

# 복잡계 네트워크 분석도구로써 k-core 알고리즘의 응용

## Application of K-core Algorithm as a Tool for Analyzing Complex Network

류제운\*, 구자을\*, 박별나\*, 조성진\*,  
 유재수\*\*, 김학용\*  
 충북대학교 생화학과\*, 충북대학교 정보통신공학부\*\*

Ryu Jea-woon, Ku Jaeul, Park byeol-na,  
 Cho seong-jin, Yoo Jae Soo, Kim hak-yong  
 Chungbuk National Univ.

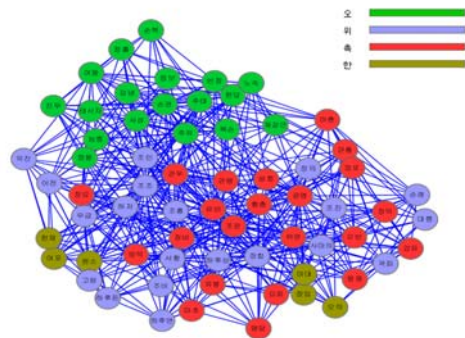
### 요약

복잡계 과학의 발달에 따라 많은 사회 네트워크들이 분석되고 있다. 우리는 연결선수, 중간성(betweenness), 결집계수와 같은 링크 수를 중심으로 네트워크의 구조적 분석에서 나아가 복잡한 네트워크 속에서 핵심 되는 중심 모듈을 찾아 분석하였다. K-core 알고리즘은 복잡한 네트워크를 가중치가 낮은 링크와 노드를 단계적으로 제거하여 복잡한 네트워크의 의미를 분석함에 있어 핵심이 되는 모듈을 얻는데 용이하다. 이에 소설, 영화, 과학 교과서, 단백질 상호작용 네트워크와 같은 다양한 분야에 이 알고리즘을 직접 적용해보았다. 그 결과, 각기 복잡한 네트워크로부터 핵심이 되는 모듈을 찾아낼 수 있었고, 전체 네트워크에서는 발견하기 힘든 유용한 정보들을 도출할 수 있음을 확인하였다. 본 연구에서 k-core 알고리즘을 통해 핵심 네트워크를 구축하여 유용한 정보를 도출할 수 있는 가능성을 제시하였다.

## I. 서론

복잡한 네트워크로부터 핵심이 되는 중심모듈을 찾는 유용한 방법 중 하나로 k-core 알고리즘이 있다 [1]. 이 알고리즘은 네트워크의 노드들이 k 이상의 링크수를 가지는 경우, core값을 k로 정의한다. 즉 (k+1)-core 네트워크는 (k+1)의 링크수를 가지는 노드들로 구성된 네트워크로 이 네트워크가 형성될 때까지 링크수가 k 이하인 노드를 제거하는 과정을 반복하게 된다. 가중치가 낮은 링크와 노드를 단계적으로 제거함으로써 핵심모듈을 얻는데 용이하다. 사회네트워크에서부터 생명체네트워크에 이르기까지 다양한 네트워크를 분석해 봄으로써 k-core 알고리즘의 유용성을 확인하였다.

소설 삼국지 등장인물 네트워크를 구축하고 k-core 알고리즘을 이용하여 핵심 네트워크를 얻었다. 삼국지에 등장하는 한, 촉, 오, 위 나라에 중심이 되는 인물들이 네트워크상에서도 함께 어우러져있는 것을 확인하였다. 또한 국가의 흥망성쇠에 따라 주요 인물의 퇴진과 등장과 같은 변화를 일목요연하게 확인할 수 있었다.



▶▶ 그림 1. 삼국지 핵심 네트워크

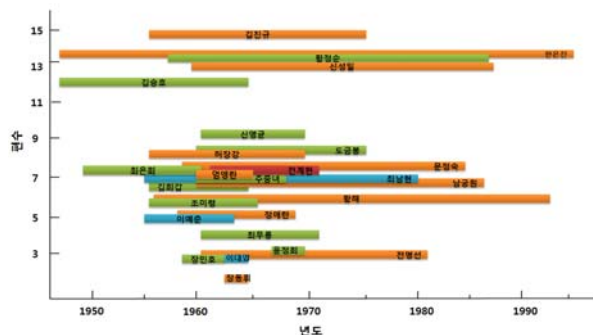
## II. 본론

### 1. 소설 네트워크

소설 삼국지의 이야기 흐름이 유비의 촉나라 중심으로 기술되어 있기 때문에, 유비와 밀접한 상호관계를 가지던 인물 중심으로 네트워크가 구성되어 있다 [2]. 전체 네트워크뿐만 아니라 시기별 세분화한 네트워크를 구축하여 네트워크의 동적변화와 등장인물의 변화에 대한 정보를 얻을 수 있었다.

