

보도자료

온라인은 5.28(목) 배포 즉시
지면 5.29(금) 조간
부터 보도해 주시기 바랍니다



자료문의

대외협력팀: 장준용 팀장, 양윤정 담당
컴퓨터공학과: 김정훈 교수

052-217-1227

052-217-2256

사기 집단· 소수 타겟 고객층만 쏙 골라낼 수 있는 네트워크 분석 기술 개발 -UNIST, 전체 네트워크 정보 없이도 관심 대상 중심으로 특정 집단 찾는 알고리즘 개발 - - 고객군 분석· 이상 거래 탐지 등 응용.. 데이터베이스분야 권위 학회 SIGMOD 채택-

복잡하게 얽힌 네트워크에서 사용자가 관심을 갖는 사람이나 대상을 중심으로 진짜 의미 있는 집단만을 효율적으로 찾아내는 알고리즘이 개발됐다. 소규모 열성 고객군에 집중하는 마케팅, 사기 의심 계좌와 연관 관련 계정 집단 추적, 생물학 단백질 관계망 분석을 통한 신약 개발 단서 발굴 등에 도움이 될 것으로 기대된다.

UNIST 컴퓨터공학과 김정훈 교수팀은 사용자가 지정한 대상을 반드시 포함하면서도, 정해진 크기 안에서 의미 있는 집단만 찾아내는 새로운 커뮤니티 탐색 기법을 개발했다고 28일 밝혔다.

커뮤니티 탐색은 방대한 네트워크 데이터 안에서 내부 연결이 강한 집단을 찾아내는 데이터 분석 기술이다. 기존 기술들은 네트워크 데이터가 크거나 개인 정보 보호 문제로 일부 관계만 볼 수 있는 상황에서는 쓰기 어렵고, 관계가 밀접하지 않음에도 집단에 포함되는 문제가 있었다.

연구팀의 기술은 전체 네트워크 정보를 모두 확보하지 않아도, 사용자가 정한 크기 안에서 내부 연결이 촘촘하고 바깥과는 비교적 잘 구분되는 집단을 골라낼 수 있다. 사용자가 지정한 노드에서 출발해 주변 후보를 차례로 확인하며 집단을 넓혀가는 방식이기 때문이다. 이때 후보를 넣었을 때 전체 결과가 좋아지는지를 계산하고, 집단이 불필요하게 커질수록 점수가 쉽게 올라가지 않도록 크기 조건을 반영하게 된다.

하나씩 고르는 과정에서 놓칠 수 있는 관계는 주변의 연결된 작은 묶음을 함께 살펴보는 방식으로 보완했다. 혼자 놓고 보면 눈에 띄지 않지만, 함께 묶었을 때 집단의 성격을 더 분명하게 하는 후보들을 반영하기 위해서다.

실제 네트워크 데이터를 이용한 실험에서도 기존 최고 성능 기법보다 F1 점수는 최대 1.39배, ARI 점수는 최대 5.95배 높게 나타났다. 이는 찾고자 한 집단을 더

정확하게 골라내면서도, 관계가 약한 대상을 덜 포함했다는 의미다.

김정훈 교수는 “현실의 네트워크 분석에서는 전체 데이터를 한 번에 확보하기 어렵고, 실제로 필요한 집단의 규모도 대개 정해져 있다”며 “이번 연구는 사용자가 관심 있는 대상 주변에서 의미 있는 관계만 빠르게 찾아내는 데 초점을 맞춘 기술로, 고객군 분석, 이상거래 탐지, 생물학 단백질 관계망 분석 등 다양한 분야에 적용될 수 있을 것”이라고 말했다.

이번 연구에는 UNIST 김다희 연구원이 제1저자로 참여했으며, 한국연구재단의 지원을 받아 수행됐다.

연구 결과는 오는 5월 31일부터 6월 5일까지 인도 벵갈루루에서 개최되는 2026 SIGMOD에 채택돼 발표될 예정이다. SIGMOD(ACM Special Interest Group on Management of Data)는 데이터베이스 분야 최고 권위 학회 중 하나다. (끝).

(논문명: *LMSC: Local Sketch Modularity Optimisation for Size-Constrained Community Search in Networks*)

붙임 : 연구결과개요, 용어설명, 그림설명, 연구자 이력사항. (끝)

연구 결과 개요

1. 연구배경

SNS(소셜 미디어), 학교·회사 인맥, 협업 네트워크처럼 수많은 사람과 관계가 얽힌 ‘관계망(네트워크)’에서는, 모든 그룹을 한꺼번에 찾기보다 “특정 사람이 속한 친한 그룹”을 알고 싶어 하는 경우가 많다. 예를 들어 “연초 모임을 열 때 누구를 초대 하면 자연스러울까?”, “이번 프로젝트를 함께할 소규모 팀을 꾸리려면 누구와 하는 게 좋을까?”와 같은 질문이 대표적이다.

