

링크-공간 변환을 통한 중첩을 허용하는 군집화 알고리즘

임성수

KAIST Applied Algorithm Lab

KAIST

Joint work with nice guys



권세정
문화기술대학원
KAIST



유승우
삼성종합기술원



정교민
전산학과
KAIST

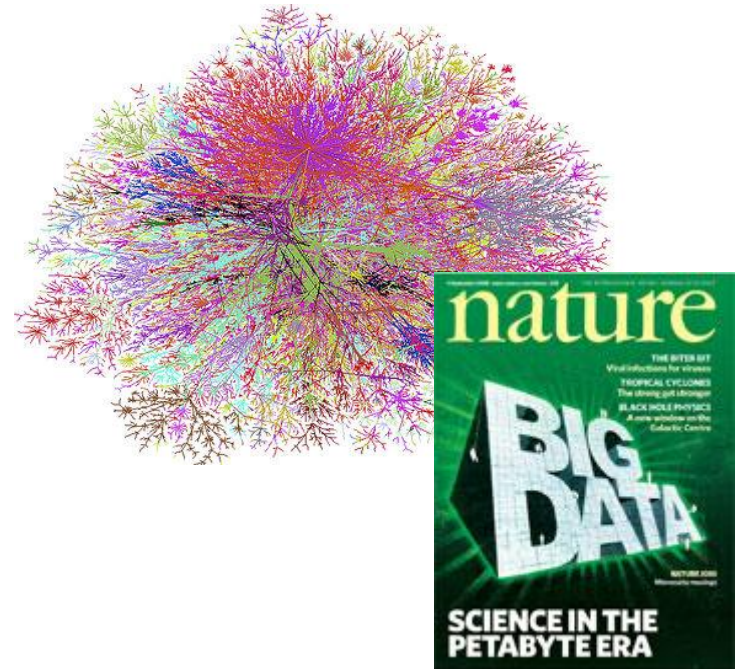


이재길
지식서비스공학과
KAIST

Big Data Analysis

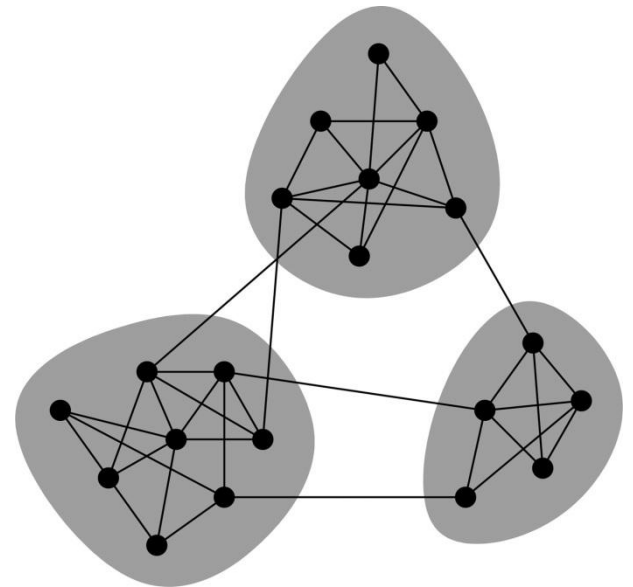
- The era of 'big data'
 - 정보화 시대 도래로 인한 데이터 폭증의 시대
 - 기존의 방식으로 분석 및 처리하기 어려운 데이터
 - 목표: **의미 있는 정보를 추출하고 사회 현상을 이해**

- 분석 기술
 - 데이터 마이닝, 기계 학습, 자연 언어 처리, 패턴 인식 등
 - 데이터로부터 추출된 특성들을 분류하고 예측하는 **알고리즘 개발**



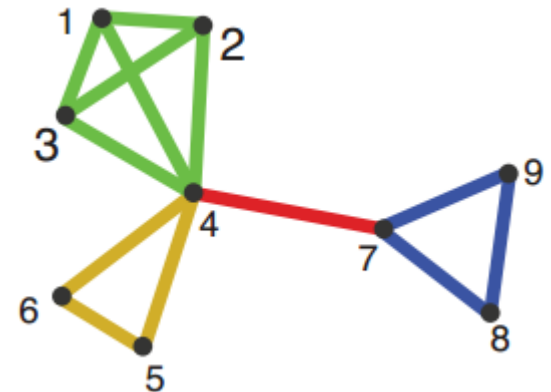
Clustering

- 네트워크
 - 노드(node): 구성요소, 행위자
 - 링크(link): 관계, 상호작용
- 네트워크 군집화
 - 노드를 여러 **군집(cluster)**으로 나눈다.
 - 동일 군집 내부에서 연결이 많고, 다른 군집 사이에는 연결이 적도록 한다.
 - 유사한 항목끼리 묶어 **구조를 발견**한다.



