Journal of Korean Library and Information Science Society
Vol. 57, No. 2

OPEN ACCESS

문헌정보학 분야 프리프린트에 관한 연구*

- arXiv를 중심으로 -

Preprints in Library and Information Science: A Focus on arXiv

조 재 인(Jane Cho)**

〈 목 차 〉

Ⅰ. 서 론
Ⅱ. 이론적 배경
Ⅲ. 연구방법
Ⅳ. 분석 결과
Ⅴ. 요약 및 논의
Ⅵ. 결 론

요 약 : 본 연구는 arXiv에 기탁된 문헌정보학 분야의 프리프린트(preprint)를 대상으로 피인용과 알트메트릭스(altmetrics)를 분석하였다. 특히 분석 대상 논문 중 기탁 후 최소 1년 반이 경과했음에도 불구하고 아직 공식 출판에 이르지 못한 프리프린트에 주목함으로써, 학술적 검증이 완료되지 않은 연구 결과가 학술 커뮤니케이션 체계 내에서 어떠한 영향력을 형성하고 있는지를 살펴보았다. 첫째, 프리프린트의 주제 분야는 전통적인 도서관학보다는 인용 및 데이터 분석을 중심으로 한 계량정보학 분야에 집중되어 있음이 확인되었다. 둘째, 공식 출판이 이루어지지 않은 미검증 프리프린트가 대중 매체를 통해 사회 전반에 확산되는 경우는 많지 않지만, 학술 커뮤니티 내에서는 조기에 공유되어 활용이 이루어지는 것으로 분석되었다. 또한 그 중 일부 프리프린트는 영향력 있는 저널에 인용되고 있는 것으로 확인되었다. 셋째, 후속 연구에 의해 피인용된 미검증 프리프린트는 인용되지 않은 프리프린트보다 AAS(Altmetric Attention Score)가 유의하게 높게 나타나 ($p < 0.001$), 소셜 미디어를 통한 사회적 파급이 더 클 수 있다는 것이 검증되었다. 이러한 현상은 검증되지 않은 사실의 확산을 초래할 수 있어, 학술지의 편집자나 연구자들의 신중한 판단과 명확한 편집 및 평가 기준이 필요함을 시사한다.

주제어 : 프리프린트, 문헌정보학, arXiv, 학술논문, 알트메트릭스, 피인용, BERTopic 모델링

ABSTRACT : This study analyzed the altmetrics and citation characteristics of preprints in the field of library and information science (LIS) deposited on arXiv. By focusing on the social attention and citation patterns of unverified preprints that had not been published despite being at least a year and a half old, we examined the impact and potential risks of preprints in scholarly communication. The analysis revealed that, first, preprints classified under LIS were predominantly associated with information science and informetrics, as identified. Second, while dissemination of unpublished preprints through mass media such as news and academic blogs was limited, early sharing and utilization occurred within the academic community. Furthermore, some were cited in subsequent papers, leading to citations in high-impact journals (SJR). Third, cited unverified preprints had significantly higher Altmetric Attention Scores (AASs) and greater social impact than uncited preprints. This phenomenon requires editors and researchers to carefully consider the value and risks of utilizing preprints, and suggests the need for active discussion on future citation criteria and utilization methods for preprints.

KEYWORDS : Preprint, Library and Information Science (LIS), arXiv, Scholarly Articles, Altmetrics, Citations, BERTopic Modeling

* 본 논문은 2025년도 인천대학교 자체연구비 지원으로 수행되었음.
** 인천대학교 문헌정보학과 교수(chojane@inu.ac.kr / ISNI 0000 0004 6016 5878)

• 논문접수: 2026년 5월 14일 • 최초심사: 2026년 6월 1일 • 게재확정: 2026년 6월 15일
• 한국도서관·정보학회지, 57(2), 257-278, 2026. http://dx.doi.org/10.16981/kliss.57.2.202606.257

※ Copyright © 2026 Korean Library and Information Science Society
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.

Ⅰ. 서 론

공식 출판 과정에서 나타나는 시간 지연을 보완하기 위해 활용되는 프리프린트(preprint)는 신속한 학술 정보 교환을 도모하는 비공식 출판본이다. 프리프린트는 연구 결과를 신속하게 보급해야 하는 위기 상황에서 중요한 역할을 할 수 있다. 따라서 글로벌 위기 상황 대처가 필요한 COVID-19 팬데믹 기간 동안 신속한 연구 결과의 보급을 위하여 기하급수적으로 증가하게 되었다(Fraser et al., 2021). 대표적인 프리프린트 플랫폼인 arXiv는 1991년 Paul Ginsparg가 만든 이래로 물리학, 수학, 컴퓨터 과학과 같은 여러 분야에서 연구 결과를 신속하게 확산하는 데 중심적 역할을 해왔다. 그밖에도 SSRN, bioRxiv 및 medRxiv 등의 프리프린트 플랫폼이 존재하며, 이들은 비영리 학술단체, 학회, 연구비지원기관이 운영하거나 영리 출판사 등에서 운영되고 있다.

학계에서는 Ingelfinger 규칙(Ingelfinger, 1977)에 의해, 한동안 동일한 연구 결과가 프리프린트에 공개되었을 경우 해당 논문을 학술지에 출판하지 않는 관행이 존재해 왔다. 이 규칙은 그동안 프리프린트의 성장과 개발을 억제해 왔는데, 최근 몇 년 동안 연구 결과의 빠른 전달과 피드백, 과학적 발견에 대한 우선 순위 확립 필요성이 증가하고 상업 출판사가 그린 오픈 액세스 아카이빙을 수용하면서 완화되었다(Larivière et al., 2014). 더불어 오픈사이언스 기조 하에 저널 출판 중심에서 소셜 미디어를 포함하는 학술 커뮤니케이션 방식의 다변화로(Okamura, 2022), 프리프린트 플랫폼은 지식 전파를 가속화하고 민주적 학술 소통을 도모하는 중요한 도구로 부상하게 되었다.

한편, 프리프린트는 연구자들 사이에서만 전달되는 것이 아니라, 사회 대중을 포함한 더 광범위한 독자 그룹으로도 전파된다. 따라서 학술적 소통을 넘어선 사회적 영향력을 포착하고 이해하는 것이 중요하다. 이를 위해서는 피인용 지표 이외에 비공식적 영향력을 추적하는 대체적 평가 지표인 알트메트릭스(altmetrics)가 활용될 수 있다. 알트메트릭스는 소셜 미디어, 언론보도, 정책문서, 참고문헌 북마크 등을 통해 학술논문이 얼마나 사회적으로 언급되었는지를 추적하여 제시하는 지표이다.

프리프린트는 최근 문헌정보학(LIS: Library & Information Science) 분야에서도 주목되기 시작하고 있지만(Wang et al., 2018), 이와 관련된 연구는 아직까지 많이 이루어지지 않고 있다. 문헌정보학 분야 프리프린트가 사회적으로 얼마나 주목되고 있는지 그리고 프리프린트 자체로도 얼마나 많이 인용되고 있는지를 조사함으로써 이 분야에서 나타나고 있는 학술 커뮤니케이션 변화를 이해할 필요가 있을 것이다. 또한 이러한 시도를 통해 문헌정보학 분야에서 프리프린트의 학술적 가치와 의미가 어떻게 형성되어 자리매김할 것인지 예측해 볼 필요가 있을 것이다.

이러한 배경에서 본 연구는 arXiv에 기탁된 문헌정보학 분야의 프리프린트를 분석해 보고자 한다. 또한 그 중 프리프린트 기탁 후 최소 1년 반이 경과했음에도 불구하고 출판되지 않고 미검증 상태로 남아 있는 논문에 주목하여, 피인용과 알트메트릭스를 분석해 보고자 한다. 이를 위하여 다음과 같은 연구 질문을 설정하였다. 첫째, arXiv에 기탁된 LIS 분야 프리프린트는 주로 어떠한

주제의 논문인가? 이 논문은 얼마나 사회적으로 주목되고 있는가? 둘째, 미검증 상태로 존재하는 프리프린트도 사회적으로 주목되고 피인용되고 있는가? 셋째, 피인용된 미검증 프리프린트는 더 많이 사회적으로 전파될 가능성을 가지는가?

