JDFR Journal of Data Forensics Research
데이터포렌식연구, 2권 1호, 2025년 6월, 19-36면
Journal of Data Forensics Research, Volume 2, Issue 1, June 2025, Pages 19-36.
https://doi.org/10.12972/JDFR.2025.2.1.2
Print ISSN 3058-9460
Online ISSN 3058-9479

원저

불법·유해 사이트 연결 실태 분석: 사회 연결망 분석(SNA)을 중심으로

이석규[1], 김태연[1], 윤지원[1], 김경종[2]

[1]경찰대학 학부생, [2]경찰대학 경찰학과 교수

교신저자: 김경종, leeyeongul@police.go.kr

요약

지능적으로 연결된 불법·유해 사이트 네트워크는 기존의 개별 사이트 차단 방식만으로는 대응에 한계를 드러내고 있다. 본 연구는 이러한 복합적인 국내 불법·유해 사이트 연결망 전체를 대규모 데이터 기반의 사회 연결망 분석(SNA)을 통해 분석함으로써, 선행 연구들이 주로 특정 유형에 국한되거나 정성적 분석에 머물렀던 한계를 극복하고자 하였다. '주소월드'를 시작점으로 웹 크롤링을 통해 약 4,000개의 연결 데이터를 수집하였으며, 이를 바탕으로 Leiden 알고리즘과 다양한 중심성 지표를 활용하여 네트워크 내 불법 도박 사이트 등의 핵심 허브, 주소 모음 사이트 및 토렌트 사이트 등의 브릿지 노드, 그리고 공동 운영 가능성이 있는 사이트 군집을 식별하였다. 본 연구는 이러한 구조적 취약점과 핵심 노드에 대한 분석 결과를 통해 개별 사이트 차단을 넘어 '풍선효과'를 최소화하고, 네트워크 전체를 효과적으로 무력화할 수 있는 데이터 기반의 표적화된 차단 전략 수립에 중요한 과학적 근거와 실질적 시사점을 제공한다는 데 의의가 있다.

주제어

불법·유해 사이트, 웹 크롤링, 사회 연결망 분석

Received: June 11, 2025
Revised: June 30, 2025
Accepted: June 30, 2025
Published: June 30, 2025

Original Article

Analysis of the connection conditions of illegal and harmful websites

Seok Kyu Lee[1], Taeyeon Kim[1], Jiwon Yoon[1], Kyungjong Kim[2]

[1]Undergraduate Student, Korean National Police University, Republic of Korea
[2]Professor, Department of Police Science, Korean National Police University, Republic of Korea

Corresponding Author: Kyungjong Kim, leeyeongul@police.go.kr

ABSTRACT

The intelligently interconnected networks of illegal and harmful websites reveal the limitations of conventional individual site-blocking methods. This study addresses these limitations by analyzing the broader, complex web of domestic illegal and harmful sites using large-scale data-driven social network analysis (SNA), aiming to overcome the constraints of prior research, which has often focused on specific site types or qualitative approaches. Based on approximately 4,000 inter-site connections crawled starting from 'Jusoworld,' we employed the Leiden algorithm along with various centrality metrics to identify key network hubs (e.g., illegal gambling sites), bridges (e.g., address collection, torrent sites), and potentially co-managed site clusters. The significance of this research lies in its provision of scientific evidence and practical insights for developing data-driven, targeted blocking strategies. By focusing on structural vulnerabilities and key nodes, these strategies can more effectively neutralize entire networks and minimize the "whack-a-mole effect," rather than merely targeting individual sites.

KEYWORDS

illegal and harmful sites, social network analysis (SNA), web crawling

I. 서론

정보통신기술의 급격한 발전은 인터넷을 통한 정보 접근성과 편의성을 극대화하였으나, 동시에 불법 유해 정보의 생성과 유통 또한 폭발적으로 증가시키는 결과를 초래하였다. 특히 다양한 형태의 불법·유해 사이트는 심각한 사회적 문제를 야기하며, 이로 인한 경제적 피해 또한 간과할 수 없는 수준에 이르고 있다. 특정 사이트를 차단하더라도 유사 사이트가 우후죽순처럼 생겨나는 이른바 '풍선효과(whack-a-mole effect)'가 빈번하게 발생하며, 사이트 운영자들은 도메인 변경, IP 주소 세탁, 해외 서버 이용 등 갈수록 지능화된 방식으로 단속을 회피하고 있다.

특히 주목할 점은 최근 불법·유해 사이트들이 단일 웹페이지 형태로 독립적으로 운영되기보다는, 배너 광고, 외부 링크 공유, 도메인 제휴 등 다양한 방식을 통해 서로 긴밀하게 연결된 네트워크 구조를 형성하며 진화하고 있다는 사실이다. 이러한 연결망은 불법 콘텐츠의 확산을 가속화하고 접근성을 높이는 통로로 기능할 뿐만 아니라, 단일 사이트 차단 시에도 네트워크 내 다른 경로를 통해 사용자를 유입시키는 등 단속 회피 전략의 핵심 요소로 작용한다. 실제로 유형이 서로 다른 불법·유해 사이트들조차 광고 수익 공유나 트래픽 교환을 목적으로 상호 연결되는 경우가 빈번하게 관찰되며, 이는 불법 콘텐츠 유통이 단순한 정보 제공을 넘어 사이트 간의 수익 창출 및 운영 협력을 중심으로 하는 복잡한 네트워크 기반 생태계로 발전하고 있음을 시사한다.

이처럼 고도화되고 네트워크화된 불법·유해 사이트의 특성을 고려할 때, 개별 사이트만을 대상으로 하는 기존의 단편적 대응 방식으로는 명확한 한계가 존재한다. 따라서 불법·유해 사이트 간의 연결 구조와 그 안에서 핵심적인 역할을 수행하는 행위자들을 식별하고, 이를 바탕으로 네트워크 전체에 효과적으로 영향을 미칠 수 있는 전략적 접근이 절실히 요구된다. 이는 기존의 점(点) 단위 대응에서 선(線)과 면(面)을 고려하는 네트워크 관점의 패러다임 전환을 의미하며, 보다 근본적이고 효율적인 대응 체계 구축의 필요성을 강조한다.

선행 연구에서도 이러한 네트워크 접근의 중요성은 일부 인식되어 왔다. 예를 들어, 송봉규(2020)는 불법·유해 사이트 간 배너광고 네트워크 분석을 통해 사이트 간 연결이 운영 및 수익 구조 차원에서 형성됨을 실증적으로 제시하였다[1]. 그는 광고 플랫폼이나 링크 경유지를 매개로 다양한 유형의 불법 사이트들이 상호 연결되어 있음을 확인하여 불법 도박 웹사이트 집중 단속, 불법 직업알선 웹사이트 차단 강화 등 불법 유해 웹사이트에 대한 경찰의 대응 방안을 모색하였다. 또한, 임경대와 이상진(2023)은 불법 저작물 스트리밍 사이트 간 연결성과 유사성 분석을 통해 동일 운영자 혹은 조직에 의한 다수 유사 사이트 생성 및 연결망을 통한 단속 우회 가능성을 지적하며, 네트워크 단위 접근의 필요성을 시사한 바 있다[2].