하지만 관계망이 커질수록 네트워크 전체를 모두 분석해 그룹을 찾는 방식은 시간과 비용 면에서 현실적인 한계가 크다. 또한 이렇게 찾은 그룹이 지나치게 크면, 가게 수용 인원이나 프로젝트 참여 인원처럼 규모 제한이 있는 상황에서는 실제로 활용하기 어렵다. 이 때문에 관심 사용자를 중심으로 하면서도, 원하는 크기 범위 내에서 의미 있는 그룹을 찾는 기술이 필요하다.

기존 커뮤니티 탐색 연구들도 이러한 문제를 다뤘왔지만, 대부분 네트워크 전체 정보인 **전역 정보**를 요구하거나, 미리 정해진 구조 조건(연결 형태, 최소 연결 수 등)에 의존해 실제 데이터에서 유연하게 적용하기 어렵다는 한계가 있었다. 특히 그룹의 ‘적정 규모’를 문제 정의 단계에서 직접 반영하지 못하는 점이 실용성의 제약으로 작용해 왔다.

2. 연구내용

연구팀은 대규모 관계망에서 관심 사용자가 속한 ‘친한 그룹’을 보다 정확하고 실용적으로 찾기 위해, 목적 함수 수준에서 새로운 이론적 기준과 지역적 탐색 전략을 결합한 방법론을 제안했다. 이 접근은 네트워크 전체 정보를 모두 보지 않고도, 원하는 규모의 그룹을 식별할 수 있도록 설계됐다.

연구팀은 실제 관계망에서 ‘친한 그룹’이란 단순히 연결이 많은 집단이 아니라, 그룹 내부는 밀집돼 있으면서도 외부와는 잘 구분되는 구조를 가져야 한다는 점에 주목했다. 기존 방법들은 결과가 불필요하게 커지면서, 실제로 그룹과 긴밀하지 않은 구성원(무임승차자)이 함께 포함되는 문제가 있었지만, 연구팀은 이를 줄이기 위해 LSM(Local Sketch Modularity)이라는 새로운 커뮤니티 품질 지표를 제안했다.

LSM은 관심 사용자 주변의 로컬 정보만을 활용해 후보 그룹의 품질을 평가하며,

그룹이 커질수록 무조건 유리해지는 기존 지표의 한계를 이론적으로 보완한다. 연구팀은 이 지표가 높아지는 방향으로 그룹을 확장하는 탐색 전략을 설계하고, 동시에 그룹 크기를 직접 제한할 수 있도록 해, 모임 인원이나 프로젝트 참여 인원 처럼 규모 제약이 있는 상황에서도 바로 활용 가능한 결과를 제공한다. 즉, “무조건 큰 그룹”이 아니라 “딱 필요한 크기의 친한 그룹”을 찾는 접근이다.

또한 대규모 관계망에서도 효율적으로 동작하도록, 한 명씩 확장하는 방식과 구조적으로 밀집된 부분을 덩어리 단위로 병합하는 방식을 함께 제시했다. 이를 통해 불필요한 탐색과 계산 중복을 줄여, 큰 관계망에서도 관심 사용자 주변의 친한 그룹을 빠르고 안정적으로 식별할 수 있음을 보였다.

3. 기대효과

이번 연구는 대규모 관계망에서도 특정 사용자를 중심으로 ‘현실에서 쓰기 좋은 크기’의 친한 그룹을 효율적으로 찾을 수 있게 해, 다양한 분야에 활용될 것으로 기대된다. 추천 서비스에서는 사용자와 가까운 사람이나 관심사 그룹을 기반으로 보다 자연스러운 추천 후보를 구성할 수 있고, 협업·조직 운영에서는 프로젝트 팀 구성처럼 인원 규모가 중요한 상황에서 후보군을 효과적으로 좁힐 수 있다. 또한 탐지·분석 분야에서는 평소 함께 움직이는 그룹과 다른 이상 패턴을 파악하거나, 관계망 구조를 이해하는 데에도 도움이 될 수 있다. 더 나아가, 본 연구에서 제안한 이론적 기준과 탐색 전략은 향후 동적 네트워크나 다양한 속성을 가진 관계망으로 확장될 가능성도 갖는다.

용 어 설 명

1. 관계망 (네트워크/그래프)

사람, 계정, 조직 같은 대상(노드)와 그 사이의 관계(엣지)를 나타내는 관계 구조

2. 커뮤니티(Community)

관계망 안에서 서로 가깝게 연결된 사람들로 이루어진 '그룹'. 내부 연결이 상대적으로 촘촘한 집단을 의미함.

