

# 네트워크 데이터 분석을 위한 링크 그래프 기반 중첩 커뮤니티 탐색 방안

## Overlapping Community Detection Based on Link-Graphs for Network Analysis

홍지원(Jiwon Hong)<sup>1</sup> 이유진(Yujin Lee)<sup>2</sup> 김상욱(Sang-Wook Kim)<sup>3</sup>

### 요 약

현실 세계의 현상 중 다수는 네트워크 데이터로 모델링될 수 있다. 특히, 최근에는 소셜 네트워크 서비스(SNS)의 발달로 다양한 형태의 소셜 네트워크가 새로이 등장하고 있다. 이러한 네트워크 데이터에서 서로 깊은 관계를 갖는 커뮤니티(community)들을 찾아내는 커뮤니티 탐색 기술은 중요한 분석 도구의 하나로 사용된다. 커뮤니티는 노드 간의 관계의 집합으로 정의될 수 있다. 본 논문에서는 노드 간의 관계를 나타내는 엣지가 다수의 커뮤니티에 속할 수 있다는 점에 착안하여 링크 그래프(link-graph)를 바탕으로 seed 확장 기법을 이용하여 엣지 중심적이며 각 엣지가 다수의 커뮤니티에 속하는 것을 허용하는 중첩 커뮤니티 탐색 방안을 제안한다. 이를 바탕으로 기존 중첩 커뮤니티 탐색 방안이 커뮤니티 간의 얇은 중첩만을 허용하는 문제를 해결하고 더 정확한 중첩 커뮤니티 탐색을 달성할 수 있다. 또한, 실험을 통해 제안하는 방법이 중첩 커뮤니티 탐색에 효과적임을 증명하였다.

주제어: 소셜 네트워크 분석, 커뮤니티 탐색, 중첩 커뮤니티 탐색

1 한양대학교 컴퓨터·소프트웨어학과, 박사과정.

2 한양대학교 컴퓨터·소프트웨어학과, 석사.

3 한양대학교 컴퓨터·소프트웨어학과, 교수, 교신저자.

+ 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 (No. NRF-2017R1A2B3004581) 및 한국연구재단-차세대정보·컴퓨팅기술개발사업의 지원 (No. NRF-2017M3C4A7083678)을 받아 수행된 연구임.

+ 논문접수: 2018년 07월 22일, 최종심사완료: 2018년 08월 24일, 게재승인: 2018년 08월 27일.

## Abstract

A number of phenomena in the real world can be modeled as network data. In particular, various types of social networks are emerging recently thanks to the rapid growth of social network services (SNS). Community detection is one of the most important analytic tools, which finds community structures that are densely connected set of nodes. A community can be defined as a set of relationships between nodes. In this paper, we focus on the fact that an edge that represents relationship between nodes can be part of multiple communities. Using link-graph, we adopt seed expansion method to achieve edge-centric overlapping community detection that allows an edge to be in multiple communities. With our proposed method, we can remedy the problem that some of existing overlapping community detection method only allows swallow overlapping and thus achieve more accurate overlapping community detection. We prove that the proposed method is efficient for searching overlapping communities from real world networks via a set of experiments.

Keywords: Social network analysis, community detection, overlapping community detection

## 1. 서론

커뮤니티 탐색(community detection)은 사회학이나 생물학을 포함하여 다양한 분야에서 현실 세계의 현상을 네트워크로 모델링했을 때 유용하게 사용될 수 있는 분석 방법이다[1, 2, 3]. 특히 최근에는 소셜 네트워크 서비스(SNS)가 다수 등장하는 추세이며, 이러한 소셜 네트워크에 대한 분석(SNA; social network analysis)이 나날이 중요해져 가고 있다. 특히, 커뮤니티 탐색은 SNA에서도 유용한 분석 도구이다.

