조화가 일어났다가 중하위권 주제어에서는 과편화가 일어날 가능성이 높다. 다만 이는 실제 뉴스 주제어 연결망 분석을 통해 확인되어야 한다.

## References

- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2), 203-226.
- Bennett, W. L. (1990). Toward a theory of press-state relations in the United States. *Journal of Communication*, 40(2), 103-125.
- Baum, M. A., & Groeling, T. (2008). New media and the polarization of American political discourse. *Political Communication*, 25(4), 345-365.
- Boczkowski, P. J. (2010). *News at work: Imitation in an age of information abundance*. University of Chicago Press.
- Brown, J. D., Bybee, C. R., Weardem, S. T., & Straughan, D. M. (1987). Invisible power: Newspaper news sources and the limits of diversity. *Journalism Quarterly*, 64(1), 45-54.
- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 591-646.
- Culbertson, H. M. (1980). Veiled attribution: An element of style? *Journalism Quarterly*, 55(3), 456-465.
- Davies, N. (2008). *Flat earth news: An award-winning reporter exposes falsehood, distortion and propaganda in the global media*. Chatto & Windus.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745.
- Entman, R. M. (2003). Cascading activation: Contesting the White House's frame after 9/11. *Political Communication*, 20(4), 415-432.
- Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2.
- Han, D., & Lim, J. (2001). *An examination of media's utilization of news sources and hegemonic struggle*. Proceedings of the Fall Conference of the Korean Journalism Association, Seoul. [한동섭·임종수 (2001). <미디어의 뉴스원 활용과 헤게모니 투쟁에 대한 고찰>. 한국언론학회 가을철 정기학술대회 발표논문집. 서울: 서강대학교.]
- Harp, D., Loke, J., & Bachmann, I. (2010). Voices of dissent in the Iraq war: Moving from deviance to legitimacy?. *Journalism & Mass Communication Quarterly*, 87(3-4), 467-483.

- Herman, E. S., & Chomsky, N. (1988). *Manufacturing consent: The political economy of the mass media*. Pantheon Books.
- Horkheimer, M., & Adorno, T. W. (1944). *Dialectic of enlightenment*. [김유동 (역) (2001). <계몽의 변증법>. 문학과지성사.]
- Iyengar, S. (1991). *Is anyone responsible? How television frames political issues*. University of Chicago Press.
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19-39.
- Kang, B. N. (2010). *Complex network science*. Jipmundang. [강병남 (2010). <복잡계 네트워크 과학>. 집문당.]
- Kim, K., & Noh, G. (2011). A comparative study of news reporting about North Korea on newspapers in South Korea. *Korean Journal of Journalism & Communication Studies*, 55(1), 361-387. [김경희·노기영 (2011). 한국 신문사의 이념과 북한 보도방식에 대한 연구. <한국언론학보>, 55권 1호, 361-387.]
- Korea Press Foundation (2023). *Newspaper industry in Korea 2023*. Korea Press Foundation. [한국언론진흥재단(2023). <2023 신문산업 실태조사>. 서울: 한국언론진흥재단.]
- Lee, J. (2022). Differences in the use of North Korean experts by conservative and progressive media: Focusing on KPF-BERT-based deep-learning analysis of expert quotes. *Korean Journal of Journalism & Communication*, 66(6), 154-194. [이종혁 (2022). 보수 언론과 진보 언론의 북한 전문가 활용 방식의 차이 탐색: 인용문에 대한 KPF-BERT 기반 딥러닝 분석을 중심으로. <한국언론학보>, 66권 6호, 154-194.]
- Lee, W., & Choi, M. (2020). Language and semantic structure of the crime news headline: Through the analysis of semantic networks for each crime type reported on the portal site 'Naver'. *Korean Journal of Communication & Information*, 101, 367-398. [이완수·최명일 (2020). 범죄뉴스 헤드라인 언어와 의미구성: 포털 사이트 '네이버'에 보도된 범죄 유형별 의미연결망 분석을 통해. <한국언론정보학보>, 101호, 367-398.]
- Levendusky, M., & Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes?. *Political Communication*, 33(2), 283-301.
- Lewis, J., Williams, A., & Franklin, B. (2008). A compromised fourth estate? UK news journalism, public relations, and news sources. *Journalism Studies*, 9(1), 1-20.

