

Structural Recommendations in Networks

10.4,5

이마태

목차

1. Link Prediction

1-1) 개요

1-2) Likelihood Measures

1-3) 추천 시스템에서의 활용

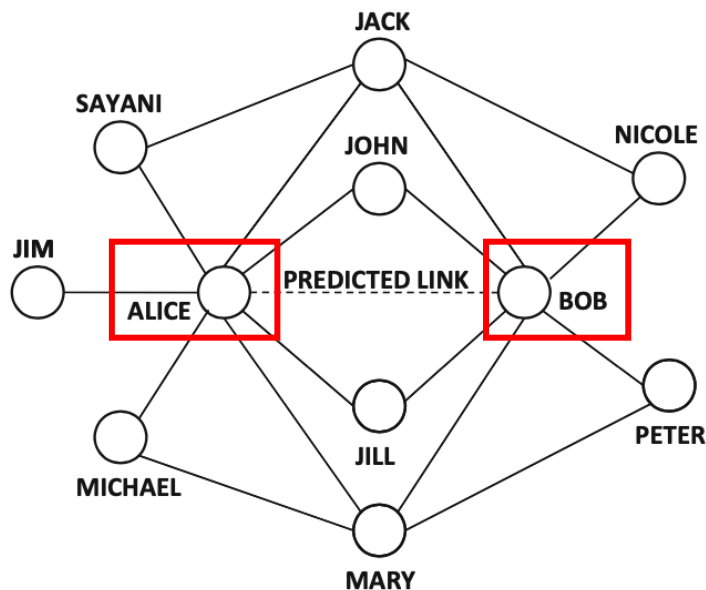
2. Social Influence Analysis와 바이럴 마케팅

1. Link Prediction: 개요

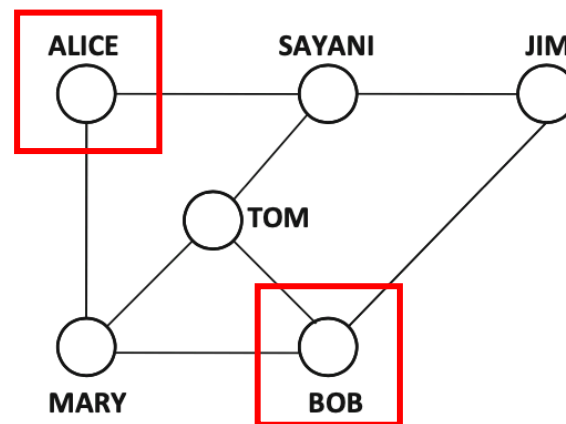
- 페이스북과 같은 SNS에서는 친구를 추천해 줌으로써 네트워크를 확장시킴
- ‘잠재적 친구’ 추천 → Potential Link
 - 아직 맺어지지 않은 관계의 Future Link 예측
- 노드 간의 관계를 측정하여, **Likelihood**가 높은 친구 추천

1. Link Prediction: Neighborhood-Based Measures

- 노드 간에 **공통적인 이웃의 개수**를 측정



(a) Many common neighbors between Alice and Bob



(b) Many indirect connections between Alice and Bob

(a): Alice와 Bob은 4명의 공통 이웃을 가짐

(b): Alice와 Bob 사이에 공통 이웃은 Mary 뿐

1. Link Prediction: Neighborhood–Based Measures

1) Common Neighbor Measure

- 각 노드의 이웃 집합의 교집합의 크기. 즉 **공통 이웃 개수**

$$CommonNeighbors(i, j) = |S_i \cap S_j|$$

S_i : node i 의 neighbor set

S_j : node j 의 neighbor set

- **약점**: 절대적인 개수 이용 \rightarrow 전체 네트워크에서 해당 숫자의 상대적인 중요도를 반영하지 못함
 - Alice와 Bob의 node degree(이웃 개수)가 작지만, 이웃이 아주 많을 경우 4개의 공통 이웃은 우연일 수 있음