대표적인 네트워크의 종류 중 하나인 소셜 네트워크에서는 주로 사람을 노드(node)로 표현하고, 사람과 사람 사이의 관계 여부를 엣지(edge)로 나타낸다. 다른 네트워크들에서도 어떤 객체를 노드로, 두 객체 사이의 어떠한 연관성을 엣지로 나타낸다. 네트워크 내에서 커뮤니티란 노드들의 집합을 의미하는데, 일반적으로 하나의 커뮤니티 내에서는 해당 커뮤니티에 속한 노드들 간에는 연관성이 많고, 서로 다른 두 커뮤니티에 속한 노드들 간에는 연관성이 적을 것으로 기대한다. 즉, 커뮤니티 내의 노드들 간의 엣지들이 서로 다른 커뮤니티에 속한 노드들 간의 엣지들에 비해 더 많이 존재할 것이라 기대한다.

다양한 네트워크에서 커뮤니티는 중첩(overlap)될 수 있다. 이는 각 노드가 여러 커뮤니티에 동시에 속하는 것이 허용됨을 의미한다. 예를 들어, 같은 지역의 고등학교 동창 커뮤니티와 중학교 동창 커뮤니티는 중복되는 구성원이 존재하여 중첩될 수 있다. 이와 같이 커뮤니티의 중첩은 소셜 네트워크에서 일상적으로 나타나는 개념이다. 대부분의 커뮤니티 탐색 기법들은 일반적으로 중첩되지 않는 커뮤니티들을 찾아내는 데에 초점이 맞추어져 있어 실제로 중첩이 존재하는 네트워크에 적용할 경우 현실적이지 못한 커뮤니티 목록만을 얻을 수 있다.

이러한 경우에는 커뮤니티 간의 중첩이 허용되는 중첩 커뮤니티 탐색(overlapping community detection) 기법[4, 5, 6, 7, 8]들을 이용해야 한다.

노드와 마찬가지로, 노드 간 관계를 의미하는 엣지 역시 여러 커뮤니티에 중복되어 속할 수 있다. 이는 두 노드 간의 관계가 하나의 커뮤니티 내의 관계로만 정의될 수 없는 경우가 존재함을 의미한다. 앞선 중학교 동창 커뮤니티와 고등학교 동창 커뮤니티의 예에서 두 커뮤니티에 모두 속한 어떤 두 구성원이 중학교와 고등학교에 걸쳐 친분 관계를 지니고 있었다면 그 관계를 나타내는 엣지 역시 양쪽 커뮤니티에 동시에 속한다 할 수 있다.

소셜 네트워크에서의 각 노드는 사람을 나타내고, 노드 간의 엣지는 사람과 사람 사이의 관계를 나타낸다. 어떤 사람이 가지는 관계는 다양한 의미를 지닐 수 있다. 가족, 친구, 직장 동료와의 관계들이 네트워크 내의 엣지로 나타나며, 이러한 엣지 중에는 두 개 이상의 관계를 포함하는 경우도 존재한다. 커뮤니티가 각 객체들의 관계 속에서 형성되는 것이라는 점에 주목하면, 그 관계를 나타내는 엣지들이 노드들 대신 커뮤니티 탐색의 대상이 되어야 한다.

본 논문에서는 이와 같이 노드뿐만이 아니라 엣지 역시 여러 커뮤니티에 중복될 수 있다는 점에 착안한 중첩 커뮤니티 탐색 방안을 제안한다. 제안하는 방안은 노드가 아닌 엣지를 중첩 커뮤니티 탐색의 대상으로 삼는다. 각 노드가 속한 커뮤니티는 해당 노드와 연결된 엣지가 속한 커뮤니티로 정의되며, 각 엣지는 여러 커뮤니티에 속할 수 있으므로 그러한 엣지를 다수 지닐 수 있는 노드 역시 여러 커뮤니티에 속할 수 있다. 이를 통해 네트워크 내에 존재하는 중첩된 커뮤니티의 구조를 더욱 잘 반영하여 더 정확한 중첩 커뮤니티 탐색이 가능하다.

본 논문의 2장에서는 제안하는 방안과 관계된 판

런 연구들에 대해 설명한다. 3장에서는 제안하는 방안을 상세히 서술하며, 4장에서 실험을 통해 제안하는 방안의 유용성을 보인다. 마지막으로 5장에서 본 논문을 끝맺는다.