Ⅱ. 이론적 배경

1. 프리프린트

프리프린트는 학술지에 정식 출판되기 전에 프리프린트 서버를 통해 공개되는 학술논문의 출판 전 버전을 의미한다. 이를 통해 연구자는 학술지의 동료심사 및 출판 과정에서 발생하는 시간적 지연 없이 연구 결과를 신속하게 공개할 수 있다(조재인, 2025). 프리프린트 플랫폼은 점점 더 보편화되면서 광범위한 주제로 확장되고 있으며, 연구 결과에 대한 대중의 접근성을 높임으로써 학계와 비학계 독자 간의 정보 격차를 줄이는 역할을 하고 있다. 그러나 프리프린트에 실린 검증되지 않은 연구 결과가 여러 미디어 채널을 통해 광범위하게 유포되면서 비판을 받기도 하고 철회되는 경우도 발생하고 있다(Oh & Kang, 2023). 본 장에서는 먼저 프리프린트 유통이 가지고 있는 장점과 단점에 대하여 정리해 본다.

먼저 장점에 대하여 기술하면 다음과 같다. 첫째, 연구자는 유사한 주제와 방법의 경쟁 속에서 우선순위 논란을 피하기 위해 독창적 아이디어와 결과를 신속히 발표하는 것이 중요하다. 이러한 상황에서 프리프린트 기탁은 연구 결과에 대한 빠른 선점을 가능하게 한다. 둘째, 프리프린트는 더 많은 피드백을 유도하게 한다. 기존의 동료 검토 시스템 내에서 저자는 소수의 검토자와 편집자로부터 제한된 횟수의 피드백만을 받을 수 있었다(Lin et al., 2020). 그러나 프리프린트 기탁을 통해 논문이 공개된다면 광범위한 학술 커뮤니티를 통해 연구 결과에 대한 토론과 피드백을 이끌어 낼 수 있다. 셋째, 프리프린트는 독자들의 주의를 환기시키는 중요한 통로가 될 수 있다. 실제로 공식 출판에 앞서 프리프린트를 기탁한 논문은 초기 단계에서부터 연구자들의 관심을 유도하며, 그 결과 그렇지 않은 논문보다 이후 더 많은 인용을 획득한다는 사실이 여러 연구자들에 의해 실증적으로 입증되고 있다(Davis & Fromerth, 2007; Feldman et al., 2018; Wang et al., 2020). 넷째, 프리프린트는 오픈액세스를 통해 독자들이 보다 손쉽게 무료로 연구 성과에 접근할 수 있도록 한다. 이는 저널 구독이 어려운 기관이나 국가의 연구자들에게도 열람의 기회를 제공한다는 점에서 긍정적으로 평가되고 있다(Rodriguez et al., 2006; Wang et al., 2018).

프리프린트는 위와 같이 연구 결과를 대중에게 더 빨리 제공한다는 점에서 고무적이지만 윤리적 위험 초래 여부는 계속적인 논쟁이 되고 있는데(da Silva, 2017), 다음과 같이 정리할 수 있겠다.

첫 번째, 많은 사용자가 여전히 프리프린트를 심사를 거친 원고와 혼동할 수 있다는 점이다. 검증되지 않은 프리프린트가 널리 배포되었다면 연구 결과가 잘못된 것으로 판명되어도 사람들은 자신이 믿었던 것을 잊는 것이 쉽지 않기 때문이다(Bertin & Atanassova, 2022). 이는 특히 잘못된 정보가 생명을 위협할 수 있는 의학, 제약 및 생물 화학 분야에서 심각한 문제가 될 수 있다(Brown, 2003). 두 번째, 스쿠핑(scooping)이 존재할 수 있다(Oh & Kang, 2023). 스쿠핑이란 연구자가 장기간 실시한 연구 결과를 공식 발표하기 전에 경쟁 연구팀이 먼저 동일 또는 유사한 연구 결과를 출판하여 우선권을 빼앗거나 아이디어 제안자, 실제 결과 산출자를 적절히 인용하지 않고 발표하는 경우를 의미한다(Icahn School of Medicine at Mount Sinai, n.d.). 프리프린트는 연구 아이디어를 조기에 공개함으로써 우선권을 확보하는 수단이 될 수 있지만, 반대로 아이디어 도용이나 데이터 스쿠핑의 위험 요인이 될 수도 있다(Yi & Huh, 2021). 세번째 동일한 과학적 산출물에 대해 두 개의 서로 다른 고유 식별자가 공존해 인용이 분산되는 문제도 존재한다(Chaleplioglou & Koulouris, 2023). 프리프린트 서버에 등록된 논문은 공개 시점에 고유 식별자를 가질 수 있으며, 추후 학술지에 공식 게재될 때에도 별도의 식별자가 새로 부여될 수 있기 때문이다. 따라서 프리프린트로서 공개된 기간 동안 누적된 인용은 공식 출판본과는 별도로 집계되어 연구 성과 평가 과정에서 혼란이 야기될 수 있다.

2. 선행연구

선행연구에서는 프리프린트의 인용과 알트메트릭스 관련 주요 연구 결과를 고찰해 보며 커뮤니티 피어리뷰 모델 등 주목되는 최신 연구 결과를 정리해 본다.

첫째, Larivière et al.(2014)은 arXiv에 게재된 많은 프리프린트가 결국 심사 평가를 거쳐 공식 출판물에 게재되었고 그 비율은 약 73%에 이른다고 보고하였다. 또한 프리프린트는 그 자체로도 공식 출판물에 참고문헌으로 인용되고 있다고 덧붙였다. Li et al.(2015)도 Scopus 등재 저널에 수록된 논문이 arXiv, RePEc, SSRN 등에 기탁된 프리프린트를 직접 인용하고 있으며 그 숫자가 꾸준히 증가하고 있다고 언급하였다. Fraser et al.(2020)도 bioRxiv에 제출된 논문이 과학 커뮤니티에서 많은 관심을 받고 있다고 하였으며, 다른 많은 연구자들도 의생명 분야에서 프리프린트를 인용하는 것은 학술 출판 사이클의 일부가 되었다고 논평하고 있다(Hoy, 2020; Penfold & Polka, 2020). 또한 프리프린트 논문은 기탁되지 않은 논문보다 훨씬 더 많은 인용을 받는다는 보고도 있다(Fraser et al., 2020; Fu & Hughey, 2019; Serghiou & Ioannidis, 2018). 그 이유로는 조기 공개에 따른 가시성 제고, 오픈액세스가 제공하는 접근성 때문이다(Kurtz et al., 2005; Moed, 2007). 또한 연구자들은 일반적으로 본인의 연구 중 품질이 우수한 논문을 먼저 프리프린트로 공개하는 자가선택편향이 있기 때문에 프리프린트로 공개된 논문이 더 많이 피인용될 수 있다고 논의되고 있다(Kurtz et al., 2005).

둘째, 프리프린트의 다운로드 횟수가 인용을 예측할 수 있는지 살펴본 연구도 있다. Brody et al.(2006)은 프리프린트 사용 횟수가 인용 횟수를 예측할 수 있다고 하였으며, 조재인(2025)도 누적된 이용 지표를 사용하면 향후 공식 출판본에서 나타날 수 있는 영향력을 예측해, 연구 결과 선별에 도움을 줄 수 있다고 하였다. 이러한 연구들은 프리프린트가 추후 발행될 공식 출판본의 인용 예측에 일정한 역할을 할 수 있음을 시사한다. 한편 프리프린트의 이용도 분산을 분석한 연구도 존재한다. Henneken et al.(2007)은 프리프린트가 유일본으로 존재할 때에는 사용도가 집중되지만 출판이 되면 공식 출판본으로 대체되는 경향을 보인다고 하였다. 또한 같은 맥락에서 Davis와 Fromerth(2007)도 프리프린트로 기탁된 공식 출판본은 기탁되지 않은 논문에 비하여 이용 횟수가 분산될 수 있다고 언급하였다.

셋째, 본 연구와 같이 문헌정보학 분야의 프리프린트와 관련된 연구를 살펴보면 다음과 같다. Wang et al.(2020)은 508개의 arXiv 기탁 논문과 5,536개의 비기탁 논문의 피인용을 비교하였다. 그들은 문헌정보학의 3개 주요 저널을 대상으로 분석한 결과, arXiv 기탁 후 정식 출판된 논문에서 상당한 수준의 인용 강화 효과를 발견했다고 밝혔다. 또한 arXiv 기탁 후 출판된 논문은 알트메트릭스 특히 Mendeley와 Twitter의 주목도가 높았다고 관찰하였다. 프리프린트의 조기 접근 및 오픈액세스 속성이 소셜 미디어에서 논문에 대한 관심을 즉시 모을 수 있기 때문이다. 또한 그들은 다른 연구(Wang et al., 2018)를 통해 arXiv 기탁 논문의 비중은 비교적 낮음에도 불구하고, 절대 량과 저자 수는 최근 몇 년 동안 증가했다고 보고하였다. 더불어 프리프린트로 기탁된 논문의 인용률은 그렇지 않은 논문보다 약 2.5배 높고 출판 후 3년 동안 더 높은 초기 인용률을 달성할 수 있다고 설명하였다. 마지막으로 Chingath(2024)는 arXiv에 기탁된 LIS 프리프린트를 대상으로 알트메트릭스를 분석하고 인용 지표간의 관련성을 확인하였다. 그는 Twitter, Facebook, 정책문서에서의 언급 빈도와 인용 지표 간의 상관관계를 분석한 결과, 양(+)의 관련성이 존재한다고 언급하면서, 프리프린트 기탁은 연구 결과의 확산에 기여할 수 있다고 주장하였다.