1. Link Prediction: Neighborhood–Based Measures

2) Jaccard Measure

- 이웃 집합의 상대적인 크기. **교집합/합집합**

$$JaccardPredict(i, j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|}$$

S_i : node i 의 neighbor set

S_j : node j 의 neighbor set

- Alice와 Bob의 이웃 개수가 많아지면, *JaccardPredict* 는 작아지게 됨
- **장점**: node degree에 따라 상대적인 크기가 조절됨
- **약점**: intermediate neighbor를 반영하지 못함
 - Alice와 Bob의 이웃은 node degree가 아주 큰 인기 있는 사람일 수 있음
 - Ex) 유명 연예인은 많은 사람들이 팔로우하고 있을 것

1. Link Prediction: Neighborhood–Based Measures

3) Adamic–Adar Measure

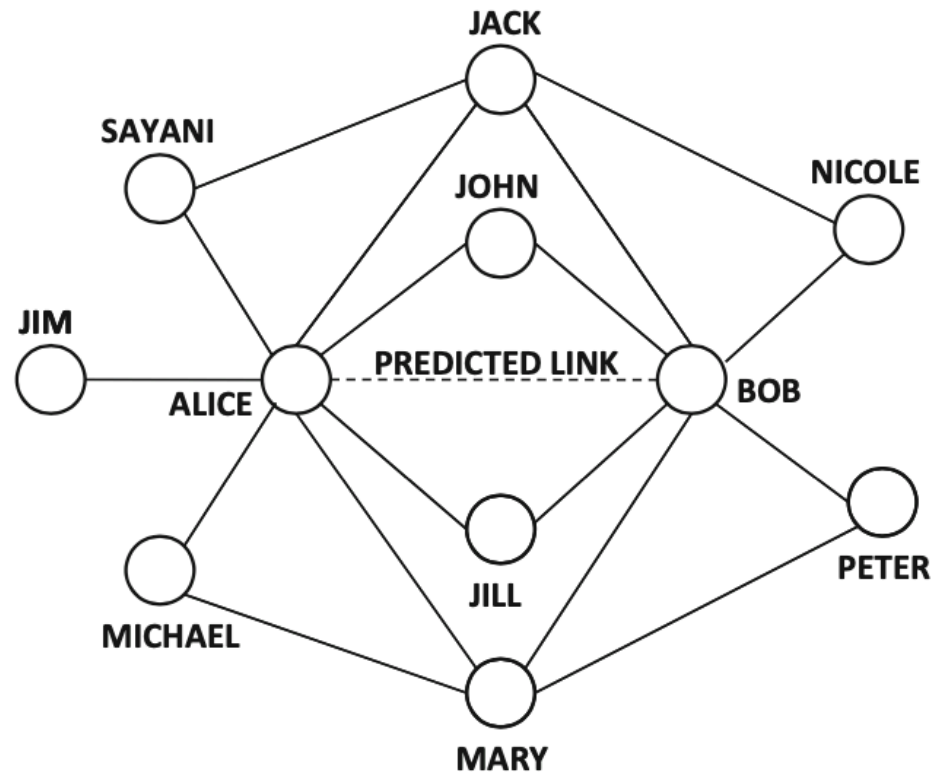
- 각 이웃들의 weight 계산

$$AdamicAdar(i, j) = \sum_{k \in S_i \cap S_j} \frac{1}{\log(|S_k|)}$$

$|S_k|$: node k의 node degree

- 이웃의 weight: $\log(\text{node degree})$ 의 역수
 - 덜 유명한(연예인 아닌 진짜 친구) 이웃일수록 중요도가 커짐

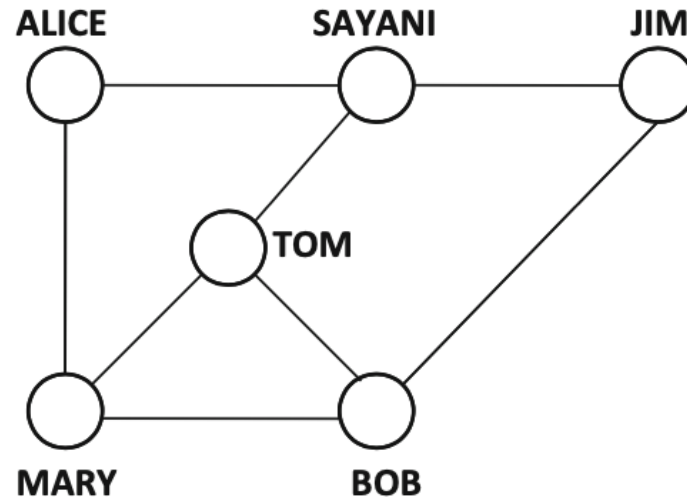
1. Link Prediction: Neighborhood-Based Measures