## 2. 관련 연구

본 장에서는 제안하는 중첩 커뮤니티 탐색 방안과 관련된 기존 연구들을 간략히 소개한다. 네트워크 분석 분야에서는 각 노드들이 한 커뮤니티에만 속한다고 가정하는 비중첩(non-overlapping) 커뮤니티 탐색 기법들에 대한 연구가 진행되어 왔다 [1, 2, 3]. 비교적 최근 연구되기 시작한 중첩 커뮤니티 탐색 기법들은 대부분 기존의 비중첩 커뮤니티 탐색 기법에 기반하여 이를 일부 변형함으로써 중첩 커뮤니티를 탐색할 수 있도록 하고 있다 [4, 5, 6, 7, 8].

LinkSCAN\* [6]은 기존의 비중첩 커뮤니티 탐색 기법인 SCAN을 변형하여 중첩 커뮤니티를 탐색한다. LinkSCAN\*은 노드가 아닌 엣지를 커뮤니티 탐색의 대상으로 삼음으로써 커뮤니티의 중첩이 가능하게 한다. 이는 주어진 네트워크를 노드와 엣지의 관계가 역전된 링크 그래프로 변환함으로써 가능해진다 [4, 6, 7]. 링크 그래프에서 SCAN 커뮤니티 탐색 기법을 적용하여 엣지들에 대해 비중첩 커뮤니티 탐색을 수행한 후, 각 노드를 해당 노드가 가진 엣지들이 속한 커뮤니티에 포함시킴으로써 서로 다른 커뮤니티에 속한 엣지들을 보유한 노드가 여러 커뮤니티에 중첩되도록 한다.

[5]에서는 중첩 커뮤니티 탐색을 위해 우선 주어진 네트워크에서 커뮤니티의 핵심이 될 수 있는 seed들을 탐색한다. 이후 이렇게 발견한 seed들을 중심으로 하는 Personalized PageRank(PPR) 점수를 계산하고, 각 seed로부터의 PPR 점수가 가장 높은 노드 순으로 각 seed를 중심으로 하는 커뮤

니티에 포함시켜 커뮤니티를 형성한다. 이와 같은 seed 중심 커뮤니티의 확장은 해당 커뮤니티가 외부와의 연결보다 내부와의 연결이 더 강해지는 방향으로 진행되지만 해당 커뮤니티가 충분히 확장된 후에는 외부와의 연결은 늘어나지만 내부와의 연결이 약해지기 시작한다. 이러한 시점을 파악하여 커뮤니티의 확장을 멈춘다. 이와 같이 커뮤니티의 seed들을 찾은 후 이를 중심으로 하는 커뮤니티를 점차 확장하는 방식으로 탐색하므로, 최종적으로 발견된 커뮤니티들은 중첩될 수 있다. 예를 들어, 비교적 가까이 위치한 두 seed로부터 형성된 두 커뮤니티는 일부 노드를 공유할 것이다.