## 2. 영화 네트워크

한국영화 100선에 등장하는 영화배우 네트워크를 구축하고, k-core 알고리즘을 이용하여 핵심 네트워크를 얻었다. 한국영화 100선은 예술성을 바탕으로 선정되었기 때문에 인지도가 많은 배우가 핵심 네트워크의 중심이 되기보다는 의외의 배우가 핵심 네트워크의 중심에 등장하였다. 특히 1960년대에 제작된 영화가 100선에 많이 포함되어 있기 때문에 당시 활약한 배우들이 대다수를 차지하고 있으며, 이들 소수의 영화배우들이 집중적으로 출연하였기에 다른 기간에 비해 매우 네트워크 상에서 뭉침계수가 높음을 확인하였다.



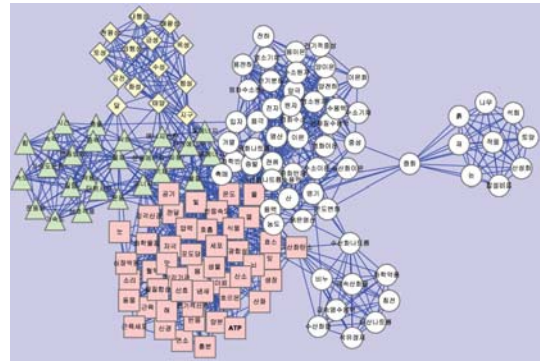
▶▶ 그림 2. 각 배우가 활동한 연도별 편수 그래프

## 3. 과학교과서 네트워크

고등학교 과학교과서에 실린 과학용어 네트워크를 구축하고 k-core 알고리즘으로 과학교과서 핵심네트워크를 얻었다. 전체 네트워크에서는 보이지 않던 물리, 화학, 생물, 지구과학 용어들이 정확하게 분리되는 모듈 구조를 보여주고 있음을 확인하였다.

본 네트워크에서 추출한 핵심 용어는 과학현상을 설명하는데 중심이 되기 때문에, 이 용어를 충분히 이해한다면 고등학교 수준에서의 과학을 이해하는데 충분하

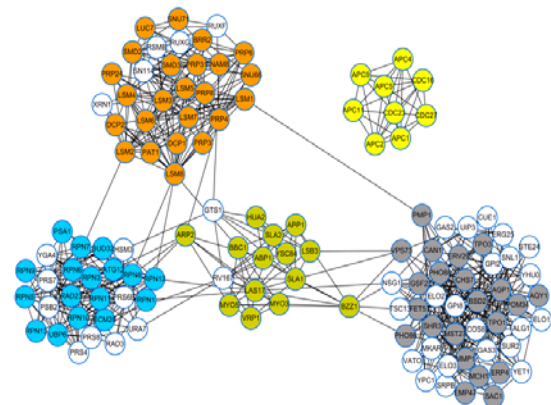
다고 보며, 아울러 본 연구에서 제시한 이들 용어들을 데이터베이스화 한다면 과학 교과서를 바탕으로 상관관계를 분석하여 다양한 학습 분야에 모델로 활용할 수 있을 것으로 기대한다.



▶▶ 그림 3. 과학 디딤돌 교과서 핵심 과학용어 네트워크

## 4. 단백질 상호작용 네트워크

효모의 단백질-단백질 상호작용 네트워크(노드:4,937, 링크:17,553)로부터 각 유전자의 localization 정보를 이용해 세포질의 로컬 네트워크(노드:3,712 링크:9,114)를 만든 후, k-core 알고리즘을 통해 핵심 네트워크를 얻었다 [3].



▶▶ 그림 4. 세포질 단백질-단백질 상호작용 핵심 네트워크

핵심 네트워크는 core 6에서 잘 분리된 5개의 모듈을 형성하고 그 각각의 모듈은 모듈의 독자적인 기능을 수행하는 단백질끼리 뭉쳐져 있으며, 모듈과 모듈을 연결하는 단백질이 효모 세포에서 매우 중요한 역할을 하고

있음을 확인하였다.

### III. 결론

본 연구에서는 복잡한 네트워크로부터 k-core 알고리즘을 사용하여 핵심이 되는 중요 모듈을 찾아 실체를 분석해 봄으로써, 구조 네트워크로부터 정보를 얻을 수 있는 정보 네트워크로 형상화하여 유용한 정보를 얻고자 하였다. 따라서 k-core 알고리즘은 앞으로 복잡한 네트워크에 숨겨진 정보를 얻기 위한 중요한 도구로 활용될 수 있을 것이다.

### ■ 참고 문헌 ■

- [1] Tong A. H. Y., "A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules", Science, Vol. 295, No. 5553, pp. 321-324, 2002.
- [2] 이윤경 "소셜 삼국지 등장인물 네트워크의 동적 변화 분석", 한국콘텐츠학회논문지, 제9권, 제4호, pp.364-371.
- [3] 류제운 "단백질 상호작용 네트워크에서 필수 단백질의 견고성 분석", 한국콘텐츠학회논문지, 제8권, 26호, pp.74-81.

\* 본 논문은 2010년 교육과학기술부와 한국산업기술재단의 "지역혁신인력양성사업"과 교육과학기술부의 "지역거점연구단육성사업/충북BIT연구중심대학사업단"의 지원을 받아 수행되었음.