본 연구는 이러한 문제의식과 선행 연구의 논의를 바탕으로, 불법·유해 사이트 간의 외부 링크 및 배너광고 기반 연결 실태를 사회 연결망 분석(Social Network Analysis, SNA) 기법을 통해 실증적으로 분석하고, 이를 통해 네트워크 기반의 효과적인 차단 전략을 제시하는 데 주된 목적을 둔다. 특히 네트워크 내에서 사이트의 중요성을 평가하기 위해 연결 중심성, 매개 중심성, 페이지랭크 중심성 등의 중심성 지표를 사용한다. 이와 더불어 클러스터링을 통해 네트워크 내에서 긴밀히 연결된 노드들을 그룹화하고 시각화하여 네트워크의 구조적 특성을 파악하고자 한다.

특히 본 연구는 다음과 같은 두 가지 측면에서 학문적 기여를 지향한다. 첫째, 불법·유해 사이트 간의 외부 링크 연결 관계를 분석하여, 네트워크 내에서 중심성을 가지는 허브 사이트 또는 중개 노드를 식별하고, 이들을 중심으로 한 우선 차단 및 집중 대응 전략의 필요성을 제시한다. 둘째, 기존의 콘텐츠 유형이나 도메인 기반의 개별 차단 방식에서 벗어나, 사이트 간 연결 구조

전체를 고려한 구조 기반 대응 체계로의 정책적 전환 가능성을 탐색한다. 궁극적으로 본 연구는 불법·유해 사이트 대응이 보다 전략적이고 선제적인 방향으로 전환될 수 있는 이론적, 실무적 기반을 마련하는 데 기여하고자 한다.

II. 이론적 배경 및 선행연구

2.1. 불법·유해 사이트

2.1.1. 불법·유해 사이트의 개념 및 정의

송봉규(2020)는 불법·유해 사이트를 “법률상 금지된 콘텐츠(불법 도박, 음란물, 저작권 침해, 마약류 등)를 포함하거나, 청소년에게 유해한 정보를 제공하는 웹사이트”로 정의하면서, 이를 불법 도박과 같이 직접적으로 범죄가 이뤄지는 웹사이트, 범죄 정보를 제공하여 간접적으로 범죄를 조장, 유인, 방조하는 웹사이트, 불법 콘텐츠를 이용해 방문자를 유인하는 웹사이트 등으로 분류하였다[1].

방송통신심의위원회는「정보통신망 이용촉진 및 정보보호 등에 관한 법률」제44조의7(불법정보의 유통금지 등)과「방송통신위원회의 설치 및 운영에 관한 법률」등을 근거로, 청소년에게 유해하거나 법적으로 금지된 정보를 제공하는 웹사이트에 대해 심의를 거쳐 차단 조치를 시행하고 있다. 대표적인 차단 대상은 음란물, 불법 도박, 저작권 침해물, 마약류 거래 정보 등이며, 유형에 따라 청소년 유해매체물 또는 불법 정보로 구분되어 대응된다.

이처럼 불법·유해 사이트는 법률적 위반 요소와 사회적 유해성을 동시에 포함하는 복합적 개념으로, 이에 대한 효과적인 대응을 위해서는 단순한 URL 차단이나 콘텐츠 삭제와 같은 기술적 차단 조치만으로는 한계가 있다. 따라서 법률적 정의에 더해, 운영 방식, 수익 구조, 사이트 간 연결성 등 구조적 특성에 대한 종합적 이해와 함께, 기술적·정책적 차원의 다층적 대응 전략이 병행되어야 할 필요가 있다.

2.1.2. 불법·유해 사이트의 유형 및 특성

불법·유해 사이트는 다양한 형태로 운영되며, 주요 유형으로는 불법 도박, 불법 성인물(음란물), 저작권 침해 웹사이트(불법 웹툰, 영화, 드라마 등), 마약류 판매 및 정보 제공 사이트, 그리고 피싱이나 스미싱을 통해 금융 정보나 개인 정보를 탈취하려는 사기 사이트 등이 있다. 이러한 사이트들은 막대한 트래픽을 유발하기도 하며(예: 음란물 사이트), 사회적으로 심각한 해악을 초래한다. 예를 들어, 불법 도박 사이트는 재정적 파탄을, 불법 성인물 사이트는 성 인식 왜곡 및 착취물 유통의 문제를, 저작권 침해 사이트는 창작자의 권리 침해 및 관련 산업 위축을, 마약류 사이트는 국민 건강과 안전을 심각하게 위협하며, 피싱 사이트는 금융 자산 및 개인정보 탈취로 이어진다.

이들 사이트의 주된 수익 창출 방식은 배너 광고를 통한 상호 트래픽 유도 및 광고 수익 확보이다. 특히, 송봉규(2020)의 연구에 따르면 불법 도박 사이트는 자체적으로 배너 광고를 게재하지 않으면서도 다른 불법 사이트의 최대 광고주 역할을 하는 특성을 보인다[1]. 또한, 마약류나 해킹 도구 등 불법적인 상품이나 서비스를 직접 판매하고 결제를 유도하거나, 피싱 등을 통해 탈취한 개인정보를 이용하여 금전적 이득을 취하기도 한다.

2.1.3. 불법·유해 사이트의 연결 구조 및 은닉/회피 전략

불법·유해 사이트들은 단순한 하이퍼링크 연결을 넘어 다양한 방식으로 서로 연결되며, 차단 조치를 회피하기 위한 전략을 사용한다. 주요 연결 방식으로는 배너 광고를 통한 상호 연결이 있으며, 이는 불법 사이트 생태계의 중요한 특징 중 하나이다. 또한, 여러 불법 사이트 주소를 모아 제공하는 '주소 모음 사이트' 역시 이들의 연결성을 강화하는 역할을 한다. 일부 사이트들은 리디렉션 체인이나 단축 URL을 사용하여 실제 목적지를 숨기기도 한다.

차단 및 단속을 피하기 위한 은닉/회피 전략도 고도화되고 있다. IP 주소나 도메인 이름을 수시로 변경하는 것은 기본적인 회피 수단이며, 콘텐츠 전송 네트워크(CDN) 서비스를 악용하여 실제 서버 위치를 은폐하기도 한다[2]. 특히 다크웹과 같은 익명 네트워크(Tor 등)를 활용하여 사이트를 운영함으로써 추적을 어렵게 만들고 접속 차단을 회피하는 경우가 많다[3]. 이러한 다크웹 사이트들은 일반적인 검색 엔진으로는 접근이 어렵고, 주소 또한 자주 변경되어 단속의 어려움을 가중시킨다. 더불어, 웹사이트 운영자들은 CAPTCHA, IP 기반 접근 차단, 로그인 요구 등의 안티 크롤링(anti-crawling) 조치를 통해 자동화된 정보 수집 및 분석 시도를 방해한다[4].

2.2. 선행 연구 검토

2.2.1. 불법·유해 사이트 및 온라인 범죄 네트워크 연구 동향

불법·유해 사이트 및 온라인 범죄 네트워크에 대한 연구는 국내외적으로 꾸준히 진행되어 왔으며, 주로 이들 사이트의 운영 방식, 수익 구조, 연결 특성 및 차단 전략에 초점을 맞추고 있다.