- $CommonNeighbors(Alice, Bob) = 4$
- $JaccardPredict(Alice, Bob) = 4/9$
- $AdamicAdar(Alice, Bob)$
$$= \frac{1}{\log(4)} + \frac{1}{\log(2)} + \frac{1}{\log(2)} + \frac{1}{\log(4)} = \frac{3}{\log(2)}$$

1. Link Prediction: Katz Measures

- 이웃 기반 방법은 robust 하지만, 노드 간에 공통된 이웃이 적을 경우에는 효과적이지 않음



(b) Many indirect connections
between Alice and Bob

- Alice–Bob, Alice–Jim 사이에는 **단 하나의 공통 이웃** 존재
- 두 관계의 prediction strength를 구별하기 어려움

1. Link Prediction: Katz Measures

- Katz Measure
 - Walk-based measure
 - 관계의 길이를 이용

$$Katz(i, j) = \sum_{t=1}^{\infty} \beta^t \cdot n_{ij}^{(t)}$$

$n_{ij}^{(t)}$: 길이가 t인 walk 개수

β^t : 길이가 긴 walk의 중요도를 낮추는 discount factor (< 1)

- $Katz(i, j)$ 가 수렴하기 위해서는 β 값이 작아야 함

1. Link Prediction: Katz Measures

- Katz Measure: $n \times n$ 행렬로 확장

$$K = \sum_{i=1}^{\infty} (\beta A)^i = (I - \beta A)^{-1} - I$$

K : Katz coefficient matrix

A : adjacency matrix of an undirected network

- β 는 A 에서 가장 큰 eigenvalue의 역수보다 작아야 함

adjacency matrix

- Link 있으면 1, 없으면 0

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

→ 멀리 떨어진 이웃들의 관계를 고려하여 이웃 간의 Likelihood를 측정할 수 있음

1. Link Prediction: 분류 문제에 활용

- 각 Link(Label)의 존재 유무에 따른 Binary Classification 문제로 볼 수 있음
- 각 행(=각 node)에서의 Features
 - 다양한 Likelihood Measures (이웃 기반, Katz 등)
 - Node-degree
- Feature를 이용하여 관계 예측
 - Positive pair(link 있는 관계)는 모두 사용
 - Negative pair(link 없는 pair)가 많기 때문에 sample 사용

1. Link Prediction: 분류 문제에 활용

1) Training Phase

- Link 있는 pair당 하나의 record
- Link 없는 pair들 중 sample record
- Feature: Likelihood Measures + Node Degree
- Label: Link 여부

2) Testing Phase

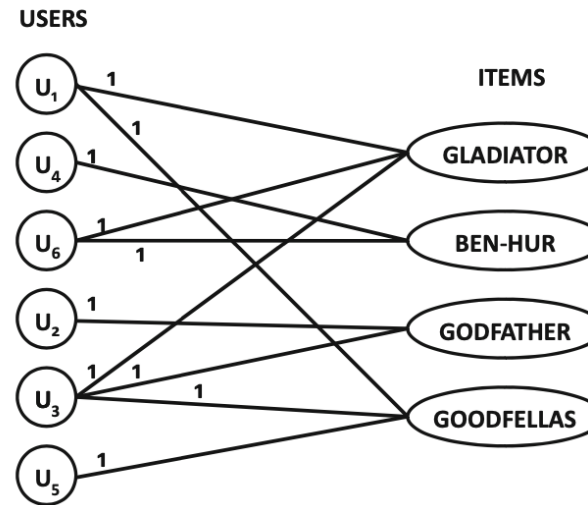
- Test node pair를 Multidimensional Record로 변환
- 로지스틱 회귀와 같은 분류 모델 사용

1. Link Prediction과 Collaborative Filtering

- Link Prediction은 CF에서 unary rating/implicit feedback 문제와 유사
 - Unary rating: *좋아요* 만 있는 시스템

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS
U ₁	1			1
U ₂		1		
U ₃	1	1		1
U ₄			1	
U ₅				1
U ₆	1		1	