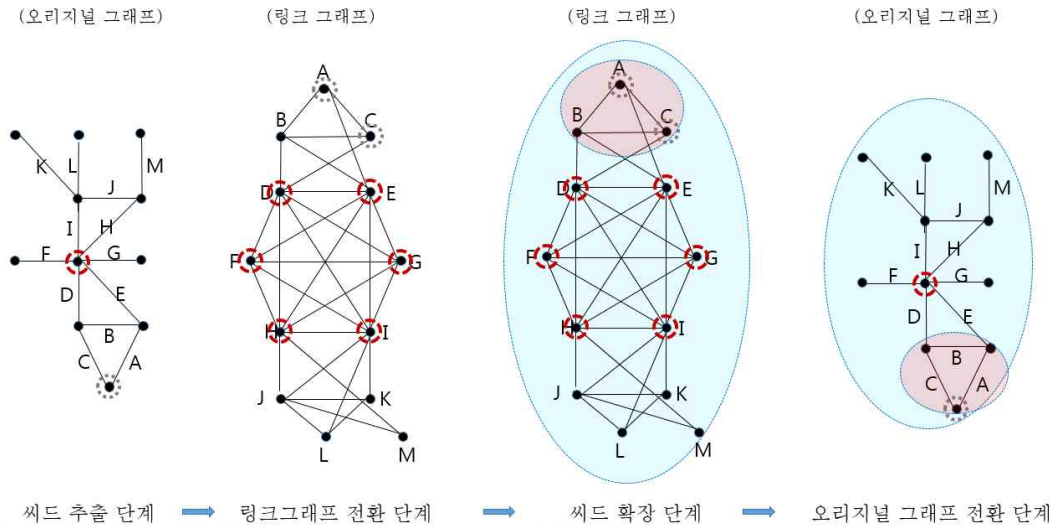
## 3. 제안하는 방안

### 3.1 개요

본 논문에서 제안하는 방안은 주어진 네트워크에서 각 노드가 가지는 관계를 나타내는 엣지가 동시에 여러 커뮤니티에 속할 수 있다는 사실에 착안하여 엣지를 대상으로 중첩 커뮤니티 탐색을 수행한다.

기존의 중첩 커뮤니티 탐색 기법인 LinkSCAN\*은 엣지를 커뮤니티 탐색의 대상으로 삼는 점에서 제안하는 방안과 유사한 착안점을 지니지만 하나의 엣지가 여러 커뮤니티에 속할 수 있는 점은 반영하지 않으며, 그에 따라 각 커뮤니티가 서로 겹치는 정도가 비교적 '얇은' 경향이 있다. 엣지들의 커뮤니티는 비중첩 커뮤니티 탐색 기법을 이용하여 탐색되는데, 이 때 나누어진 커뮤니티를 노드에 반영했을 때, 엣지들을 나눈 커뮤니티 간의 '경계'에 걸친 노드들이 주로 복수의 커뮤니티에 속하게 되며, 다른 노드들은 하나의 커뮤니티만을 갖는 경향이 있다.

또 다른 기존의 중첩 커뮤니티 탐색 기법인 [5]



[그림 1] 제안하는 알고리즘의 진행 단계

에서는 우선 커뮤니티의 중심이 될 수 있는 seed를 찾고, 이로부터 커뮤니티를 점차 확장하여 중첩된 커뮤니티들을 탐색한다. [5]는 LinkSCAN\*과 같이 각 커뮤니티가 겹치는 정도가 얇은 문제는 가지지 않는다.

제안하는 방안은 LinkSCAN\*이 갖는 얇은 중첩 문제를 해결하며, 네트워크의 특성을 더욱 잘 반영하여 기존 seed 확장 기법[5]에 비해 더 정확한 중첩 커뮤니티 탐색을 수행한다.

### 3.2 알고리즘

제안하는 방안은 각 엣지가 다수의 커뮤니티에 속할 수 있도록 하는 중첩 커뮤니티 탐색을 수행한다. 이를 위해 주어진 그래프를 엣지를 중심으로 하는 링크 그래프로 변환한 후 [5]에서 제안된 중첩 커뮤니티 탐색을 적용한다. 그림 1은 제안하는 알고리즘의 수행 순서를 나타낸다.

우선, 주어진 그래프로부터 커뮤니티의 중심이 될 수 있는 seed들을 추출한다. [5]에서는 시드를 추

출하기 위한 ‘spread hubs’와 ‘GraClus centers’ 방안을 소개한다. ‘GraClus centers’ 방법은 효율적인 그래프 분할 알고리즘을 사용하여 대략적인 커뮤니티를 파악한 후 이를 기반으로 seed를 선정하는 것이다. 즉, NEO-k-means를 통해 k개의 클러스터들을 찾고 각 클러스터의 centroid와 가장 가까운 노드들을 선택한다.

‘Spread hubs’ 방법은 seed를 확장시켰을 때, 높은 커버리지를 가질 수 있도록 하는 방법이다. 즉, Degree가 높은 노드 순서대로 k개의 seed를 뽑는 방식이다. 해당 방법은 다른 노드들과 많은 연결을 가지고 있는 노드를 고르는 것이므로 그래프에서 seed 집합을 확장했을 때 높은 커버리지를 가질 수 있고 커뮤니티를 이루는 중요한 핵심이 될 확률이 높다. [5]에서는 실험을 통해 ‘spread hubs’ 방식의 seed 추출이 ‘GraClus centers’ 방식보다 우수함을 보인다. 그러므로 본 논문에서는 ‘spread hubs’ 방법을 이용하여 seed들을 선택한다.