넷째, 프리프린트와 관련된 주목되는 최근 연구 결과를 살펴보도록 한다. Bertin과 Atanassova(2022)는 연구자가 주로 논문의 어떠한 부분에서 프리프린트를 인용하게 되는지 분석하였다. 그 결과 22-44%가 방법 섹션에서 프리프린트를 주로 인용하고 있다고 하였다. 또한 공식 출판본을 인용할 때에는 'show', 'suggest', 'argue'와 같이 학술적 기여에 대한 주장을 표현하는 동사를 사용하는 반면, 프리프린트 인용시에는 'use', 'perform'과 같은 기술적, 방법론적 기여를 나타내는 동사를 사용해 완화된 표현을 한다고 주장하였다. 이는 연구자들이 프리프린트를 완성된 학술 성과라기보다 연구 과정에서 참고 가능한 잠정적 지식이나 방법론적 자원으로 인식하고 있음을 시사한다. 한편 Wang과 Zhan(2019)은 프리프린트에 적용 가능한 피어리뷰 모델로 Self-Organizing Peer Review(SOPR)를 제안해 주목된다. 이 모델은 리커트 척도 기반의 이중맹검 방식으로 운영되며, 프리프린트를 제출한 저자가 다른 프리프린트 3편의 리뷰를 의무적으로 수행하도록 설계되

어 있다. 이러한 구조는 프리프린트를 대상으로 한 동료 평가의 지속 가능성을 확보할 수 있으며, 커뮤니티 피어 리뷰에서 나타날 수 있는 불공정성과 관심 편중 문제를 보완할 수 있다고 주장하였다.

Ⅲ. 연구방법

1. 데이터 수집

본 장에서는 연구에 필요한 데이터 수집과 분석 방법에 대하여 기술한다. 먼저, arXiv에 기탁된 프리프린트의 메타데이터와 알트메트릭스 및 피인용 지표를 추출하기 위한 방법을 설명한다. arXiv의 주제 분류 체계에서는 문헌정보학에 해당하는 독립된 분류가 제공되지 않기 때문에, 이 플랫폼에서는 해당 분야 프리프린트를 정확히 선별하는 것이 불가능하다. 그러한 이유로 기계학습 기반 문서 단위 주제 분류(ANZSRC For: Field of Research)를 제공하는 Dimensions(https://www.dimensions.ai/)를 활용해 arXiv에 아카이빙된 LIS 프리프린트를 추출하였다. Web of Science보다 3배 많은 저널을 포괄하는 인용 색인 데이터베이스인 Dimensions는 2024년 11월 기준 1억 4천만 건 이상의 출판물을 포함하며, arXiv도 색인에 포함하고 있다(Dimensions, 2019; Singh et al., 2021). 또한 주제 전문가가 분류한 데이터를 활용한 기계 학습 알고리즘을 통해 새로운 논문을 분류하고 있기 때문에(Dimensions, 2022), LIS 분야 프리프린트를 효과적으로 선별할 수 있다. 본 연구에서는 Dimensions에서 source title을 arXiv로 설정한 뒤, “Library and Information Studies”에 분류되고 2022년에 기탁된 문헌정보학 프리프린트 433건을 최종 분석 대상으로 선정하였다. 수집 및 분석 시점이 2024년임에도 불구하고 arXiv 등록 기준을 2022년도로 설정한 이유는 기탁 이후 공식 출판본이 생성될 수 있는 충분한 시간이 경과한 프리프린트를 선별하기 위함이다. 일반적으로 학술지는 원고가 제출되면, 초기 검토, 동료 심사, 수정 및 재심사, 최종 승인 후 편집 및 제작 과정을 거친다. 학문분야에 따라 상이하지만 인문사회과학 분야의 경우, 약 12-18개월, 경우에 따라 더 지연되면 2년 가까이 소요되기도 한다(Björk & Solomon, 2013). 본 연구는 안정적으로 프리프린트의 공식 출판 여부를 식별하기 위하여 2022년도 최초 기탁이 이루어져 2024년 중반 까지 18개월에서 30개월이 경과한 프리프린트를 분석 대상으로 선정하였다.

한편 추출된 433건을 대상으로 공식 출판본과 해당 논문을 피인용한 논문의 존재 여부를 다음과 같이 확인하였다. Dimensions에서는 공식 출판본이 존재할 경우 ‘resulting publication’ 링크를 통해 해당 논문의 출판 후 버전을 연계하고 있다. 본 연구에서는 먼저 이 링크의 존재 여부를 확인하였으며, 이어 교차 검증 차원에서 Altmetrics Explorer을 통해 arXiv 이외의 또 다른 출처를 보유한 동일 논문이 존재하고 있는지 일일이 확인하였다. 한편 Dimensions에서는 공식 출판본이

존재하지 않는 프리프린트라 하더라도 해당 프리프린트 자체가 인용된 경우 인용 논문에 대한 정보(publication citation)를 제공하고 있다. 본 연구에서는 이 정보를 활용하여 해당 프리프린트를 인용한 저널의 명칭과 영향력 지표로서 SJR(SCImago Journal Rank)도 함께 수집하였다.

2. 분석 방법

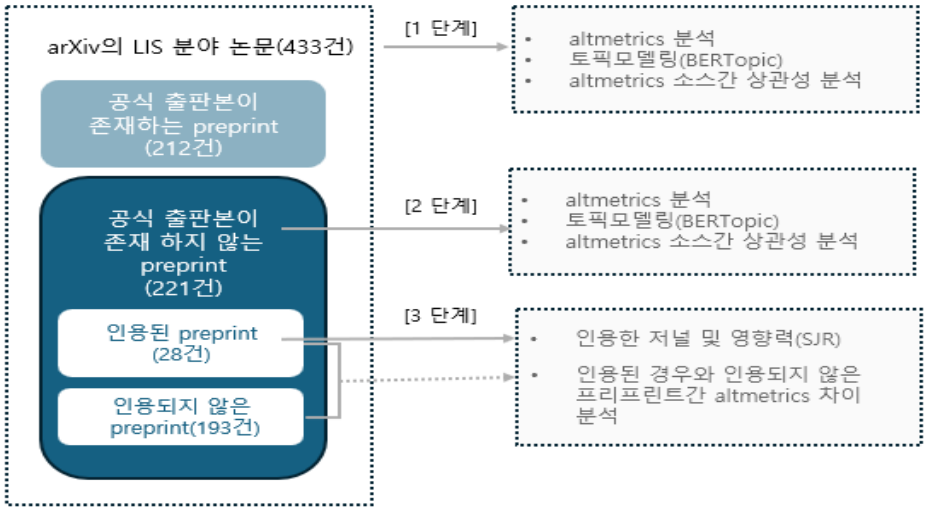
수집된 정보는 다음과 같이 분석되었다.

먼저, 1단계에서는 2022년 arXiv에 기탁된 문헌정보학(LIS) 분야 프리프린트 433편을 대상으로 알트메트릭스 및 피인용 정보를 통계적으로 분석하였다. 또한 프리프린트의 주제 경향을 탐색적으로 분석하기 위하여 논문의 제목을 활용해 Python 환경에서 BERTopic(BERT-based topic modeling) 모델을 활용한 토픽 모델링을 수행하였다. Sentence-Transformers 기반 사전학습 언어모델로 문서를 벡터화한 후, UMAP(Uniform Manifold Approximation and Projection)을 통해 차원을 축소하였다(n_neighbors=15, n_components=5, min_dist=0.0, metric=cosine). 이후 KMeans 군집화 기법을 적용하여 문서를 8개의 토픽으로 분류하였으며, 군집별 대표 키워드 추출을 위해 c-TF-IDF(class-based Term Frequency-Inverse Document Frequency)를 적용하였다. 본 연구의 토픽 모델링은 프리프린트의 세부 주제 구조를 정밀하게 규명하기 위한 목적이 아니라 분석 대상 문헌의 전반적인 주제 경향을 탐색적으로 파악하기 위한 목적으로 수행되었다. 이에 따라 연구 주제를 가장 압축적으로 표현하는 제목을 분석 대상으로 활용하였다. 한편 알트메트릭스 데이터는 Altmetric사의 연구자 지원 프로그램을 통해 제공된 Altmetrics Explorer 라이선스를 활용해 2024년 7월 기준으로 수집하였음을 밝힌다(Altmetric, 2024)[1)]

둘째, 2 단계에서는 추출된 전체 프리프린트 중 공식 출판본이 존재하지 않는 221편을 대상으로 기술 통계 분석과 토픽 모델링을 통한 주제 분석을 실시하고, 1단계 분석 결과와 어떠한 차이가 나타나는지 확인하였다. 더불어 알트메트릭스 소스간 상관관계를 다시 산출하고 역시 전체 데이터 셋과의 차이를 비교하였다.