국내 연구 중 송봉규(2020)는 불법·유해 웹사이트의 주요 수익원인 배너광고를 중심으로 연결 실태를 분석하였다. 해당 연구는 음란물, 불법 웹툰, 불법 영화/드라마, 불법 도박, 불법 직업 알선 사이트 등을 대상으로 조사하여, 트래픽이 많은 사이트가 다른 불법 사이트로 연결되는 허브 역할을 하며, 특히 불법 도박 사이트는 자체 광고는 없으나 다른 불법 사이트의 주요 광고주임을 밝혔다. 또한, 이러한 불법 사이트 운영의 배후 세력 존재 가능성을 언급하였다[1]. 임경대·이상진(2023)은 불법 저작물 스트리밍 사이트 간의 연관성 분석을 통해 해당 사이트를 판별하는 기법을 제시하고, 이를 기반으로 특정 유형의 불법 사이트 네트워크 특성을 분석하고자 하였다[2]. 이 외에도 장준영 외(2022)는 HTML 및 URL 특징을 이용한 유해사이트 수집 시스템을 제안하는 등 데이터 수집 단계에서의 기술적 연구도 이루어지고 있다[5].

온라인 불법 도박 네트워크와 관련해서는 사회 연결망 분석(SNA)을 활용하여 사이버 도박 범죄 조직의 구조를 분석하고 수사 단서를 도출하려는 연구가 진행되었다[6]. 이러한 연구는 계좌 거래내역, 통화내역 등을 기반으로 조직원 간의 관계를 파악하고 핵심 인물을 식별하는 데 SNA를 적용한다.

다크웹(Dark Web)과 관련된 연구는 국제적으로 활발히 진행되고 있으며, 주로 다크웹 마켓플레이스의 구조, 판매 상품, 이용자 특성 및 수사 기법 개발에 집중된다. Nunes 등(2016)은 다크넷 및 딥넷에서 사이버 위협 정보를 수집하고 분석하는 시스템을 개발하여, 악성코드나 제로데이 공격과 같은 위협 정보를 선제적으로 탐지하고자 하였다[3]. 이들은 웹 크롤러를 통해 데이터를 수집하고 기계학습을 이용해 관련 정보를 분류하며, 여러 플랫폼에 걸쳐 활동하는 사용자들을 식별하였다. Kühn 등(2024)은 다크웹의 수동 및 반자동 평가를 통해 CTI 관련 정보의 양과 접근성의 어려움(예: CAPTCHA, 사용자 상호작용 요구)을 지적하며, 다크웹이 특수한 정보원이 될 수 있음을 시사했다[4]. 이러한 다크웹 연구들은 데이터 수집의 어려움, 사이트의 휘

발성, 익명성으로 인한 분석의 한계 등을 공통적으로 지적한다.

피싱 사이트와 관련해서는 주로 악성 앱 유포 방식이나 메신저를 통한 피싱 범죄의 실태 분석이 이루어지고 있으나[7], 피싱 사이트 간의 명시적인 연결망 구조를 SNA로 분석한 연구는 상대적으로 부족한 실정이다.

이러한 선행연구들은 불법·유해 사이트 및 온라인 범죄 네트워크의 다양한 측면을 조명하지만, 특정 유형의 사이트에 집중하거나(예: 다크웹 마켓, 불법 도박), 특정 연결 방식(예: 배너 광고)에 한정되거나, 혹은 정성적 분석이나 기술 개발에 머무르는 경우가 많았다. 또한, 데이터 수집의 어려움과 분석 대상 네트워크의 방대함으로 인해 전체적인 연결 구조 파악 및 핵심 노드 식별에는 한계를 존재한다.

2.2.2. SNA를 활용한 유사 네트워크 분석 연구

사회 연결망 분석(SNA)은 다양한 분야에서 복잡한 관계 구조를 파악하고 핵심 행위자를 식별하는 데 유용하게 활용되어 왔다. 특히 범죄 및 테러 네트워크, 정보 확산 패턴 분석 등에서 그 적용 가능성과 유용성이 입증되었다.

범죄 조직 네트워크 분석에서 SNA는 조직원 간의 통화 기록, 자금 거래 내역, 공범 관계 등을 기반으로 네트워크를 구축하고, 중심성 분석을 통해 조직 내 핵심 인물이나 하위 그룹(sub-group)을 식별하는 데 활용된다. 예를 들어, Pete 등(2020)은 6개의 다크웹 포럼을 대상으로 SNA를 적용하여 비교 분석하였으며[8], 또 다른 연구에서는 페이스북 내 온라인 활동가 그룹의 핵심 인물을 식별하기 위해 연결 중심성과 매개 중심성 등의 SNA 지표를 활용하여 네트워크 교란 전략 수립에 시사점을 제공했다[9]. 국내에서도 김광진 등(2024)은 COVID-19 역학조사 사례에서 SNA를 활용하여 확진자 간 접촉 네트워크를 분석하고 전파 경로를 예측하였으며, 이러한 분석 기법이 보이스피싱, 마약 거래, 조직폭력 수사 등에도 효과적으로 적용될 수 있음을 제언하였다[10].

테러 네트워크 분석에서도 SNA는 핵심적인 방법론으로 사용된다. Bermingham, Conway 등(2009)은 유튜브 내 특정 그룹을 대상으로 SNA와 감성 분석을 결합하여 온라인 급진화 가능성을 탐색하고, 네트워크 내 영향력 있는 인물을 식별하고자 하였다[11]. 또한, 미국 국가안보국(NSA)은 테러리스트 조직을 파악하기 위해 통화 기록 등의 메타데이터에 SNA를 적용하는 것으로 알려져 있다. Conway(2002)는 테러리스트들이 인터넷을 선전, 모금, 조직원 모집, 통신 등에 활용하며 네트워크화된 조직 형태를 띤다고 분석하며, 이들의 온라인 활동 패턴 이해의 중요성을 강조했다[12]. 이처럼 SNA는 다양한 유형의 네트워크에서 구조적 특성을 파악하고 핵심 노드를 식별하며, 정보 흐름의 동적인 과정을 이해하는 데 강력한 분석 도구로 활용되고 있다.

2.2.3. 본 연구의 차별성

앞서 검토한 선행연구들은 불법·유해 사이트 및 온라인 범죄 네트워크의 특성과 분석 방법론에 대해 중요한 기초를 제공한다. 그러나 몇 가지 한계점과 연구 공백이 존재한다. 첫째, 국내 연구의 경우 송봉규(2020)의 연구와 같이 특정 연결 방식(예: 배너 광고)에 초점을 맞추거나[1], 특정 유형의 불법 사이트(예: 불법 스트리밍 사이트[2], 불법 도박 조직[6])를 대상으로 하는 경우가 많아, 다양한 유형의 불법·유해 사이트를 포괄하는 전체적인 네트워크 구조에 대한 이해는 부족하였다. 둘째, 국제적으로 다크웹 분석 연구는 활발하지만[3,4], 데이터 수집의 극심한 어려움과 네트워크의 역동성으로 인해 포괄적이고 안정적인 네트워크 구조를 파악하는 데 한계

가 있었다. 셋째, 일부 연구는 정성적 분석이나 특정 기술 개발에 중점을 두어, 실제 네트워크 데이터에 기반한 대규모 정량 분석과 이를 통한 구체적인 차단 전략 제시는 미흡한 측면이 있었다.