- Napoli, P. M. (2011). *Audience evolution: New technologies and the transformation of media audiences*. Columbia University Press.
- Neuman, R., Guggenheim, W., Mo Jang, S., & Bae, S. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193-214.
- Noelle-Neumann, E. (1993). *The spiral of silence: Public opinion – Our social skin* (2nd ed.). University of Chicago Press.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Park, D. (2013). News source network analysis as big data analytics of news articles. *Korean Journal of Journalism & Communication*, 57(6), 234-262. [박대민 (2013). 뉴스 기사의 빅데이터 분석 방법으로서 뉴스 정보원 연결망 분석. <한국언론학보>, 57권 6호, 234-262.]
- Park, D. (2014). Biases by bursts of quoting sources in journalism. *Communication Theories*, 10(1), 295-324. [박대민 (2014). 뉴스 정보원 인용에서의 폭발성과 언론의 편향성. <커뮤니케이션 이론>, 10권 1호, 295-324.]
- Park, D. (2015). A study of double validity claims in quotations: News source network analysis of news on the four major rivers project in the Dong-A Ilbo and the Hankyoreh. *Korean Journal of Journalism & Communication Studies*, 59(5), 121-151. [박대민 (2015). 사실기사의 직접인용에 대한 이중의 타당성 문제의 검토: 동아일보와 한겨레신문의 4대강 추진 논란 기사에 대한 뉴스 정보원 연결망 및 인용문 분석. <한국언론학보>, 59권 5호, 121-151.]
- Park, D. (2016). Automated time series content analysis with news big data analytics: Analyzing sources and quotes in one million news articles for 26 years. *Korean Journal of Journalism & Communication*, 60(5), 353-407. [박대민 (2016). 장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26년. <한국언론학보>, 60권 5호, 353-407.]
- Park, D. (2023). Topology of media bias: Fat-tailed distribution as universal distribution of quotation by analyzing news source networks with 16.5 million articles. *Korean Journal of Journalism & Communication*, 67(6), 185-218. [박대민 (2023). 편향의 위상학: 1650만 건 기사의 뉴스 정보원 연결망 분석을 통해 파악한 인용 방식의 보편적 분포로서 두터운 꼬리 분포. <한국언론학보>, 67권 6호, 185-218.]
- Park, D.-M., & Lee, H.-J. (2024). Literature review of ai hallucination research since the advent of ChatGPT: Focusing on papers from arXiv. *Informatization Policy*, 31(2), 3-38. [박대민·이한중 (2024). 챗GPT 등장 이후 인공지능 환각 연구 문헌 검토. <정보화정책>, 31권 2호, 3-38.]

- Park, J. (2020). A comparative study on the 'Corona19' news frame based on ideological orientation of media. *Korean Journal of Journalism & Communication*, 64(4), 40-85. [박주현 (2020). 언론의 이념 성향에 따른 '코로나19' 보도 프레임 비교 연구. <한국언론학보>, 64권 4호, 40-85.]
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
- Reese, S. D., & Danielian, L. H. (1989). Intermedia influence and the drug issue: Converging on cocaine. In P. J. Shoemaker (Ed.), *Communication campaigns about drugs: Government, media, and the public* (pp. 29-46). Erlbaum.
- Ross Arguedas, A., Robertson, C., Fletcher, R., & Nielsen, R. (2022). *Echo chambers, filter bubbles, and polarisation: A literature review*. Reuters Institute for the Study of Journalism.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2), 143-186.
- Sohn, D. (2022). Spiral of silence in the social media era: A simulation approach to the interplay between social networks and mass media. *Communication Research*, 49(1), 139-166.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556-576.
- Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press.
- Strogatz, S. H. (2012). *Sync: How order emerges from chaos in the universe, nature, and daily life*. Hachette Books.
- Sunstein, C. (2001). *Republic.com*. Princeton University Press.
- Swain, B. M., & Robertson, M. J. (1995). The Washington Post and the Woodward problem. *Newspaper Research Journal*, 16(1), 2-21.
- van Dijk, T. A. (1988). *News as discourse*. Lawrence Erlbaum.
- Wilson, R. J. (1979). *Introduction to graph theory*. Pearson Education India.

최초 투고일 2024년 10월 04일  
 게재 확정일 2025년 02월 01일  
 논문 수정일 2025년 02월 02일