셋째, 3 단계에서는 공식 출판본이 존재하지 않는 데이터 셋 가운데 피인용이 발생한 논문 28건을 선별하였다. 그리고 해당 논문을 피인용한 저널명과 저널의 영향력(SJR: SCImago Journal Rank) 지수를 조사하였다. 이를 통해 검증 이전 단계의 프리프린트가 영향력 있는 저널의 학술 커뮤니케이션 과정에서 어떻게 수용되고 있는 지를 살펴보았다. 또한 피인용이 발생한 미검증 프리프린트와 그렇지 않은 프리프린트간에 사회적 파급력에 차이가 나타는지 Mann-Whitney 비모수 분석을 통해 양 집단의 AAS(Altmetric Attention Score) 차를 검증하였다. 위에서 언급한 분석 프레임워크를 도식화하면 아래 〈그림 1〉과 같다.

1) The authors wish to thank Altmetric for providing this study's data free of charge for research purposes.



〈그림 1〉 분석 프레임워크

Ⅳ. 분석 결과

1. arXiv에 기탁된 LIS 분야 프리프린트 분석

1단계로 분석 대상 프리프린트 전체의 알트메트릭스를 분석한 결과를 제시하면 다음 〈표 1〉과 같다. 알트메트릭스 지수(AAS) 평균은 11.58로 나타났으며, X(구 Twitter)에 한번이라도 언급된 논문은 388건으로 전체 논문의 90%로 나타났다. X는 언급 평균 또한 12.50회로 가장 높게 나타났지만, 리트윗이나 출판사 홍보 활동과 같은 요인으로 인해 다소 부풀려 나타났을 수도 있다. 한편 뉴스에 언급된 경우는 27건으로 전체의 6%에서만 나타났으며 알트메트릭스 소스 중 평균(0.25회)이 가장 작았다. 한편, 높은 AAS로 주목되는 프리프린트를 살펴보면 다음과 같다. 22년 6월에 아카이빙된 학술지 편집자에 관한 계량정보학 연구가 대표적이다. 이 논문은 언론보도와 학술블로그, X 등에 다수 언급되면서 사회적으로 주목된 것으로 나타났으며, 결국 23년 1월에 공식 출판된 것으로 확인되었다. 또한 22년 9월 아카이빙된 대면 학술회의의 지식 확산에 대한 연구도 역시 X를 통해 회자되면서 높은 AAS가 누적되었는데 이 경우는 전자와 달리 공식 출판 기록은 부재한 것으로 확인되었다. 그 밖에 AAS 상위 10편의 프리프린트의 출판 여부를 확인한 결과, 그중 5편은 공식 출판본이 존재하였으나 나머지 5편은 확인되지 않아, 사회적 주목을 받은 프리프린트라 하더라도 모두 공식 학술출판으로 이어지는 것은 아님을 보여준다.

〈표 1〉 arXiv에 아카이빙된 프리프린트의 알트메트릭스

알트메트릭스 소스	N	언급된 논문수	최대값	평균
AAS	433	388	402	11.58
뉴스	433	27	20	.25
블로그	433	77	10	.28
X	433	361	400	12.50
Mendeley	433	239	80	6.66

둘째, LIS의 어떠한 연구 주제가 상대적으로 더 많이 프리프린트로 아카이빙되고 있는지 확인하기 위하여 〈표 2〉와 같이 프리프린트의 제목을 활용해 BERTopic 기반 모델링을 수행하였다. 분석 결과, 학술평가・인용・동료심사(80편), 디지털 도서관 및 정보관리 시스템(72편), 과학계량학 및 연구성과 분석(62편), 오픈데이터 및 연구데이터 생태계(58편), 학술출판 및 오픈액세스(50편) 등이 주요 토픽으로 도출되었다. 예를 들어 학술평가・인용・동료심사 토픽에는 「Are We Standing on Unreliable Shoulders? The Effect of Retracted Papers Citations on Previous and Subsequent Published Papers」와 같은 연구가 포함되었으며, 과학계량학・연구성과 분석 토픽에는 「Bibliometric analysis of the scientific production found in Scopus and Web of Science about business administration」과 같은 연구가 대표 문헌으로 확인되었다. 또한 학술출판・출판행태 연구 토픽에는 「Library and Information Science Scholarly Journals Publishing Simulation」와 같이 학술지 출판과 학술 커뮤니케이션을 다루는 연구가 다수 포함되어 있었다. 즉 학술평가, 인용, 오픈데이터 등 계량정보학 및 오픈사이언스와 밀접하게 관련된 주제가 토픽의 대부분을 차지하고 있는 것으로 나타났다. 이는 arXiv에 등록된 LIS 연구가 전통적 도서관 운영이나 제도적 관리 연구보다는 학술 커뮤니케이션 구조, 연구성과 측정, 데이터 공유와 같은 지식의 생산・유통・평가 체계를 다루는 정보학 중심 영역에 집중되어 있음을 시사한다. 나아가 데이터 기반 분석과 계량적 방법을 활용하는 연구가 주요 흐름을 형성하고 있다는 점에서, 공학 및 컴퓨터과학과의 학제적 연계를 갖는 정보학 계열 연구자들이 프리프린트 기탁 관행에 상대적으로 적극적일 가능성을 추론해 볼 수 있을 것이다.

〈표 2〉 프리프린트의 BERTopic 기반 모델링

토픽	N	주요 키워드	토픽	N	주요 키워드
1. 학술평가, 인용, 동료심사	80	peer, quality, review, author, citation	5. 오픈 데이터・연구데이터 생태계	58	data, open, research
2. 디지털 도서관 및 정보관리 시스템 연구	72	library, digital, information, software	6.학술출판・출판행태 연구	50	publishing, scholarly, publications, scientific
3. 과학계량학・연구성과 분석	62	research, scientific, impact, collaboration, productivity	7. 인용 데이터 ・자연어 기반 학술 데이터 분석	33	citation, language, bibliometric, data, detecting
4. 그래프 기반 지식표현 및 머신러닝 응용 연구	60	knowledge, learning, graph, networks	8. 천문학 분야의 데이터・코드・연구 인프라 연구	18	astronomy, astrophysics, library

셋째, 일반적으로 소셜 미디어간에는 높은 관련성이 나타나, 언론에 자주 보도된 논문이 학술 블로그에도 많이 포스팅되며, SNS에서도 다수 회자될 수 있다. arXiv에 등록된 LIS 프리프린트에서도 비슷한 현상이 나타나는지 확인한 결과, 블로그와 뉴스 간에는 r = 0.388(p 〈 0.01)의 유의한 상관관계가 나타났다(〈표 3〉). 이는 학술적 발견이 뉴스 보도와 학술 블로그를 통해 동시에 언급되면서 대중에게 확산되는 경향을 반영할 수 있다. 이와 관련해 Fraumann과 Colavizza(2022)도 뉴스에 등장한 논문이 학술 블로그에서도 자주 다루어지고 있어, 이 두 매체가 과학적 발견을 대중에게 전달하는 주요 경로임을 시사한 바 있다. 한편 X와 Mendeley 간에 나타난 관련성도 살펴볼 필요가 있다. 여기에서도 r = 0.555(p 〈 0.01)의 비교적 높은 상관관계가 나타나 X에서 활발히 언급되는 논문일수록 Mendeley에서도 북마크될 가능성이 높은 것으로 나타났다. 또한 Mendeley 북마크는 피인용과의 관련성이 이미 여러 연구에서 확인된 바(Ebrahimy et al., 2016; Thelwall & Wilson, 2016), 이러한 결과는 X에서의 대중적 관심이 이후 학술적 관심과 피인용으로 이어질 가능성을 시사할 수 있겠다(Shuai et al., 2012). 앞서 언급한 바와 같이 1단계의 데이터셋에는 미검증 프리프린트와 공식 출판본이 존재하는 프리프린트가 혼재되어 있는데, 위의 분석 결과는 출판이 완료된 학술성과물 간에서 나타난 일반적 상관성(Fraumann & Colavizza, 2022)과 크게 다르지 않은 것으로 확인되었다. 이에 미검증 프리프린트만을 대상으로 제한해 그 차이를 살펴볼 필요가 있는데, 그 결과는 다음 챕터에서 제시해 보도록 하겠다.