본 연구는 이러한 선행연구들의 공백을 메우고 다음과 같은 차별성을 확보하고자 한다. 첫째, 본 연구는 특정 유형에 국한되지 않고 불법 도박, 불법 성인물, 저작권 침해 사이트 등 다양한 유형의 불법·유해 사이트를 포괄하는 대규모 네트워크를 분석 대상으로 한다. 둘째, 웹 크롤링을 통해 수집된 실제 사이트 간 연결 데이터를 기반으로 SNA 기법을 적용하여 정량적으로 네트워크의 구조적 특성을 분석한다. 셋째, 연결 중심성, 매개 중심성, 페이지랭크 중심성 등 다양한 중심성 지표와 커뮤니티 탐지 알고리즘(Leiden 알고리즘)을 활용하여 네트워크 내에서 영향력 있는 핵심 사이트와 긴밀하게 연결된 하위 그룹을 식별하고, 이를 통해 효과적인 차단 우선순위를 제시하고자 한다. 이는 기존 연구에서 충분히 다루어지지 않았던, 다양한 불법·유해 사이트로 구성된 복합적인 네트워크 전체를 조망하고, 데이터 기반의 실질적인 차단 전략 수립에 기여한다는 점에서 본 연구의 주요 차별점이라 할 수 있다. 이러한 접근은 불법·유해 사이트의 확산을 효과적으로 억제하고 안전한 인터넷 환경을 조성하는 데 기여할 것으로 기대된다.

III. 연구설계

3.1. 데이터설계

본 연구는 불법·유해 사이트 간의 네트워크 구조를 분석하기 위해 웹 크롤링 기법을 활용하여 데이터를 수집하였다. 파이썬의 Selenium 라이브러리를 사용하여 각 사이트의 HTML 소스에서 외부 링크를 추출하였으며, 이들 링크는 사이트 간 연결 관계를 나타내므로 네트워크 분석의 기초 자료로 활용되었다. 국내에서도 특정 유형의 불법 사이트(예: 도박사이트)의 특징점을 이용한 링크 수집 및 분류, 탐지 및 모니터링 시스템 구현에 대한 연구가 진행된 바 있다 [13,14]. 데이터 수집 과정에서는 구글, 네이버, 페이스북 등 주요 검색 엔진과 SNS 플랫폼의 링크를 제외하였으며, Cloudflare와 같은 보안 서비스로 차단된 링크 및 JavaScript 오류가 발생하는 링크도 필터링하여 분석의 정확성을 높이고자 하였다.

크롤링 과정에서 시작 사이트는 불법·유해 사이트 주소를 모아놓은 대표적인 사이트인 '주소 월드'로 설정하였다. 또한, 불법·유해 사이트 간의 직접적 연결뿐만 아니라 간접적 연결까지 포함하여 네트워크의 연결성을 파악하기 위해 크롤링 깊이를 2로 설정하였다. 크롤링 깊이 2는 시작 사이트에서 연결된 모든 사이트(1단계)를 수집한 뒤, 수집된 사이트들로부터 다시 연결된 사이트(2단계)를 추가로 수집하는 방식으로, 최대 두 단계 떨어진 사이트까지의 모든 연결을 수집할 수 있도록 하였다.

<Figure1> The structure of linked sites

이러한 방식으로 시작 사이트로부터 두 단계 이내의 모든 사이트 간 연결을 수집한 결과, 약 8,000개의 연결이 확보되었다. 이후 구글, 네이버, 페이스북 등 주요 사이트와 JavaScript, Cloudflare 관련 오류를 제거하여 최종적으로 약 4,000개의 연결을 분석 대상으로 선정하였다.

<Table 1> Source URL and target URL

Source URL	Target URL
https://jusowd.com/	https://toonkor370.com/
https://jusowd.com	https://blacktoon306.com/
https://jusowd.com	https://newtoki341.com/
https://jusowd.com	https://joatoon34.com/
https://jusowd.com	https://booktoki341.com/
https://jusowd.com	https://agit362.com/
https://jusowd.com	https://sektoon65.org/
https://jusowd.com	https://yatoon157.com/
https://jusowd.com	https://funbe390.com/
https://jusowd.com	https://newtoon228.com/
https://jusowd.com	https://tv43.wiki/
https://jusowd.com	https://xlqlahswnth.site/
https://jusowd.com	https://hoohootv65.xyz/
https://jusowd.com	https://tvhot.wiki/
https://jusowd.com	https://daiso1.org/
https://jusowd.com	https://coktv.live/
https://jusowd.com	https://mzgvod01.com/
https://jusowd.com	https://s57.sonagitv.live/
https://jusowd.com	https://mvking.org/

3.2. 분석 방법

3.2.1. 중심성 분석

중심성 분석은 네트워크 내에서 중요한 노드와 엣지를 식별하기 위해 사용되는 기법으로, 네트워크의 구조적 특성을 이해하고 각 노드가 네트워크 내에서 차지하는 중요도를 평가하는 데 도움을 준다. 사회 연결망 분석의 원리를 범죄 수사에 적용하여 그 활용 방안을 모색하는 연구는 SNA 기법의 수사적 가치를 뒷받침한다 [15]. 본 연구에서는 연결 중심성, 매개 중심성, 페이지랭크 중심성 지표를 활용하여 불법·유해 사이트 네트워크 내의 주요 사이트를 특정하고자 하였다.

3.2.2. 연결 중심성(Degree Centrality)

연결 중심성은 특정 노드에 직접 연결된 엣지의 수를 나타내며, 연결 중심성이 높은 노드는 다른 노드와 직접적으로 많은 연결을 가지므로 네트워크 내에서 중요한 허브 역할을 한다 (Freeman, 1978)[16]. 이러한 노드는 불법·유해 사이트 네트워크에서 중요한 연결 지점으로, 차단 시 네트워크 흐름에 중대한 영향을 줄 수 있다. 연결 중심성은 다음 수식으로 정의된다.

$$C_{D(v)} = \deg(v)$$

여기서 C_D(v)는 노드 v의 연결 중심성 값이며, deg(v)는 노드 v에 연결된 엣지의 수를 의미한다.

3.2.3. 매개 중심성(Betweenness Centrality)

매개 중심성은 네트워크 내의 다른 노드 쌍 사이의 최단 경로에 특정 노드가 얼마나 자주 포함되는지를 측정한다. 매개 중심성이 높은 노드는 정보 흐름의 중재자로, 해당 노드를 차단함으로써 네트워크 내 정보 전파를 효과적으로 방해할 수 있다(Brandes, 2001)[17]. 매개 중심성은 다음 수식으로 정의된다.

$$C_{B(v)} = \sum_{s \neq v \neq t} \frac{\sigma_{st(v)}}{\sigma_{st}}$$

3.2.4. 페이지랭크 중심성(PageRank Centrality)

페이지랭크 중심성은 링크의 수와 질을 기반으로 노드의 중요성을 평가하는 알고리즘으로, 페이지랭크 중심성이 높은 노드는 다른 중요한 노드들로부터 많이 참조되며 네트워크 내에서 권위 있는 역할을 한다(Page et al., 1999)[18]. 이 지표는 불법·유해 사이트의 확산 및 영향력을 파악하는 데 유용하다. 페이지랭크 중심성은 다음 수식으로 정의된다.