(a) Unary ratings matrix



(b) Unsigned link prediction

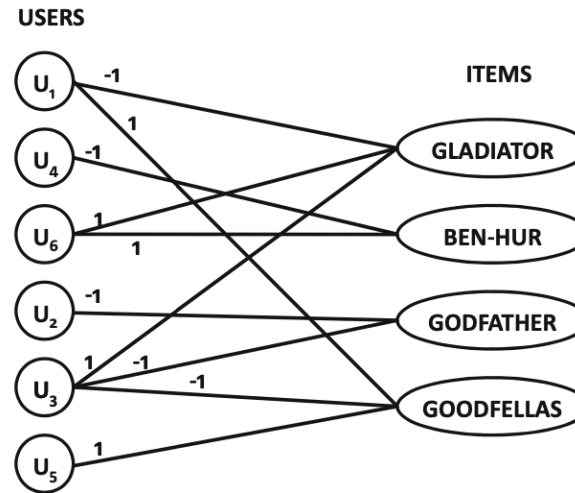
- 노드 간의 Likelihood 예측값 = 사용자가 item을 얼마나 좋아할지 예측값
- 유저 기준으로 보면, 추천해 줄 top-k 아이템 정할 수 있음
- 아이템 기준으로 보면, 특정 아이템을 홍보할 top-k 유저 집단을 정할 수 있음

1. Link Prediction과 Collaborative Filtering

- Binary rating은 Link의 특성이 있는 Signed Network
 - Binary rating: *좋아요/싫어요* 있는 시스템

	GLADIATOR	GODFATHER	BEN-HUR	GOODFELLAS
U ₁	-1			1
U ₂		-1		
U ₃	1	-1		-1
U ₄			-1	
U ₅				1
U ₆	1		1	

(c) Binary ratings matrix

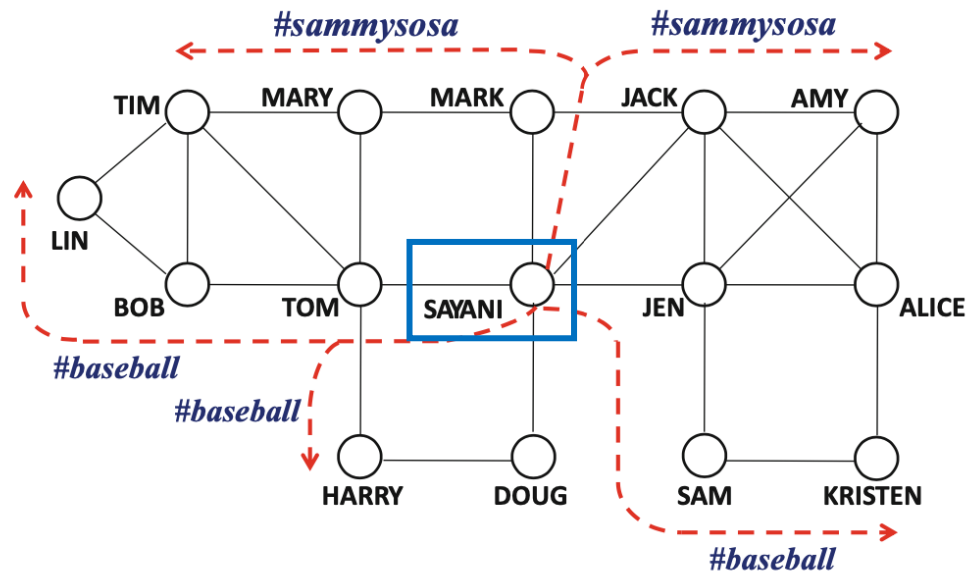


(d) Signed link prediction

- 노드 간의 Likelihood 예측값 = 유저가 item을 얼마나 좋아할지/얼마나 싫어할지 예측값
- Link의 특성(+1/-1)이 있기 때문에, 앞서 설명한 내용과는 조금 다름

2. Social Network Analysis와 바이럴 마케팅

- SNS에서 각 사람들의 영향력이 다름
 - 배우가 트윗 → 팔로워들에게 노출 → 팔로워들의 리트윗
 - 이런 과정을 통해 정보가 퍼지게 됨
- Propagation process



→ Sayani: 야구 관련된 글에서 영향력이 큼

감사합니다