〈표 3〉 아카이빙 프리프린트의 알트메트릭스 소스 간 상관성 분석 결과

		뉴스	블로그	X	Mendeley
뉴스	상관계수	1.000	.338**	.244**	.132**
	유의확률 (양측)	.	〈.001	〈.001	.006
	N	433	433	433	433
블로그	상관계수	.338**	1.000	.363**	.233**
	유의확률 (양측)	〈.001	.	〈.001	〈.001
	N	433	433	433	433
X	상관계수	.244**	.363**	1.000	.555**
	유의확률 (양측)	〈.001	〈.001	.	〈.001
	N	433	433	433	433
Mendeley	상관계수	.132**	.233**	.555**	1.000
	유의확률 (양측)	.006	〈.001	〈.001	.
	N	433	433	433	433

**p 〈 .01

2. 공식 출판본이 존재하지 않는 프리프린트 분석

본 장에서는 2단계로 프리프린트로 아카이빙된 이후 최소 1년 반이 경과된 논문 중 공식 출판본이

존재하지 않는 사례를 집중적으로 분석하였다. 공식 출판본이 존재하지 않는 경우는 221건(52%)으로 나타나, 분석 대상 프리프린트 중 절반 이상은 공식 학술지에 게재되지 않은 채 프리프린트 상태로 남아 있는 것으로 확인되었다. 다만, 게재되지 않은 논문이라고 하여 반드시 투고에 실패했거나 게재가 거절된 경우로 보기는 어렵다. 일부는 심사 중이거나 출판을 대기 중인 상태일 수 있으며, 또 다른 일부는 공식 출판을 목적으로 하지 않고, 대중적 공유를 위해 공개된 경우도 존재하기 때문이다(Abdill & Blekhman, 2019; Bertin & Atanassova, 2022; Larivière et al., 2014; Sevryugina & Dicks, 2021).

공식 출판본이 존재하지 않는 프리프린트의 알트메트릭스 지표를 분석한 결과는 다음 〈표 4〉와 같다. 뉴스 0.12회, 블로그 0.20회, X 9.72회, Mendeley 북마크 3.60회로 나타나, 모든 항목에서 앞서 제시한 전체 데이터셋 평균치보다 낮은 수준을 보였다. 이는 공식 출판본이 없는 프리프린트가 사회적으로 확산되는 과정에서 대중 매체를 통한 노출 빈도가 상대적으로 제한적임을 시사할 수 있다. 한편 이러한 결과는 언론이나 학술 블로그가 동료심사를 거치지 않은 연구 결과를 상대적으로 신중하게 다루기 때문일 가능성이 있다. 그러나 본 연구에서는 뉴스 및 블로그 게시 시점과 기사 작성 맥락을 직접 분석하지 않았으므로, 해당 결과는 프리프린트의 주제 특성, 표본 규모의 제한, 데이터 수집 시점의 영향 등에 의해 설명될 가능성도 있다. 따라서 이에 대한 해석에는 신중함이 요구된다.

〈표 4〉 공식 출판본이 존재하지 않는 프리프린트의 알트메트릭스

알트메트릭스 소스	N	최소값	최대값	평균	표준편차
Altmetric지수	221	0	194	8.31	21.949
뉴스	221	0	12	.12	.990
블로그	221	0	3	.20	.579
x	221	0	284	9.72	30.238
Mendeley	221	0	76	3.60	8.420

둘째, 앞에서 살펴본 바와 같이, 미검증 프리프린트는 전반적으로 낮은 사회적 언급 수치를 보였으나, 일부는 예외적으로 높은 AAS를 기록하였다. 앞에서 언급한 대면 학술회의의 지식 확산 효과를 다룬 프리프린트 이외에도 개인의 전자기록 관리 실태, 미·중 연구 협력 감소 현상을 분석한 프리프린트가 언론 보도와 X에 언급되면서 주목을 끌었다. 더불어 지식그래프인 OpenAlex를 소개한 논문도 Mendeley에서 다수 북마크되면서 큰 관심을 끌었다.

셋째, 미검증 프리프린트를 대상으로 BERTopic 기반 토픽 모델링을 수행한 결과를 〈표 5〉와 같이 살펴보았다. 분석 결과, 인용 분석과 학술 영향력(44)에 해당하는 토픽이 가장 큰 비중을 차지하였으며, 인공지능·머신러닝 연구의 학술적 영향 분석(37)과 관련된 토픽도 큰 비중을 차

지하였다. 예를 들어 인용 분석과 학술 영향력 토픽에는 「The effect of the rapid growth of COVID-19 publications on citation indicators」와 같이 인용지표와 학술영향력을 분석하는 연구가 포함되었으며, 인공지능・머신러닝 연구의 학술적 영향 분석 토픽에는 「Scientific Impact of Graph-Based Approaches in Deep Learning Studies」와 같은 연구가 대표 문헌으로 확인되었다. 그밖에도 오픈사이언스(35), 도서관 연구 인프라(27)와 관련된 토픽이 도출되었다. 토픽의 구조는 전반적으로 앞의 〈표 2〉에서 분석한 결과와 유사한 양상을 보였으나, 세부적으로는 기술 기반 연구의 응집성이 보다 뚜렷하게 나타나는 특징을 보였다. 특히 인공지능・머신러닝 관련 연구는 미검증 프리프린트 집합에서 계량정보학과 동시에 주요 축을 형성하며 독립 토픽으로 나타났다는 점이 주목된다. 또한 연구데이터 및 오픈데이터 관리, 학술 인프라와 관련된 주제가 분화된 형태로 나타나, 데이터・계량정보・AI 기반 연구가 주도하는 지식 생태계 내에 더욱 가까이 있음을 시사할 수 있겠다.

〈표 5〉 공식 출판본이 존재하지 않는 프리프린트 BERTopic 기반 모델링 결과

토픽	N	주요 키워드	토픽	N	주요 키워드
1. 인용 분석과 학술 영향력 연구	44	citation, scientific, bibliometric, papers, academic	5. 디지털 미디어와 정보사회 연구	24	digital media, social media, government, smart
2. 인공지능・머신러닝 연구의 학술적 영향 분석	37	machine learning, impact, peer review	6. 연구데이터 및 학술 메타데이터 분석 연구	22	research data, community, science
3. 오픈사이언스와 학술 커뮤니케이션 연구	35	open science, scientific, publications, author	7. 오픈데이터와 연구데이터 관리 체계 연구	19	open data, blockchain, research, supporting, tool
4. 도서관 및 연구 인프라와 재현성 연구	27	library, source, python, empirical study	8. 연구정보관리 및 학술 인프라 운영 연구	13	research information, metadata, journals, management

넷째, 알트메트릭스 간 상관관계 양상은 〈표 6〉과 같이 전체 데이터셋과 다소 상이한 모습을 보였다. 특히 전체 데이터셋을 대상로 한 1단계 분석에서 비교적 높은 상관성을 보인 뉴스와 X간의 관련성이 여기에서는 유의하지 않은 것으로 나타났으며($r = 0.07$, $p > 0.05$), 뉴스와 블로그 간의 관련성도 여기에서는 유의미하게 낮아진 것으로 나타났다($r = 0.217$, $p < 0.01$). 이는 피어리뷰를 거치지 않은 미검증 상태의 연구 결과에 대해 언론과 학술 블로그가 언급에 신중을 기하고 있음을 나타낸다. 다시 말해 검증되지 않은 프리프린트가 특정 매체를 통해 단독으로 소개될 수는 있겠으나, 다중 매체에 의해 연쇄적으로 보도됨으로써 사회적으로 파급되는 현상은 제한적이라고 말할 수 있을 것이다. 한편, X와 Mendeley 북마크 간의 상관관계는 공식 출판본이 존재하는 프리프린트에서 관찰된 수준과 유사한($r = 0.5$, $p < 0.01$) 정도를 보였는데, 이는 연구자들의 관심을 끄는 연구 결과는 검증 여부와 무관하게 소셜 미디어를 통해 공유되고, 학술적, 실무적 목적을 위해 북마크되고 있음을 나타낼 수 있을 것이다.