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in M(v)} \frac{PR(u)}{\deg(u)}$$

3.3. 클러스터링

본 연구에서는 불법·유해 사이트 네트워크 내 유사한 특성을 가진 사이트들을 그룹화하기 위해 Leiden 알고리즘을 활용하여 클러스터링을 수행하였다. 클러스터링은 네트워크의 구조적 특성을 파악하고, 네트워크 내에서 밀접히 연결된 그룹(커뮤니티)을 도출하는 데 효과적이다.

3.3.1. Leiden 알고리즘

Leiden 알고리즘은 모듈러리티 최적화를 통해 네트워크 내 노드들을 잘 연결된 그룹으로 분류한다. Louvain 알고리즘의 개선된 버전으로, 커뮤니티가 일관되게 유지되고 잘 연결되도록 설계되었다(Traag et al., 2019)[19]. 이를 통해 불법·유해 사이트가 어떤 그룹으로 묶여 있으며, 특정 커뮤니티가 차단될 경우 네트워크에 미치는 영향을 분석할 수 있다. 이때 사용되는 모듈러리티(Modularity) Q는 네트워크가 커뮤니티로 나뉘었을 때 각 커뮤니티 내 노드들이 얼마나 밀접하게 연결되어 있는지를 평가하는 지표이다. 모듈러리티는 높은 값일수록 커뮤니티 내의 연결 밀도가 높고, 다른 커뮤니티 간의 연결은 상대적으로 적음을 나타낸다. 모듈러리티 Q는 다음과 같은 수식으로 정의된다.

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) * \delta(c_i, c_j)$$

3.4. 시각화

불법·유해 사이트 네트워크의 구조를 시각적으로 이해하기 위해 NetworkX와 Matplotlib 라이브러리를 사용하여 네트워크 시각화를 수행하였다. 시각화는 복잡한 네트워크 데이터를 직관적으로 이해하게 함으로써 네트워크 내에서의 핵심 사이트와 클러스터를 한눈에 파악할 수 있게 한다.

3.4.1. NetworkX 및 Matplotlib

NetworkX는 네트워크의 생성, 조작, 시각화 기능을 제공하는 파이썬 라이브러리로 다양한 알고리즘 적용이 가능하며, Matplotlib는 시각화를 위한 대표적인 파이썬 라이브러리로 다양한 형태의 그래프 생성이 가능하다.

3.4.2. Kamada-Kawai 레이아웃

본 연구에서는 Kamada-Kawai 레이아웃을 사용하여 네트워크 시각화를 수행하였다. 이 알고리즘은 에너지 최소화 원리를 기반으로 노드 간 거리가 실제 연결 강도를 반영하도록 배치하여 네트워크의 구조적 특성을 명확히 보여준다(Kamada & Kawai, 1989)[20]. 에너지 함수 E는 다음과 같이 정의된다.

$$E = \sum_{i<j} \left(k_{ij} (d_{ij} - l_{ij})^2 \right)$$

Ⅳ. 분석 결과

4.1. 중심성 분석

4.1.1. 연결 중심성

연결 중심성 분석에서는 링크의 방향성을 고려하여 In-Degree Centrality와 Out-Degree Centrality로 나누어 분석하였다. In-Degree Centrality 상위 20개의 사이트를 분석한 결과, 불법 도박 사이트가 대다수를 차지하였는데, 이는 외부에서 이들 사이트로의 링크가 많다는 것을 의미하며, 주로 배너 광고 등을 통해 사용자 유입이 이루어지는 것으로 추정된다.

반면, Out-Degree Centrality 분석에서는 주소 모음 사이트의 값이 높게 나타냈으며, 이는 각종 불법 사이트의 주소를 모아 제공하는 특성 때문으로 판단된다. 주소 모음 사이트를 제외하면, 먹튀 검증 사이트의 Out-Degree Centrality가 높은 편인데, 이는 먹튀 검증 사이트가 주로 불법 도박 사이트로 연결되는 링크를 제공하기 때문으로 해석된다.

<Table 2> In-degree and out-degree degree of degree centrality

Top 20 In-Degree Centrality Nodes	URL In-Degree Central	Top 20 Out-Degree Centrality Nodes	URL Out-Degree Central
https://wn-st.com/	37	https://jusodude09.com/	927
https://1bet1.vip/	35	https://www.ygy01.com/	298
https://ww-ot.com/	32	https://ygy01.com/	298
https://xn—1-v78es76b.com/	24	https://linkssakda1.com/	258
https://dis-bb.com/	20	https://jusowd.com/	80
https://spst-ddd.com/	19	https://jusoya.com/	64
https://mcl-ddff.com/	18	https://mttotoinfo.com/	63
https://play-tt.com/	18	https://www.bobaelink55.xyz/	62
https://gc-bbb.com/	18	https://amylink.net/	61
https://bsbs-777.com/	18	https://nsfwkr.net/	61
https://snc-rr.com/	18	https://totosave.com/	56
https://mrc-kk.com/	17	https://totogun.com/	56

https://upup-rr.com/	17	https://tostar01.com/	56
https://btbt-777.com/	17	https://yajuso.com/	56
https://tenca-10.com/	16	https://toinkuk.com/	55
https://fst-ccc.com/	16	https://linksome.net/	53
https://hr-rr.com/	16	https://save-1.net/	53
https://happytoon01.com/	16	https://kr67.sogirl.so/	44
https://pg-kkk.com/	15	https://kr25.4kjav.co/	43
https://lv-ca.com/	15	https://xhamster.com/	38

4.1.2. 매개 중심성

매개 중심성(Betweenness Centrality) 분석에서도 상위 20개 사이트를 검토한 결과, 주소 모음 사이트와 일부 토렌트 사이트가 높은 매개 중심성을 보였다. 주소 모음 사이트는 여러 사이트와 연결이 많아 높은 매개 중심성을 보이는 것이 예상된 결과이다. 주목할 만한 점은, 연결 중심성은 낮지만 매개 중심성은 높은 토렌트 사이트의 경우, 네트워크 내에서 중재자 역할을 수행하며, 정보 흐름의 중심에 있는 것으로 분석된다. 이러한 사이트들은 네트워크의 핵심 중재자 역할을 하므로, 차단 시 네트워크 연결 흐름을 방해하는 효과가 있다.