두 번째 단계는, 주어진 원본 그래프를 링크 그래

프로 변환하는 단계이다. 커뮤니티는 유사한 특성들을 가진 관계들이 모여 이루어지므로 엣지들을 대상으로 커뮤니티 탐색을 진행해야 한다. 이를 달성하기 위해 제안하는 방안에서는 원본 그래프를 링크 그래프로 변환한다. 원본 그래프 내 노드들을 연결하는 엣지들을 각각 링크 그래프 내의 노드들로 변환한다. 이후 원본 그래프 내에서 인접했던 (즉, 같은 노드를 공유했던) 엣지들에 대해 링크 그래프에서 해당 노드들을 연결하는 엣지를 생성하여 그래프를 재구성한다. 그림 1의 두 번째 단계는 링크 그래프를 생성하는 단계이다. 이 때 원본 그래프의 두 개의 seed와 인접한 엣지들이 링크그래프에서 두 개의 seed 집합이 된다. 오리지널 그래프 내 한 seed와 연결된 엣지들인 링크 그래프 내 D, E, F, G, H, I노드들이 하나의 커뮤니티를 도출하기 위한 seed가 된다. 그리고 나머지 하나의 seed와 연결된 엣지들인 링크 그래프 내 노드 A, C가 다른 하나의 커뮤니티를 도출하기 위한 seed로 사용된다.

세 번째 단계는, 첫 번째 단계에서 추출한 seed로부터 커뮤니티를 확장하는 단계이다. Seed 추출 단계에서 추출된 seed들은 각 커뮤니티들의 핵심 노드라고 간주한다. 그리고 seed들로부터 커뮤니티의 특성을 가지는 커뮤니티가 구성될 때까지 커뮤니티의 범위를 확장해 나가야 한다. 'spread hubs' 방법을 통해 뽑은 seed 집합을 초기 노드들(initial nodes)로 삼고 personalized PageRank를 계산한다. 이로부터 각 seed를 기준으로 하는 그 외 노드들의 PPR 점수를 획득한 후 PPR 점수의 내림차순으로 정렬한다. 이후 PPR 점수가 높은 노드 순으로 각 seed의 커뮤니티와의 연관성이 높다고 간주하고 해당 커뮤니티에 포함시킨다. 커뮤니티 내 노드들 간의 연결은 강하고 커뮤니티 외부 노드와의 연결은 약한 커뮤니티의 특성을 저하시키지 않을 때까지 더한다. 커뮤니티의 특성은 conductance 값

인  $\Phi(S)$ 를 이용한다:

$$\Phi(S) = \frac{c_s}{\min(\text{Vol}(S), \text{Vol}(\bar{S}))}$$

$u$ 와  $v$ 가 각각 하나의 노드를 의미할 때,  $c_s = |(u, v) : u \in S, v \notin S|$ 는 커뮤니티 S안에 한 노드만 속해 있는 엣지들의 수를 의미한다.  $d(u)$ 는 노드  $u$ 의 degree를 의미할 때  $\text{Vol}(S) = \sum_{u \in S} d(u)$ 는 커뮤니티 S안에 속해 있는 노드들의 degree의 합이라 할 수 있다.  $\text{Vol}(\bar{S})$ 는 반대로 커뮤니티 S에 속하지 않는 노드들의 degree의 합이다. 즉, conductance는 커뮤니티 내 모든 노드와 연결되어 있는 엣지 대비 커뮤니티 밖으로 향하는 엣지의 비율을 의미한다.