〈표 6〉 공식 출판본이 존재하지 않는 프리프린트의 알트메트릭스 소스간 관련성

		뉴스	블로그	x	Mendeley
뉴스	상관계수	1.000	.217**	.076	-.052
	유의확률 (양측)	.	.001	.261	.438
	N	221	221	221	221
블로그	상관계수	.217**	1.000	.287**	.192**
	유의확률 (양측)	.001	.	〈.001	.004
	N	221	221	221	221
x	상관계수	.076	.287**	1.000	.500**
	유의확률 (양측)	.261	〈.001	.	〈.001
	N	221	221	221	221
Mendeley	상관계수	-.052	.192**	.500**	1.000
	유의확률 (양측)	.438	.004	〈.001	.
	N	221	221	221	221

**p 〈 .01

3. 공식 출판본이 존재하지 않지만 인용된 프리프린트 분석

마지막으로, 공식 출판본 없이 프리프린트 상태로 피인용된 사례에 대해 살펴보도록 한다. COVID-19 관련 프리프린트는 다른 주제보다 최소 15배 이상 더 많이 인용되고 확산되는 경향이 있는데(Fraser et al., 2021), 이는 사회적으로 주목받는 주제를 다룬 논문일수록 프리프린트 상태에서도 상당한 수준의 인용이 발생할 수 있음을 시사한다. 이러한 경향이 문헌정보학 분야에서도 관찰되는지 확인하기 위하여 프리프린트의 피인용 현황을 분석해 보았다. 분석 결과 〈표 7〉과 같이 총 221건의 논문 중 피인용이 발생한 경우는 28건(12.7%)이었으며, 나머지 193건(87.3%)은 인용되지 않은 것으로 확인되었다.

〈표 7〉 공식 출판본이 존재하지 않은 프리프린트의 피인용 여부와 세부 인용 횟수분포

피인용 여부	빈도	퍼센트	세부 피인용 횟수(빈도)
없음	193	87.3	
있음	28	12.7	1회(20), 2회(5), 3회(1), 4회(1), 29회(1)
전체	221	100.0	

먼저, 피인용된 논문은 정보 및 컴퓨터과학, 데이터관리 등 학술커뮤니케이션과 오픈사이언스 인프라 전반에 걸친 다학제 영역에 분포한다. 조금 더 구체적으로 살펴보면, 학술 생태계와 연구협력, 문제 논문 자동 탐지 및 출판 윤리, AI 기반 연구 평가 등과 관련된 논문이 피인용되었다. 가장 높은 피인용 횟수(29회)를 기록한 논문은 개방형 과학 지식 그래프(Open Scientific Knowledge Graph)인 OpenAlex에 대한 프리프린트로 이는 학술지 출판을 목적으로 하지 않는 기술보고서 성격의 문서로 보인다. 이 논문은 정보과학, 컴퓨터과학, 문헌정보학, 건강과학, 인간사회 연구,

데이터 관리 등 다양한 학문 분야에서 인용되었으며, 구체적으로는 SN Computer Science(SJR 0.6), Scientometrics(SJR 1.02), Patterns(SJR 1.34) 등의 저널에서 인용된 것이 확인되었다. 한편, 4회 인용된 프리프린트는 화상회의 피로 요인을 다룬 연구로 해당 논문은 정보 및 컴퓨터 과학, 인간 중심 컴퓨팅, 경영 및 상업, 관광・서비스 산업 등에서 인용되었다. 인용 저널에는 Sustainability(SJR 0.664) 등이 포함된다. 또한 AI 기반의 연구자 및 논문 영향력 평가에 관한 프리프린트도 계량정보학 대표 저널인 Scientometrics(SJR 1.02), Quantitative Science Studies (SJR 1.62) 등을 통해 3회 인용된 것으로 확인되었다. 그 밖에 출판 윤리, 논문 품질 및 부정행위 탐지, 피어 리뷰 등 AI를 활용한 연구 및 출판 과정 자동화 등의 주제가 인용된 것으로 확인되었다.

둘째, 프리프린트가 인용된 저널을 분석한 결과, 총 45개의 상이한 저널에서 인용이 이루어졌으며, 문헌정보학 분야에서는 Scientometrics가 대표적 인용 저널로 확인되었다. 그 이외에도 〈표 8〉과 같이 컴퓨터공학, HCI, 데이터사이언스 분야의 저널에서 미검증 프리프린트가 인용된 것으로 확인되었다. 의학 분야에서는 프리프린트 간 상호 인용이 빈번히 관찰되었으며, 이는 COVID-19 시기 연구 성과의 신속한 확산 필요성에서 비롯된 특수한 현상으로 논평된 바 있다(조재인, 2025). 그러나 문헌정보학 분야에서는 이와 같은 프리프린트 간 인용 현상은 뚜렷하게 나타나지 않았다. 종합적으로 보면, 도서관 실무보다는 정보 및 데이터 관련 분야의 저널에서 미검증 프리프린트가 피인용되는 경향이 나타났고, 이는 실증적, 계량적 접근이 이루어지는 연구 분야에서 프리프린트의 활용도가 높게 나타날 수 있다는 점을 시사하고 있다.

〈표 8〉 공식 출판본이 없는 프리프린트를 2회 이상 인용한 저널

저널	인용횟수	SJR
Open Research Europe	2	0
PeerJ Computer Science	2	0.638
Proceedings of the 5th International Conference on Conversational User Interfaces	2	-
Royal Society Open Science	2	0.84
Scientometrics	2	1.02
Swarm and Evolutionary Computation	2	2.65

셋째, 고영향 저널에서도 미검증 프리프린트가 인용되고 있는지 확인한 결과, 〈표 9〉와 같이 SJR 1 이상인 다수의 저널이 프리프린트 상태의 논문을 인용한 것으로 확인되었다. 참고로 SJR 지수는 전 세계 학술지의 평균값을 1로 설정하고 있어, SJR ≥ 1은 평균 이상의 영향력을 가진 저널, SJR 〈 1은 평균 이하의 저널로 해석할 수 있다. 〈표 9〉를 보면 정보학, 과학계량학, 컴퓨터 및 데이터 과학 등의 상위 저널들이 포함되어 있는데, 이는 데이터와 기술 중심의 정량 분석을 중시하는 우수 저널에서도 동료평가를 거치지 않은 미검증 프리프린트를 인용하고 있음을 의미한다. 논문의 주제를 자세히 살펴보면, AI 기반 연구 평가, 출판 윤리와 같이 엄격한 검증이 요구되

는 민감한 주제가 포함되어 있는 것으로 나타났다. 이는 정책적, 제도적 함의가 커 엄격한 평가가 요구되지만, 미검증 상태에서 인용되고 있는 것이다. 이렇게 고영향 저널에서 미검증 프리프린트를 인용하고 있다는 것은 신속한 정보 공유가 필요한 주제를 다룬 프리프린트의 학술적 활용이 확대되고 있음을 보여준다. 그러나 동시에 아직 충분한 동료 검토를 거치지 않은 결과가 학술 담론에 조기에 편입될 수 있다는 우려도 내포하고 있다.

〈표 9〉 공식 출판본이 없는 프리프린트를 2회 이상 인용한 우수 저널

저널	SJR
Swarm and Evolutionary Computation	2.65
British Journal of Educational Technology	2.12
Higher Education	1.95
Quantitative Science Studies	1.61
Patterns	1.34
The CRISPR Journal	1.15
Journal of Construction Engineering and Management	1.15
Economic Analysis and Policy	1.1
EPJ Data Science	1.07
Scientometrics	1.02
Journal of the Association for Information Science and Technology	1.01
Frontiers in Environmental Science	1

넷째, 미검증 프리프린트 중 피인용된 논문이 그렇지 않은 논문에 비해 더 많은 사회적 주목을 받았는지 검증하기 위해 비모수 분석을 실시하였다. 집단 간 차이 분석에 앞서 Shapiro-Wilk 검정을 실시한 결과, Altmetric Attention Score 및 세부 지표들의 분포가 정규성을 충족하지 않는 것으로 나타났다($p<.001$). 이에 따라 출판 여부에 따른 차이 분석에는 비모수 검정인 Mann-Whitney U 검정을 적용하였다. 그 결과, 〈표 10-11〉과 같이 AAS는 피인용이 이루어진 논문 군에서 유의하게 높은 값을 나타냈으며($p < 0.01$), 세부 지표 중에서는 특히 X 언급 수가 통계적으로 유의미한 차이를 보였다($p < 0.01$). 이러한 결과는 인용이 이루어진 프리프린트는 더 큰 사회적 파급력을 보이는 경향이 있음을 시사한다. 비록 검증 이전의 프리프린트라도 후속 연구에 의해 인용되었다면 소셜 미디어에서의 노출을 유도해 사회적 확산 수준 역시 높아질 수 있다는 것이다. 따라서 프리프린트 인용은 단순한 정보 활용을 넘어 검증 이전 지식의 정당화와 확산을 매개할 수 있다는 점에서 신중한 접근이 요구된다. 따라서 미검증 프리프린트가 인용된 논문이 학술지에 투고된다면, 저널 측은 혹시 잘못된 정보가 학문적 권위를 통해 다양한 매체로 사회에 널리 확산되는 것은 아닐지 주의를 기울일 필요가 있을 것이다. 다만 위 분석에서 피인용이 발생한 프리프린트는 28편에 불과하므로 본 연구의 결과를 일반화하는 데에는 신중한 해석이 필요함을 언급한다.

