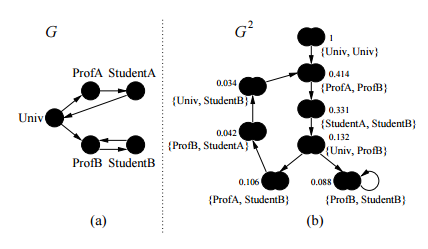
****

****

**SimRank: a measure of structural-context similarity ---** Glen Jeh and Jennifer Widom

Object 간 유사성을 측정하는 문제는 다양한 분야에서 일어난다. 그리고 documents에서 text를 매칭하거나 item set 간 overlap을 계산하는 domain-specific 방법들이 개발되어 왔다. 이 논문에서는 object와 object간 관계가 있는 어느 영역에서도, object가 발생된 structural context의 유사성을 측정하기 위해 적용할 수 있는 보완적인 방법을 제시하고 있다(다른 object들과의 관계를 기반으로). 논문에서는 두 오브젝트가 동일한 오브젝트들과 관련되어 있다면 서로 비슷하다고 말한다. 이렇게 간단한 그래프 이론 모델에 기반을 둔 similarity measure가 SimRank이다. 주어진 영역에서, SimRank는 다른 영역의 similarity measure와 결합할 수 있다. 그리고 이 논문에서는, SimRank 스코어를 효율적으로 계산하는 방식을 소개한다.

웹에서는 페이지 간 하이퍼링크가 있을 때 두 페이지가 관련되어 있다고 말한다. 논문-인용도 비슷한 예가 될 수 있다. 추천 시스템의 경우에는, 사용자-아이템 간 관계에서 아이템에 대한 사용자의 선호도가 중요 지표가 된다. 그리고 이런 영역들은 objects를 표현하는 노드와, 관계를 표현하는 엣지로 모델링된다. 논문에서는 이러한 데이터 셋으로부터 만들어진 그래프를 분석해서 노드가 나타나는 structural context를 기초로 해, 노드 간 similarity score를 계산한다. 알고리즘 intuition을 살펴보면, 오브젝트 a와 b는 만약 그들이 비슷한 오브젝트인 c와 d와 관련되어 있다면 비슷한 오브젝트일 것이다.



그래프 G에서는 Univ에 하이퍼링크 되어 있는 ProfA와 ProfB가 비슷할 것이라고 추론할 수 있다. 또한 이 사실로부터 ProfA와 ProfB와 연결된 Student A와 B도 유사할 것이라고 추론할 수 있다. 그리고는 G2을 계산해서 노드를 G에서의 노드들의 쌍으로 표현하여 스코어를 표시했다. G2을 반복적으로 돌리면 node-pairs에 대한 SimRank 스코어를 얻을 수 있다. 그리고 이 스코어는 두 노드 간 유사성을 확인할 수 있는 지표가 된다. 스코어는 노드와 그 이웃들 간의 flowing으로 생각되며 각 iteration에 의한 스코어는 엣지를 따라 퍼지며 스코어가 일정한 값에 수렴할 때까지 반복이 계속된다.

Contributions:

1. 임의의 그래프에 대해서 SimRank similarity scoring를 정의한다. 그리고는 SimRank 스코어를 계산하는 알고리즘을 소개한다.
2. SimRank의 사용과 계산을 위한 수학적 통찰력을 주는 그래프 이론 모델을 제시한다.
3. 2개의 read data sets을 이용해 SimRank의 effectiveness와 feasibility 실험. 검증