각 seed를 중심으로 각 노드를 커뮤니티에 추가 하면서 해당 커뮤니티의 conductance의 값이 커지기 시작할 때 해당 커뮤니티 구성을 멈춘다. [3]에서 커뮤니티 외부로 연결되는 엣지 정도를 측정하는 conductance와 반대로 내부로 연결되는 엣지 정도를 측정하는 modularity를 다룬다. 해당 논문에서는 특정 커뮤니티에 노드들이 추가될수록 modularity가 커지다가 modularity가 줄어들기 시작하는 지점부터는 네트워크 내 모든 노드가 커뮤니티에 추가 되는 시점까지 줄어든다는 것을 확인한다. 그러므로 conductance를 네트워크 내 모든 노드가 커뮤니티에 추가되는 경우까지 측정하지 않아도 된다. 그림 1의 seed 확장 단계는 위의 방법에 따라 커뮤니티가 확장되어 두 개의 커뮤니티가 도출된다. 이와 같은 상황은 학교 커뮤니티 안에 동아리 커뮤니티가 존재하는 상황과 유사하다.

마지막 단계는, 링크 그래프를 일반적인 그래프로 다시 변환하는 단계이다. 위의 세 단계를 거쳐 엣지들이 속한 커뮤니티들을 추출한 다음에는 링크 그

래프 내 엣지들을 대상으로 추출된 커뮤니티를 원본 그래프의 노드 대상의 커뮤니티로 변환한다. 그림 1의 링크 그래프로 변환단계를 보면 최종적으로 노드들이 속하는 두 개의 중첩된 커뮤니티를 확인할 수 있다.

## 4. 실험

본 장에서는 실험을 통해 제안하는 방법의 효용성을 보이하고자 한다.

### 4.1 실험 환경

실험을 위해 두 개의 실세계(real-world) 네트워크 데이터를 준비하였다. 데이터는 Stanford Network Analysis Project에서 제공한 Amazon 네트워크와 DBLP 네트워크를 사용한다[9, 10].

Amazon 데이터는 Amazon 웹사이트로부터 크롤링하여 수집한 네트워크이다. Amazon 웹사이트에서 만약 상품  $i$ 가 상품  $j$ 와 빈번히 함께 구매된다면, 그래프에서 노드  $i$ 와  $j$ 는 비방향성(undirected)의 엣지가 연결된다. Amazon 네트워크의 정답(ground-truth)으로 사용되는 커뮤니티는 Amazon에 의해 나누어진 각 상품의 카테고리이다. 상품 카테고리에서 나타나는 각 connected component를 각 정답 커뮤니티로 간주하였다. 또한 세 개보다 적은 노드들로 이루어진 정답 커뮤니티의 경우 제거되었다.

DBLP 네트워크는 DBLP computer science bibliography에서 제공하는 컴퓨터 과학 분야의 연

구 논문들의 리스트를 기반으로 구성된다. 만약 두 저자들이 하나 이상의 논문에서 공동 집필했다면 두 저자는 엣지로 연결이 되는 방식으로 공저자(co-authorship) 네트워크가 구성된다. DBLP 네트워크의 정답 커뮤니티는 저널이나 컨퍼런스와 같은 publication venue로 구성된다. 즉, 특정 저널 또는 컨퍼런스에 같이 참여한 저자들로 하나의 커뮤니티를 구성한다. 또한, 세 개보다 적은 노드들로 구성된 정답 커뮤니티들의 경우 제거되었다.

실험은 제안하는 방안과 기존의 두 가지 방안인 [5]와 LinkSCAN\*이 도출하는 커뮤니티의 퀄리티를 보기 위해 정답 커뮤니티를 얼마나 잘 찾아내는지 정확도를 비교한다. [5] 방안은 seed 추출을 위해 ‘spread hubs’ 방식을 이용한다.

제안하는 방안과 기존 seed 확장 방안에서의 커뮤니티 개수  $k$ 는 25, 000이고 LinkSCAN\*의  $\epsilon$ 는 0.17,  $\mu$ 는 0.7로 설정하였다. 알고리즘 실행 결과 도출된 커뮤니티들  $C$ 와 정답 커뮤니티들  $S$ 를 가지고 정확도 측정을 위해  $F_1$ -measure와  $F_2$ -measure를 사용한다.  $F$ -measure는 precision과 recall이 가지는 trade-off 관계를 극복하기 위해 두 값을 통합한 지표로, 가중치를 가진 조화 평균(weighted harmonic mean)이라고도 한다.  $F_1$ -measure는 precision과 recall의 중요성을 동등하게 평가하는 것이고  $F_2$ -measure는 recall의 중요성을 precision보다 높게 평가하는 것이다.  $F$ -measure의 수식은 아래와 같다:

[표 1] 실험 그래프

그래프	노드 수	엣지 수	Ground-truth 커뮤니티 개수
DBLP	317,080	1,049,866	13,477
Amazon	334,863	925,872	75,149

$$F_{\beta}(S_i) = (1 + \beta^2) \frac{precision(S_i) \times recall(S_i)}{\beta^2 \times precision(S_i) + recall(S_i)}$$

이 때,  $\beta$ 는 non-negative real value이고  $S_i \in S$ 에 대한 precision과 recall을 통합한다.

정밀도(precision)와 재현율(recall)은 다음과 같이 정의된다:

$$precision(S_i) = \frac{|C_j \cap S_i|}{|C_j|}$$

$$recall(S_i) = \frac{|C_j \cap S_i|}{|S_i|}$$

$C_j \in C$ 이고,  $j^* = \operatorname{argmax} F_{\beta}(S_i, C_j)$ 일 때,  $F_{\beta}(S_i) = F_{\beta}(S_i, C_{j^*})$ 이다.

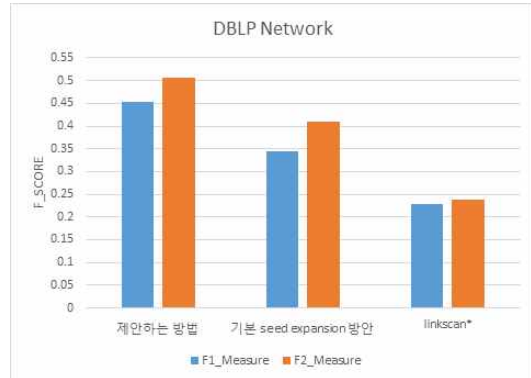
### 4.2 실험 결과

본 논문에서 제안하는 방안의 효용성을 확인하기 위해 [5]에서 제안된 seed 확장 방안과 [6]의 LinkSCAN\* 방안과의 정확도를 비교하였다.

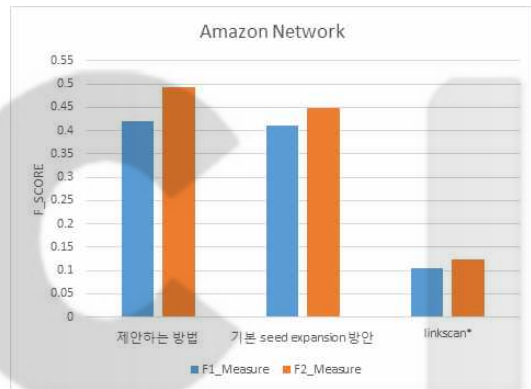
그림 2는 DBLP 네트워크에서 각 방안의  $F_1$ 과  $F_2$  점수를 보인다. 그림 2에 따르면 제안하는 방안이 seed 확장 방안에 비해 높은 정확도를 보이며, LinkSCAN\*에 비해서는 약 두 배에 달하는 높은 정확도를 보인다.

그림 3은 Amazon 네트워크에서의 각 방안의 정확도를 보인다. 여기서도 제안하는 방안이 가장 높은 정확도를 가지며 LinkSCAN\*에 비해서는 약 4 배에 달하는 높은 정확도를 가진다.

이를 통해 제안하는 방안의 정확도가 기존 방안보다 우수함을 확인할 수 있다. 특히, LinkSCAN\*과 같이 얇은 중첩 문제가 발생하지 않는 seed 확



[그림 2] DBLP 네트워크에서 정확도



[그림 3] Amazon 네트워크에서의 정확도

장 기법보다 높은 정확도를 보인 점으로 미루어 옛지를 커뮤니티 탐색의 대상으로 하는 관점이 효과적임을 확인할 수 있다.