〈표 10〉 프리프린트의 피인용 여부에 따른 알트메트릭스 현황

피인용 여부		Altmetric지수	뉴스	블로그	x	Mendeley
없음	평균	7.23	3.17	1.42	10.98	7.21
	N	193	6	24	152	72
있음	평균	15.79	8.00	1.83	19.16	13.85
	N	28	1	6	25	20
전체	평균	8.31	3.86	1.50	12.14	8.65
	N	221	7	30	177	92

〈표 11〉 프리프린트의 피인용 여부에 따른 알트메트릭스 차이에 대한 비모수 분석 결과

항목	피인용 여부	N	평균 순위	순위합	Mann-Whitney U	p
Altmetric 지수	없음	193	104.32	20133.5	1412.5	〈.001
	있음	28	157.05	4397.5		
뉴스	없음	6	3.67	22	1	0.295
	있음	1	6	6		
블로그	없음	24	14.79	355	55	0.303
	있음	6	18.33	110		
X	없음	152	83.21	12647.5	1019.5	〈.001
	있음	25	124.22	3105.5		
Mendeley	없음	72	44.39	3196	568	0.148
	있음	20	54.1	1082		

Ⅴ. 요약 및 논의

일부 학문 분야에서는 프리프린트가 적극적으로 공유되고 있지만, 문헌정보학 분야는 아직까지 프리프린트를 공유하는 관행이 확립되어 있지는 않다. 그러나 프리프린트가 빠르게 증가하고 있는 환경 속에 점차 프리프린트 기탁에 관심이 나타나고 있다(Wang et al. 2018). 본 연구는 이러한 측면에서 문헌정보학 분야의 프리프린트를 분석하였으며, 그 중 1년반 이상이 경과되었음에도 불구하고 출판이 이루어지지 않은 미검증 프리프린트 상태의 논문에 주목하여 분석을 수행하였다.

첫 번째, BERTopic 모델링 결과 arXiv에 등록된 문헌정보학 분야 프리프린트의 주제는 전통적인 도서관 실무 분야보다는 정보학, 특히 계량정보학 및 오픈사이언스와 관련된 주제가 높은 비중을 차지하는 것으로 나타났다. 또한 프리프린트를 인용한 저널 역시 순수 문헌정보학의 실무 및 서비스를 다루는 저널보다는 정보학 및 학제적 연구 성격을 가진 저널에 상대적으로 많이 분포하는 것으로 확인되었다. 이러한 결과는 arXiv에 등록된 문헌정보학 분야 프리프린트가 학술 커뮤니케이션, 연구성과 측정, 데이터 공유 및 활용과 같은 정보학 중심 주제와 밀접하게 연관되어 있음을

시사한다. 다만 본 연구는 arXiv에 등록된 문헌정보학 분야 프리프린트를 대상으로만 수행되었으므로, 이러한 결과를 문헌정보학 분야 전체 프리프린트의 활용 행태로 일반화하는 데에는 주의가 필요할 것이다.

두 번째, arXiv에 아카이빙된 프리프린트의 언론보도와 소셜 미디어 노출 정도는 높지 않은 것으로 확인되었다. 더불어 공식 출판본이 존재하지 않는 미검증 프리프린트는 알트메트릭스 소스별 언급 횟수가 더 낮으며, 다양한 매체에서 다발적으로 노출되면서 사회적으로 강하게 파급되지는 않는 것으로 나타났다. 그러나 공식 출판본 존재 여부와 무관하게 프리프린트는 학술 커뮤니티에서는 조기에 확산되어(Shuai et al., 2012), 활용되고 있는 것으로 분석되었다.

세 번째, 미검증 논문 중에도 영향력 있는 저널에서 인용되고 그로 인해 사회적 주목으로 이어지는 경우가 존재하였다. 특히 정책적, 제도적 함의가 큰 주제의 연구는 검증 이전 단계에서 인용될 경우, 학술적 권위를 매개로 학계에 빠르게 파급될 가능성이 있다. 따라서 학술지 측은 미검증 프리프린트가 인용된 논문이 투고되었을 때, 이를 통해 불완전한 정보의 확산이 가속화되는 것은 아닌지 신중한 판단이 필요할 것이다. 또한 온라인에서 먼저 화제가 된 논문이 정식 출판을 위해 학술지에 제출되었을때에도 마찬가지이다. 저널 측은 해당 연구의 학술적 타당성과 기여도뿐 아니라 이미 형성된 사회적 영향과 정보 확산의 적절성까지 함께 검토할 필요가 있을 것이다. 프리프린트는 공개 직후 과도한 관심을 받을 경우 정보 과부하와 무질서가 발생하며, 이 과정에서 부정확하거나 미숙한 보도가 확산되어 왜곡된 정보가 마치 진실인 것처럼 인식될 위험성이 있기 때문이다. 즉 검증되지 않은 상태에서 형성된 사회적 주목이나 오해가 학술 공동체의 신뢰 기반 속에서 강화되지 않도록, 제출된 원고의 내용과 그 주변의 정보 환경을 종합적으로 고려한 신중한 판단이 요구될 것이다.

Ⅵ. 결 론

본 연구는 arXiv에 기탁된 문헌정보학 분야의 프리프린트 중 아카이빙 후 충분한 시간이 경과되었지만 출판에 이르지 못한 프리프린트에 초점을 맞춰 피인용과 알트메트릭스를 분석하였다. 그 결과 미검증 프리프린트가 사회 전반에 활발히 확산되지는 않지만, 연구자들에게는 조기에 공유되어 활용될 수 있으며, 영향력 있는 학술지에 인용되고 있는 경우도 존재하는 것으로 확인되었다. 또한 영향력 있는 학술지에 인용된 프리프린트는 잠재적 한계를 지닐 수 있음에도 학술적 권위를 통해 빠르게 확산될 가능성이 있는 것으로 확인되었다.

프리프린트는 공식 출판 과정을 우회하여 연구 결과를 빠르게 배포할 수 있지만, 아이디어 스쿠핑, 내용과 증거의 변경에 따른 연구의 일관성과 투명성 훼손, 그리고 이중 맹검 시스템을 약화시킬

수 있는 문제점도 가지고 있다(Oh & Kang, 2023). 더불어 프리프린트가 언론이나 소셜 미디어를 통해 조기에 사회적으로 확산될 경우, 동료심사를 거치지 않은 연구 결과가 학술적 검증 이전에 널리 공유될 수 있으므로 이에 대한 신중한 해석이 필요하다. 따라서 전통적인 학술 출판 시스템과 프리프린트가 공존하는 현 시점에서, 저널측은 이미 사회에 확산된 연구에 대하여 어떠한 기준과 절차로 학문적 타당성을 판단하고 학술 담론으로 수용할지에 대한 원칙을 정립할 필요가 있을 것이다. 또한 학문적 무결성을 유지하기 위하여 출판 윤리와 편집 정책을 재정비하고 프리프린트와 공식 출판물 간의 관계를 명확히 규정하는 노력도 병행해야 할 것이다.

본 연구는 기탁 후 일정 기간이 경과하였음에도 공식 출판되지 않은 프리프린트에 초점을 맞추어 분석을 수행하였다. 그러나 향후 해당 논문이 정식 출판될 경우 피인용 및 알트메트릭스 지표가 변동할 수 있으므로, 본 연구의 결과 역시 달라질 가능성이 있다. 또한 피인용이 발생한 미출판 프리프린트의 수가 제한적이므로 관련 결과를 일반화하는 데에는 신중한 접근이 필요하다. 한편 본 연구의 토픽 모델링은 프리프린트의 전반적인 주제 경향을 탐색적으로 파악하기 위해 논문 제목을 활용하였으나, 향후에는 초록을 포함한 보다 풍부한 텍스트를 활용할 필요가 있으며, 아울러 SocArXiv 등 다양한 프리프린트 플랫폼을 포함한 비교 연구를 수행함으로써 연구 결과의 일반화 가능성을 보다 폭넓게 검토할 필요가 있을 것이다.