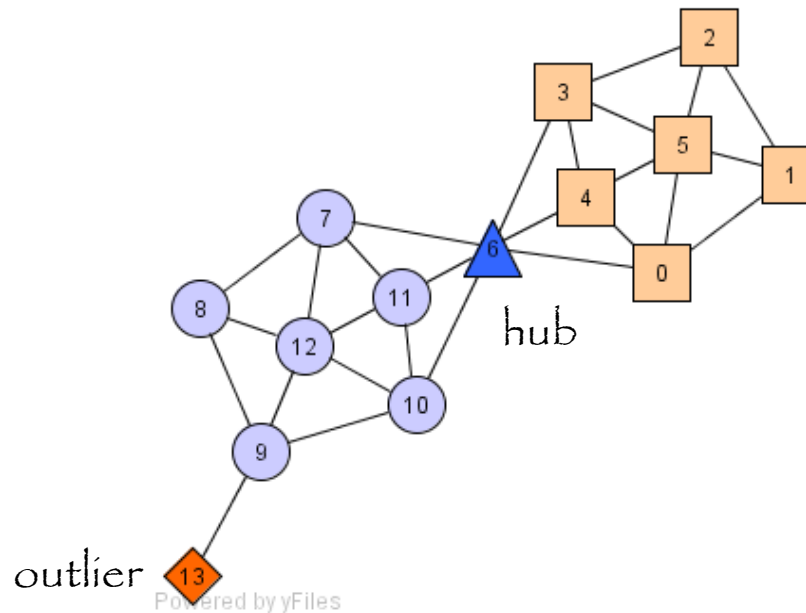
Link Clustering

- 기존의 군집화 방법
 - 각 노드가 하나의 군집에 속하도록 한다 (partition).
 - 관찰: **한 노드는 여러 개의 군집에 속할 수 있다** (Nature'05).
 - 예: 사람들은 가족, 학교, 직장 등 여러 그룹에 속한다.
- 링크 군집화 (PRE'10, Nature'11)
 - 노드는 둘 이상의 속성을 쉽게 가지는 반면, 링크는 하나 이하로 가지는 경우가 많다.
 - 예: 4와 2는 직장 동료, 4와 5는 학교 동창
 - **모든 링크에 대해 하나의 군집 배정**
 - 예: 4, 7번 노드는 중첩되는 노드들



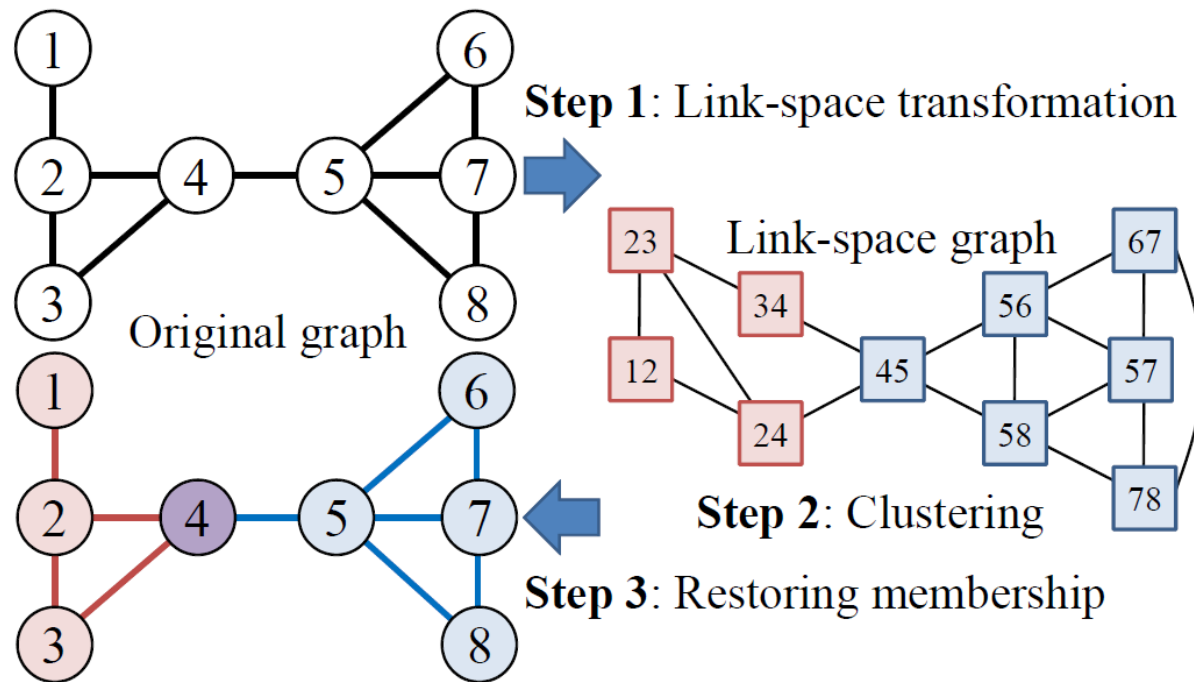
Structural Clustering

- 구조적 군집화 (KDD'07)
 - 어느 군집에 속하는 것도 합당하지 않은 노드는 배정하지 않음.
 - **활발한 노드(hub)**는 서로 다른 군집을 이어준다.
 - **동떨어진 노드(outlier)**는 어느 군집과도 유사하지 않다.



Link-Space Transformation

- 링크-공간 변환 (Lim et al. '13)
 - 구조: 링크와 링크의 연결 관계로 구성
 - 유사도: 링크-링크 유사도 측도를 정의하여 사용
 - 변환된 공간에서 구조적 군집화 알고리즘 적용



Experiments

- 합성 네트워크: LFR benchmarks (5K)
- 실세계 네트워크: DBLP (1M), Amazon (0.3M)
- 평가 척도
 - NMI, modularity for overlapping clustering, coverage 등