<Table 3> Table centrality and PageRank centrality

Top 20 betweenness centrality nodes	URL betweenness central	Top 20 PageRank nodes	URL PageRank
https://www.ygy01.com/	0.004187	https://happytoon01.com/	0.001444
https://jusodude09.com/	0.003065	https://wn-st.com/	0.001231
https://yagong77.com/	0.003042	https://ww-ot.com/	0.001171
https://drugharm.life/	0.002047	https://t.me/h2csc	0.001097
https://linkssakda1.com/	0.000735	https://pf.kakao.com/_xdqxiys	0.001020
https://ygy01.com/	0.000672	https://t.me/s/newtoki5	0.001019
https://www.todava78.asia/home.php	0.000350	https://tocatoca.net/bbs/board.php?bo_table=event_rsp&wr_id=1	0.001007
https://torrentqq319.com/	0.000252	https://t.me/tvspeed10	0.000997
https://sometv27.com/video?category1=1	0.000231	https://eoqka24.com/	0.000988
https://www.bobaelink55.xyz/	0.000228	https://1bet1.vip/	0.000988
https://torrentsome147.com/	0.000221	https://spst-1111.com/	0.000929
https://mttotoinfo.com/	0.000172	https://optalk-juso.com/	0.000854
https://jogaeparty133.com/	0.000169	https://whos.amung.us/stats/w4n4vgyo4n/	0.000831
https://totosave.com/	0.000158	https://www.rotwire.com/	0.000825
https://totogun.com/	0.000157	https://xn—gl-qv9iw2zsia.com/	0.000825
https://kr30.topgirl.co/	0.000127	https://hepd43.xyz/default.aspx	0.000819
https://coktv.live/	0.000121	https://bcrvs.com/	0.000795
https://www.qbittorrent.org/download.php	0.000111	https://chamstv8.com/	0.000787
https://blacktoon306.com/	0.000105	https://xn—z27bt9c1e.com/	0.000779
https://www.torrentreel83.site/	0.000104	https://toons.info/	0.000779

4.1.3. 페이지랭크 중심성

페이지랭크 중심성(PageRank Centrality) 분석 결과, 사이트 유형에 따른 명확한 차이는 보이지 않았으나, 불법 웹툰 사이트가 가장 높은 페이지랭크 중심성을 보였으며, 그 뒤를 불법 도박 사이트들이 차지하였다. 페이지랭크 중심성이 높다는 것은 해당 사이트가 네트워크 내에서 중요한 위치를 차지하며, 많은 사이트로부터 참조되거나 연결되어 있음을 의미한다. 따라서 이러한 사이트들은 네트워크 유지에 큰 영향을 미칠 수 있어 차단 우선순위가 높다고 간주될 수 있다.

추가적으로, 페이지랭크 중심성이 높은 사이트들의 세부 정보를 파악하기 위해 Website Informer라는 서비스를 사용하여 도메인 소유 정보를 조회하였다. 그 결과, 페이지랭크 중심성에서 상위 두 번째와 세 번째에 해당하는 사이트가 동일한 이메일 주소로 등록된 것으로 확인되었다. 이를 더 조사한 결과, 해당 이메일 계정을 통해 총 29개의 불법 도박 사이트가 개설되어 있는 사실도 발견되었다. 이는 특정 관리자가 여러 불법 사이트를 관리하며, 불법 사이트 간에 매우 긴밀한 연결이 이루어지고 있음을 보여주는 사례이다.

4.2. 클러스터링

Leiden 알고리즘을 통한 클러스터링 결과, 총 23개의 클러스터가 탐지되었다. 각 클러스터 내의 사이트들은 서로 높은 연결 밀도를 가지며, 이는 긴밀하게 연결된 사이트들의 집합으로 해석할 수 있다.

클러스터의 특성을 더 구체적으로 분석하기 위해서는 각 클러스터에 속한 사이트들이 어떤 유형인지 파악하는 것이 이상적이지만, 본 연구에서는 클러스터별 주요 사이트의 영향력을 평가하는 데 중점을 두어 매개 중심성과 페이지랭크 중심성 지표를 사용하여 분석을 진행하였다. 매개 중심성은 네트워크 내에서 중재자 역할을 하는 사이트를, 페이지랭크 중심성은 다른 사이트로부터 참조되는 빈도가 높은 사이트를 식별할 수 있어, 각각 클러스터 내에서 정보 전달과 연결 중심 역할을 수행하는 핵심 사이트를 도출할 수 있다. 분석 결과, 일부 클러스터에서는 매개 중심성과 페이지랭크 중심성이 모두 높은 사이트가 존재하였으며, 이러한 사이트는 클러스터 내에서 중요한 연결 지점으로 해석된다. 따라서 이들 사이트는 네트워크 차단 시 우선적으로 고려할 대상이 될 수 있다.

<Table 4> Degree centrality clustering

Top betweenness centrality nodes per cluster	
Cluster	Top betweenness Centrality URL
0	https://runpeople02.com/
1	https://www.ygy01.com/
2	https://mrc-kr.com/
3	https://xn-z52bt9duvy.wiki/
4	https://www.utorrent.com/intl/ko/desktop/
5	https://www.webnovel.com/book/
6	https://www.xn—ok0b68ytra.com/
7	https://www.bobaelink55.xyz/
8	https://ygp-ask.com/

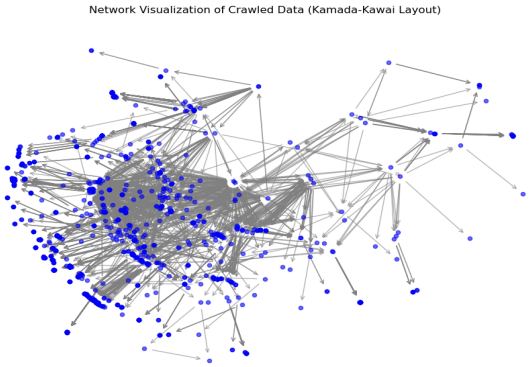
9	https://mco-ccc.com/
10	https://kimsungjune19.wixsite.com/my-site
11	https://yako.net/
12	https://hi.xhamster.com/
13	https://xn-qn1b01gy2eh05a.com/
14	https://coktv.live/
15	https://bugs.qbittoerrent.org/
16	https://qr.kakao.com/talk/4oQ7kasxJxulzhDEmPWKzcxXTQYw-
17	https://www.xn—om2bk9ff1j.com/
18	https://t.me/torrentsee_vip
19	https://api.shatsapp.com/send?text=https%3A%2F%2Fmhthemes.com%2F%20Premium$20WordPress%20Themes
20	https://store.shale.naver.com/detail/pgbmogfigpngnbilikhdjphemfpdmnhl
21	https://info.xvideos.com/
22	https://twitter.com/black16400364

<Table 5> PageRank clustering

Top PageRank nodes per cluster	
Cluster	Top PageRank URL
0	https://runpeople02.com/
1	https://open.kako.com/o/sDK6nov
2	https://spst-1111.com/
3	https://tocatoca.net/bbs/board.php?bo_table=event_rsp&wr_id=1
4	https://eoqka24.com/
5	https://happytoon01.com/
6	https://www.mydomaincontact.com/index.php?domain_name=mango17.me
7	https://www.mydomaincontact.com/index.php?domain_name=bobaelink80.top
8	https://t.me/h2csc
9	https://t.me/vbank77
10	https://hepd43.xyz/default.aspx
11	https://info.xvideos.red/legal/tos
12	https://hi.xhamster.com/
13	https://xn-qn1b01gy2eh05a.com/
14	https://coktv.live/
15	https://bugs.qbittoerrent.org/
16	https://qr.kakao.com/talk/4oQ7kasxJxulzhDEmRWKzcXYQYw-
17	https://avmov.best/
18	https://twitter.com/ICS_kr
19	https://api.shatsapp.com/send?text=https%3A%2F%2Fmhthemes.com%2F%20Premium$20WordPress%20Themes
20	https://store.shale.naver.com/detail/pgbmogfigpngnbilikhdjphemfpdmnhl
21	https://info.xvideos.com/
22	https://twitter.com/black16400364