참 고 문 헌

조재인 (2025). 프리프린트의 피인용과 사회적 주목에 관한 연구: medRxiv를 중심으로. 한국문헌정보학회지, 59(2), 25-42. https://doi.org/10.4275/KSLIS.2025.59.2.025

Abdill, R. J. & Blekhman, R. (2019). Tracking the popularity and outcomes of all bioRxiv preprints. eLife, 8, e45133. https://doi.org/10.7554/eLife.45133

Altmetric (2024). Altmetric database in CSV format. Unpublished raw data.

Bertin, M. & Atanassova, I. (2022). Preprint citation practice in PLOS. Scientometrics, 127(11), 6895-6912. https://doi.org/10.1007/s11192-022-04388-5

Björk, B.-C. & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. Journal of Informetrics, 7(4), 914-923. https://doi.org/10.1016/j.joi.2013.09.001

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. Journal of the American Society for Information Science and Technology, 57(8), 1060-1072. https://doi.org/10.1002/asi.20373

Brown, C. (2003). The role of electronic preprints in chemical communication: Analysis

of citation, usage, and acceptance in the journal literature. Journal of the American Society for Information Science and Technology, 54(5), 362-371. https://doi.org/10.1002/asi.10223

Chaleplioglou, A. & Koulouris, A. (2023). Preprint paper platforms in the academic scholarly communication environment. Journal of Librarianship and Information Science, 55(1), 43-56. https://doi.org/10.1177/09610006211058908

Chingath, V. (2024). Library and Information Science (LIS) preprints and their social media presence: An altmetric study based on arXiv preprint repository. Online Conference on Libraries, Education and AI: Bridging the Gap between Knowledge and Technology. Available: https://www.researchgate.net/publication/389447172

da Silva, J. A. T. (2017). Preprints: Ethical hazard or academic liberation? KOME- an International Journal of Pure Communication Inquiry, 5(2), 73-80. https://doi.org/10.17646/KOME.2017.26

Davis, P. M. & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? Scientometrics, 71(2), 203-215. https://doi.org/10.1007/s11192-007-1661-8

Dimensions (2019). Are preprints included in Dimensions? Available: https://help.dimensions.ai/en/articles/9783937

Dimensions (2022). Which research categories and classification schemes are available in Dimensions? Available: https://help.dimensions.ai/en/articles/9781825

Ebrahimy, S., Mehrad, J., Setareh, F., & Hosseinchari, M. (2016). Path analysis of the relationship between visibility and citation: The mediating roles of save, discussion, and recommendation metrics. Scientometrics, 109(3), 1497-1510. https://doi.org/10.1007/s11192-016-2130-z

Feldman, S., Lo, K., & Ammar, W. (2018). Citation count analysis for papers with preprints. arXiv preprint, https://arxiv.org/abs/1805.05238

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A.(2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLoS Biology, 19(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. Quantitative Science Studies, 1(2), 618-638.

https://doi.org/10.1162/qss_a_00043

Fraumann, G. & Colavizza, G. (2022). The role of blogs and news sites in science communication during the COVID-19 pandemic. Frontiers in Research Metrics and Analytics, 7, 824538. https://doi.org/10.3389/frma.2022.824538

Fu, D. Y. & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. eLife, 8, e52646. https://doi.org/10.7554/eLife.52646

Henneken, E. A., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Thompson, D., Bohlen, E., Murray, S. S., Ginsparg, P., & Warner, S. (2007). E-prints and journal articles in astronomy: A productive co-existence. Learned Publishing, 20(1), 16-22. https://doi.org/10.1087/095315107779490661

Hoy, M. B. (2020). Rise of the Rxivs: How preprint servers are changing the publishing process. Medical Reference Services Quarterly, 39(1), 84-89. https://doi.org/10.1080/02763869.2020.1704597

Icahn School of Medicine at Mount Sinai (n.d.). Preprints: The Basics: Scooping and Other Common Misconceptions. Available: https://libguides.mssm.edu/preprints/scooping (Accessed: March 31, 2025)

Ingelfinger, F. J. (1977). The general medical journal: For readers or repositories? New England Journal of Medicine, 296(22), 1258-1264.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The effect of use and access on citations. Information Processing & Management, 41(6), 1395-1402. https://doi.org/10.1016/j.ipm.2005.03.010

Larivière, V., Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv e-prints and the journal of record: An analysis of roles and relationships. Journal of the Association for Information Science and Technology, 65(6), 1157-1169. https://doi.org/10.1002/asi.23044

Li, X., Thelwall, M., & Kousha, K. (2015). The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication. Aslib Journal of Information Management, 67(6), 614-635. https://doi.org/10.1108/AJIM-03-2015-0049

Lin, J., Yu, Y., Zhou, Y., Zhou, Z., & Shi, X. (2020). How many preprints have actually been printed and why: A case study of computer science preprints on arXiv. Scientometrics, 124(1), 555-574. https://doi.org/10.1007/s11192-020-03430-8

Moed, H. F. (2007). The effect of "open access" on citation impact: An analysis of arXiv's condensed matter section. Journal of the American Society for Information Science and Technology, 58(13), 2047-2054. https://doi.org/10.1002/asi.20663

Oh, H.-C. & Kang, H. (2023). Current concerns on journal article with preprint: Korean Journal of Internal Medicine perspectives. The Korean Journal of Internal Medicine, 38(3), 332-337. https://doi.org/10.3904/kjim.2023.099

Okamura, K. (2022). Scientometric engineering: Exploring citation dynamics via arXiv eprints. Quantitative Science Studies, 3(1), 122-146. https://doi.org/10.1162/qss_a_00174

Penfold, N. C. & Polka, J. K. (2020). Technical and social issues influencing the adoption of preprints in the life sciences. PLoS Genetics, 16(4), e1008565. https://doi.org/10.1371/journal.pgen.1008565

Rodriguez, M. A., Bollen, J., & Van de Sompel, H. (2006). The convergence of digital libraries and the peer-review process. Journal of Information Science, 32(2), 149-159. https://doi.org/10.1177/0165551506062327

Serghiou, S. & Ioannidis, J. P. A. (2018). Altmetric scores, citations, and publication of studies posted as preprints. JAMA, 319(4), 402-404. https://doi.org/10.1001/jama.2017.21168

Sevryugina, Y. V. & Dicks, A. J. (2021). Publication practices during the COVID-19 pandemic: Biomedical preprints and peer-reviewed literature. bioRxiv. https://doi.org/10.1101/2021.01.21.427563

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. PLoS ONE, 7(11), e47523. https://doi.org/10.1371/journal.pone.0047523

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics, 126(6), 5113-5142. https://doi.org/10.1007/s11192-021-03948-5

Thelwall, M. & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. Journal of the Association for Information Science and Technology, 67(8), 1962-1972. https://doi.org/10.1002/asi.23501

Wang, L. & Zhan, Y. (2019). A conceptual peer review model for arXiv and other preprint databases. Learned Publishing, 32(3), 213-219. https://doi.org/10.1002/leap.1229

Wang, Z., Glänzel, W., & Chen, Y. (2018). How self-archiving influences the citation impact of a paper: A bibliometric analysis of arXiv papers and non-arXiv papers in the field of information science and library science. In Proceedings of the 23rd International Conference on Science and Technology Indicators (STI2018), 323-330.

Wang, Z., Glänzel, W., & Chen, Y. (2020). The impact of preprints in Library and Information Science: An analysis of citations, usage and social attention indicators. Scientometrics, 125(2), 1403-1423. https://doi.org/10.1007/s11192-020-03612-4

Yi, H. & Huh, S. (2021). Korean editors' and researchers' experiences with preprints and attitudes towards preprint policies. Science Editing, 8(1), 4-9. https://doi.org/10.6087/kcse.223

• 국한문 참고문헌의 영문 표기
(English translation / Romanization of references originally written in Korean)

Cho, Jane (2025). A study on citations and social attention of preprints: Focusing on medRxiv. Journal of the Korean Society for Library and Information Science, 59(2), 25-42. https://doi.org/10.4275/KSLIS.2025.59.2.025