## 5. 결론

본 논문에서는 중첩 커뮤니티 탐색에 있어 옛지를 커뮤니티 탐색의 대상으로 삼으며 각 옛지가 다수의 커뮤니티에 속하는 것을 허용하는 것이 더 현실적인 소셜 네트워크에서의 중첩 커뮤니티 탐색에 효과적임을 보였다. 본 논문에서 제안하는 중첩 커뮤니티 탐색 방안은 주어진 네트워크를 링크 그래프로 변환함으로써 커뮤니티 탐색의 대상을 옛지로 삼고, 이



링크 그래프에 기존 중첩 커뮤니티 탐색 방안 중 하나인 seed 확장 방안을 적용하였다. 이에 따라 각 엣지가 다수의 커뮤니티에 속하는 것이 허용되며 얇은 중첩 문제도 해결될 수 있다. 본 논문에서는 실험을 통해 제안하는 방안이 기존 중첩 커뮤니티 탐색 방안에 비해 더 정확하게 중첩 커뮤니티를 찾아냄을 확인하였으며, 엣지를 커뮤니티 탐색의 대상으로 삼는 것이 효과적으로 실제 네트워크에서의 정확한 커뮤니티 탐색을 가능케 함을 확인하였다.

## 참고 문헌

- [1] M. E. J. Newman, "Modularity and Community Structure in Networks," In Proc. of the National Academy of Sciences of the United States of America, vol. 103, no. 23, pp 8577-8582, 2006.
- [2] M. E. J. Newman, "Fast Algorithm for Detecting Community Structure in Networks," Physical Review E, vol. 69, 066113, 2004.
- [3] A. Clauset, M. E. J. Newman, and C. Moore, "Finding Community Structure in Very Large Networks," Physical Review E, vol. 70, 066111, 2004.
- [4] T. S. Evans and R. Lambiotte, "Line Graphs, Link Partitions, and Overlapping Communities," Physical Review E, vol. 80, 016105, 2009.
- [5] J. J. Whang, D. F. Gleich, and I. S. Dhillon, "Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion," In Proc. of IEEE Trans. on Knowledge and Data Engineering, vol. 28, no. 5, pp. 1272-1284, 2016.
- [6] S. Lim, S. Ryu, S. Kwon, K. Jung, and J.-G. Lee, "LinkSCAN\*: Overlapping Community Detection using the Link-Space Transformation," In Proc. of IEEE Int'l. Conf. on Data Engineering, pp. 292-303, 2014.
- [7] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping Community Detection in Networks: The State-of-the-art and Comparative Study," ACM Computing Surveys, vol. 45, no. 4, p. 43, 2013.
- [8] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai, "Overlapping Communities in Dynamic Networks: Their Detection and Mobile Applications," In Proc. of ACM Int'l. Conf. on Mobile Computing and Networking, pp. 85-96, 2011.
- [9] <https://snap.stanford.edu/data/com-DBLP.html>
- [10] <https://snap.stanford.edu/data/com-Amazon.html>



홍 지 원

2009년 한양대학교 정보통신대학  
컴퓨터전공 학사  
2009년~현재 한양대학교 대학원  
컴퓨터·소프트웨어학과 박사과정  
재학

관심 분야: 사회연결망 분석, 보안, 데이터 마이닝



이 유 진

2014년 숙명여자대학교 컴퓨터 과  
학과 학사  
2017년 한양대학교 대학원 컴퓨터·  
소프트웨어학과 석사

관심 분야: 사회연결망 분석, 추천 시스템



김 상 옥

1989년 서울대학교 컴퓨터공학과  
학사  
1991년 KAIST 전산학과 석사  
1994년 KAIST 전산학과 박사

1991년~1991년 미국 Stanford University,  
Computer Science Department, 방문 연구원  
1994년~1995년 KAIST 정보전자 연구소 전문 연구원  
1999년~2000년 미국 IBM T.J. Watson Research  
Center, Post-Doc.  
1995년~2003년 강원대학교 정보통신공학과 부교수  
2003년~현재 한양대학교 공과대학 컴퓨터공학부 교수  
2009년~2010년 미국 Carnegie Mellon University,  
Visiting Scholar  
관심 분야: 데이터베이스 시스템, 데이터 마이닝, 멀티미  
디어 검색, 사회연결망 분석, 웹 데이터 분석, 딥 러닝