4.3. 시각화

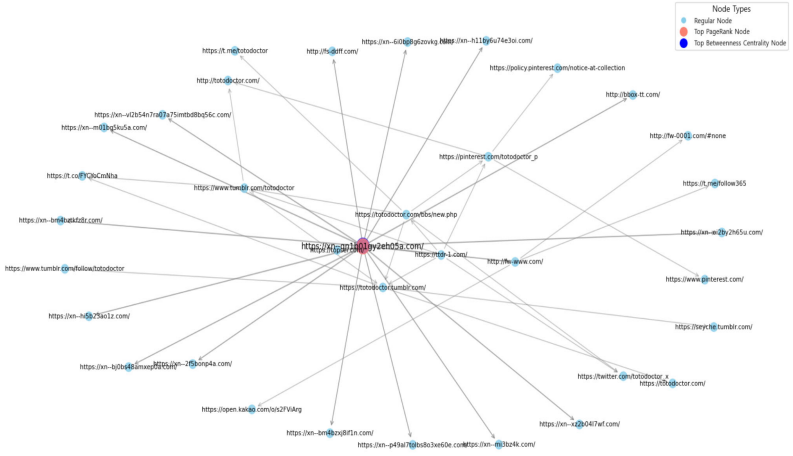
앞서 설명한 방식으로 불법·유해 사이트의 연결 네트워크를 시각화하였다. <Figure 2>는 불법·유해 사이트 간의 연결 구조를 Kamada-Kawai 레이아웃을 통해 시각화한 결과이다. 이 레이아웃은 노드 간의 최단 경로를 고려하여 배치함으로써 네트워크의 구조적 특성을 명확하게 드러낸다. 각 노드는 불법·유해 사이트를, 엣지는 이들 간의 연결을 나타낸다.



<Figure 2> 크롤링 데이터의 시각화 네트워크 분석(Kamada-Kawai layout)

<Figure 3>은 특정 클러스터 내에서 매개 중심성과 페이지랭크 중심성이 모두 높은 노드를 강조하여 시각화한 결과이다. 이 노드는 클러스터 내에서 중요한 연결 지점으로서 역할하며, 높은 페이지랭크 중심성을 통해 다수의 참조를 받고, 높은 매개 중심성을 통해 클러스터 내 사이트 간 정보 전달의 중재자 역할을 수행하고 있다. 이러한 특성을 지닌 사이트들은 네트워크 차단 전략 수립 시 우선적인 차단 대상으로 고려될 수 있다.

특히, 페이지랭크와 매개 중심성이 모두 높은 사이트는 클러스터 내부의 다른 사이트들과 강하게 연결되어 있어, 해당 노드를 차단할 경우 클러스터 전체의 연결성과 정보 흐름에 중대한 영향을 미칠 수 있다. 따라서 <Figure 3>은 이러한 주요 노드가 클러스터 내에서 차단의 우선순위가 높아야 함을 시각적으로 제시하며, 불법·유해 사이트 네트워크 관리에 있어 효율적인 차단 전략 수립에 도움을 줄 수 있다.



<Figure 3> 클러스터 내 매개 중심성과 페이지랭크 중심성이 높은 주요 노드 시각화

V. 결론

본 연구는 불법·유해 사이트 간의 복잡한 연결 네트워크를 사회 연결망 분석(SNA) 기법을 통해 실증적으로 분석하고, 이를 기반으로 효과적인 차단 전략 수립의 토대를 마련하고자 하였다. 웹 크롤링을 통해 수집된 데이터를 바탕으로 네트워크 내에서 핵심적인 역할을 수행하는 주요 노드(사이트)와 긴밀하게 연결된 하위 그룹, 즉 클러스터를 식별하는 데 중점을 두었다.

주요 분석 결과, 첫째, 중심성 분석을 통해 네트워크 내 영향력 있는 사이트들을 다각도로 파악할 수 있었다. 불법 도박 사이트들은 다수의 외부 사이트로부터 링크를 받는 높은 내향 연결 중심성(In-Degree Centrality)을 나타냈으며, 반대로 주소 모음 사이트와 먹튀 검증 사이트 등은 다른 사이트로 다수의 링크를 제공하는 높은 외향 연결 중심성(Out-Degree Centrality)을 보였다. 정보 흐름의 중재자 역할을 하는 매개 중심성(Betweenness Centrality)은 주소 모음 사이트와 일부 토렌트 사이트에서 높게 나타났다. 특히, 페이지랭크 중심성(PageRank Centrality)이 높은 사이트들을 추적하는 과정에서 동일 이메일 계정으로 다수의 불법 도박 사이트가 개설된 정황을 포착함으로써, 특정 운영 주체에 의한 불법 사이트 간 긴밀한 연계 가능성을 확인할 수 있었다. 둘째, Leiden 알고리즘을 활용한 클러스터링 분석을 통해 네트워크는 다수의 군집으로 구성되어 있음을 확인하였고, 각 클러스터 내에서 매개 중심성과 페이지랭크 중심성이 높은 주요 사이트들을 식별하여 선별적 차단 전략의 기초를 마련하였다. 마지막으로, 이러한 분석 결과를 네트워크 시각화를 통해 직관적으로 제시함으로써, 복잡한 연결 구조와 핵심 행위자들의 역할을 명확히 이해할 수 있도록 하였다.

본 연구의 결과는 단순히 개별 사이트를 차단하는 기존 방식의 한계를 넘어, 네트워크 전체의 구조적 특성을 고려한 보다 정교하고 효과적인 불법·유해 사이트 차단 전략 수립에 중요한 학술적·실무적 인사이트를 제공한다. 이는 한정된 단속 자원의 효율적 배분과 불법 유해 콘텐츠의 확산 방지, 나아가 안전한 인터넷 환경 조성에 기여할 수 있을 것으로 기대된다. 이처럼 본 연구는 불법·유해 사이트 네트워크 분석에 대한 유의미한 시사점을 제공하였음에도 불구하고, 다음과 같은 몇 가지 한계점을 지니고 있다.

첫째, 데이터 수집의 완전성이 부족하다는 점이다. 웹 크롤링 과정에서 기술적인 문제(예: JavaScript 기반 동적 로딩, 안티 크롤링 기술)나 사이트 자체의 폐쇄성으로 인해 일부 사이트의 정보가 누락되거나, 분석 목적과 직접적인 관련성이 낮은 사이트가 포함되었을 가능성을 배제할 수 없다. 이는 네트워크의 전체적인 규모나 특정 구조적 속성을 과소 또는 과대평가했을 가능성을 내포한다.