3. 전역 정보(Global information)

네트워크의 사용자 및 연결의 전체적인 정보

4. 관심 사용자(쿼리, Query)

식별하고자 하는 커뮤니티에 기준이 되는 사용자.

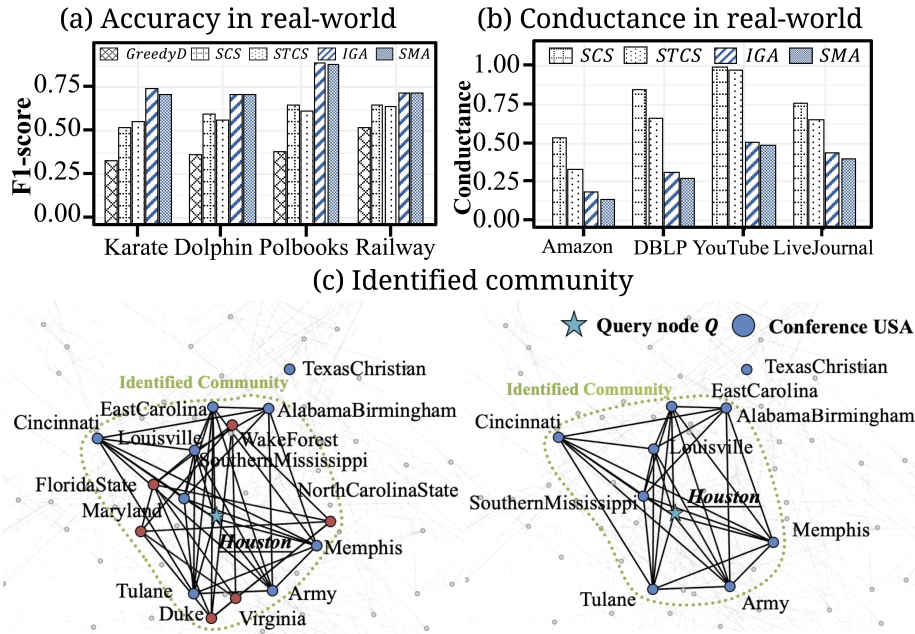
5. 무임승차자(Free-rider)

그룹에 약하게 연결돼 있으면서도 결과에 포함돼, 실제로 '친한 그룹'의 의미를 약화시키는 구성원

6. 로컬 정보(Local information)

관심 사용자 주변의 일부 연결만 활용하는 정보. 전체를 다 보지 않고도 활용할 수 있어 대규모 관계망 분석에 유리함.

그림 설명



그림설명. 실제 관계망 데이터에서 검증한 커뮤니티 식별 성능 비교

(a) 가라테 동아리 관계망, 돌고래 사회관계망, 정치 도서 관계망, 철도망처럼 실제 그룹 구분을 확인할 수 있는 데이터에서 커뮤니티 탐색 정확도를 비교한 결과. F1-score가 높을수록 알고리즘이 찾아낸 그룹이 실제 알려진 그룹과 더 잘 맞는다는 뜻이다.

(b) Amazon, DBLP, YouTube, LiveJournal 등 대규모 실제 관계망에서 찾아낸 그룹의 구조적 품질을 비교한 결과. Conductance는 그룹 안쪽 연결에 비해 바깥과 이어진 연결이 얼마나 적은지를 나타내는 지표로, 값이 낮을수록 내부 구성원 끼리는 촘촘하고 외부와는 뚜렷하게 구분된 집단을 찾았다는 뜻이다.

(c) 관심 사용자를 기준으로 찾아낸 커뮤니티 예시. 빨간색 점은 실제 커뮤니티에 속하지 않는 노드를 나타내고, 파란색 점은 실제 커뮤니티에 속한 노드를 나타내며, 빨간색 점이 적고, 파란색이 많을수록 식별한 커뮤니티의 품질이 우수함을 나타낸다.

김정훈 [교신저재] 교수 이력사항

1. 인적사항

- 소 속 : 울산과학기술원 (UNIST)
- 전 화 : 010-2178-9052
- 이메일 : junghoon.kim@unist.ac.kr



2. 학력

- 2013년: 서울시립대학교 학사
- 2015년: 한국과학기술원 석사
- 2022년: 난양공과대학교 박사

3. 경력사항

- 2015년 ~ 2016년: LG전자 미래IT 연구소
- 2022년 ~ 현재: 울산과학기술원

4. 전문 분야 정보

- 데이터마이닝
- 소셜 네트워크 분석