둘째, 시작 사이트(seed site) 선택에 따른 표본 편향 가능성 한계이다. 본 연구는 특정 주소 모음 사이트를 시작점으로 데이터 수집을 진행하였으므로, 분석된 네트워크는 해당 시드 사이트와 직·간접적으로 연결된 생태계를 중심으로 구성되었을 가능성이 높다. 따라서 본 연구의 결과가 국내에서 접근 가능한 전체 불법·유해 사이트 네트워크의 특성을 완벽하게 대표한다고 일반화하기에는 신중한 접근이 필요하다.

셋째, 사이트 유형 분류의 부재로 인한 분석의 심층성 제한이다. 수집된 개별 사이트의 콘텐츠 유형(예: 도박, 음란물, 저작권 침해 등)을 체계적으로 분류하는 라벨링 작업을 수행하지 못하였다. 이로 인해, 식별된 클러스터가 특정 불법 활동을 중심으로 형성되었는지, 또는 어떤 유형의 사이트들이 네트워크 내에서 특정 중심적 역할을 수행하는지에 대한 보다 정밀하고 심층적인 분석을 수행하는 데 한계가 있었다.

상기한 한계점들을 극복하고 연구의 깊이를 더하기 위해 다음과 같은 향후 연구 및 보완 계획을 제언한다. 첫째, 데이터 수집 전략의 고도화 및 다각화가 필요하다. 단일 시드 사이트에 의존

하기보다는, 다양한 출처(예: 다크웹 포럼, 다른 주소 모음 사이트, 법 집행기관 제공 정보 등)로부터 시드 목록을 확보하고, 주기적인 크롤링을 통해 데이터의 최신성과 포괄성을 높여야 한다. 또한, 동적 콘텐츠 수집 및 안티 크롤링 기술에 효과적으로 대응할 수 있는 고급 크롤링 기법의 도입을 고려해야 한다.

둘째, 사이트 유형 자동 분류 시스템 도입을 통한 분석의 정교화가 요구된다. 자연어 처리(NLP) 및 머신러닝 기법을 활용하여 수집된 사이트의 콘텐츠, 키워드, 메타 정보 등을 기반으로 자동으로 유형을 분류 체계를 구축한다면, 각 불법 활동 유형별 네트워크 특성 비교, 클러스터의 주제적 특성 파악 등 훨씬 심도 있는 분석이 가능해질 것이다. 이는 본 연구의 직접적인 보완하고, 불법·유해 사이트 대응 전략 수립에 실무적 기여를 높일 수 있는 유의미한 후속 과제라 할 수 있다.

참고문헌 (References)

[1] Song BG. 2020. The Status and Countermeasures of Connecting Illegal and Harmful Websites: Focusing on Banner Advertising [불법·유해 웹사이트 연결 실태와 대응방안: 배너광고를 중심으로]. The Korean Association of Police Science Review, 22(1), 121-150.

[2] Lim KD, Lee SJ. 2023. Presentation of Discrimination techniques through Association analysis between Harmful streaming sites [불법 저작물 스트리밍 사이트 간 연관성 분석을 통한 판별 기법 제시], Journal of Digital Forensics, 17(3), 37-51. http://doi.org/10.22798/KDFS.2023.17.3.37

[3] Nunes E, Diab A, Gunn A, et al. 2016. Darknet and deepnet mining for proactive cybersecurity threat intelligence. 2016 IEEE Conference on Intelligence and Security Informatics (ISI), 7-12. https://doi.org/10.1109/ISI.2016.7745435

[4] Kühn P, Wittorf K, Reuter C. 2024. Navigating the Shadows: Manual and Semi-Automated Evaluation of the Dark Web for Cyber Threat Intelligence. IEEE Access, 12, 100897-100915. https://doi.org/10.1109/ACCESS.2024.3448247

[5] Jang JY, Lim KD, Lee SJ. 2022. An Harmful site collection system using Characteristic of HTML and URL [HTML 및 URL 특징을 이용한 유해사이트 수집 시스템]. Journal of Digital Forensics, 16(1), 54-63 . http://doi.org/10.22798/KDFS.2022.16.1.54

[6] Moon BH, Kwahk KY. 2024. A case study on the investigation of cyber gambling criminal organizations using social network analysis. Information Systems Review, 26(3), 1-22. http://doi.org/10.14329/isr.2024.26.3.001

[7] Kim HC, Yoon JW. 2020. A case of cyber financial crime investigation through social network analysis (2-mode concepts). Journal of Digital Forensics, 14(4), 449-465.

[8] Pete I, Hughes J, Chua YT, et al. 2020. A social network analysis and comparison of six dark web forums. 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, pp. 484-493. http://doi.org/10.1109/EuroSPW51379.2020.00071

[9] Nouh M, Nurse JRC. 2015. Identifying key-players in online activist groups on the Facebook social network. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, pp. 969-978, http://doi.org/10.1109/ICDMW.2015.88

[10] Kim G, Kim J, Kim JM, et al. 2024. Epidemiological investigation on COVID-19 clusters using geographic information system (GIS) and social network analysis (SNA), data forensic techniques applied in criminal investigations. Journal of Data Forensics Research, 1(1), 125-144. https://doi.org/10.12972/JDFR.2024.1.1.8

[11] Bermingham A, Conway M, McInerney L, et al. 2009. Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. 2009 International Conference on Advances in Social Network Analysis and Mining, Athens, Greece, pp. 231-236. https://doi.org/10.1109/ASONAM.2009.31

[12] Conway M. 2002. Reality bytes: Cyberterrorism and terrorist 'use' of the Internet. First Monday, 7(11), 1001. https://doi.org/10.5210/fm.v7i11.1001

[13] Lee KS, Lim GM, Cho HM. 2021. Technique of Collecting and Classification for Domestic Gambling Websites using Characteristics [국내 도박사이트 특징점을 이용한 링크 수집 및 분류 기법]. Proceeding of 2021 Korean Institute of Information Scientists and Engineers, Jeju, pp. 871-873.

[14] Lee K, Lim G, Ko GH, et al. 2024. Implementation of Detection and Monitoring System of Illegal Gamble Website [불법 도박사이트 탐지 및 모니터링 시스템 구현]. KIISE Transactions on Computing Practices, 30(5), 205-211.

[15] Kim JO. 2019. A Study on the Application of Social Network Analysis Principles to Criminal Investigation [사회연결망 분석원리의 범죄 수사상 활용방안에 관한 연구]. Journal of Digital Forensics, 13(2), 87-107. http://doi.org/10.22798/kdfs.2019.13.2.87

[16] Freeman LC. 1978. Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239. https://doi.org/10.1016/0378-8733(78)90021-7

[17] Brandes U. 2001. A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology, 25(2), 163-177. https://doi.org/10.1080/0022250X.2001.9990249

[18] Page L, Brin S, Motwani R, et al. 1999. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, Stanford, CA, USA.

[19] Traag VA, Waltman L, van Eck NJ. 2019. From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports, 9, 5233. https://doi.org/10.1038/s41598-019-41695-z

[20] Kamada T, Kawai S. 1989. An algorithm for drawing general undirected graphs. Information Processing Letters, 31(1), 7-15. https://doi.org/10.1016/0020-0190(89)90102-6