

Week5-1

1. 학습정리

1. NetworkX

1. 그래프 분석 파이썬 라이브러리

2. 랜덤그래프

1. 확률적 과정을 통해 생성한 그래프

2. 에르되스-레니 랜덤 그래프

1. 두 정점 사이 간선 존재 여부는 동일한 확률 분포에 의해 결정

에르되스-레니 랜덤 그래프(Erdős-Rényi Random Graph)

임의의 두 정점 사이에 간선이 존재하는지 여부는 동일한 확률 분포에 의해 결정됩니다

에르되스-레니 랜덤그래프 $G(n, p)$ 는

✓ n 개의 정점을 가집니다

✓ 임의의 두 개의 정점 사이에 간선이 존재할 확률은 p 입니다

✓ 정점 간의 연결은 서로 독립적(Independent)입니다

3. 경로, 거리, 지름

1. 길이

1. 경로의 간선 수

2. 거리

1. 최단 경로의 길이

3. 지름

1. 정점 간 거리의 최댓값

4. 작은 세상 효과

1. 여섯 단계 분리 실험

1. 임의의 두 사람을 골랐을 때, 몇 단계의 지인을 거쳐 연결 되었는가

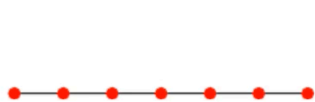
2. 평균적으로 6단계

2. 작은세상효과

1. 임의의 두 정점 사이의 거리가 작음

3. 랜덤 그래프에도 존재

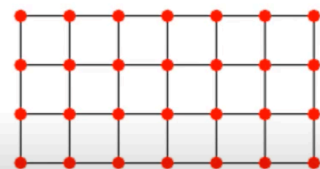
4. 체인, 사이클, 격자 그래프는 작은 세상 효과 x



체인
(Chain Graph)



사이클 그래프
(Cycle Graph)



격자 그래프
(Grid Graph)

5. 정점의 연결성

1. 정점에 연결된 간선 수

6. 연결성의 두터운 꼬리 분포

1. 연결성이 높은 사람 소수

2. 랜덤 그래프의 연결성 분포는 정규분포와 유사

1. 이 경우 허브가 없음

실제 그래프의 연결성 분포는 **두터운 꼬리(Heavy Tail)**를 갖습니다

즉, 연결성이 매우 높은 **허브(Hub)** 정점이 존재함을 의미합니다



7. 연결 요소

1. 연결요소에 속하는 정점들은 경로로 연결 가능

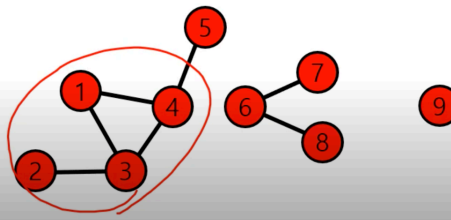
연결 요소(Connected Component)는 다음 조건들을 만족하는 정점들의 집합을 의미합니다

- (1) 연결 요소에 속하는 정점들은 경로로 연결될 수 있습니다
- (2) (1)의 조건을 만족하면서 정점을 추가할 수 없습니다

예시 그래프에는 3개의 연결 요소가 존재합니다

{1,2,3,4,5}, {6,7,8}, {9}

{1,2,3,4}는 조건 (2)를 위배합니다
{6,7,8,9}는 조건 (1)을 위배합니다



- 거대 연결 요소

- 실제 그래프는 거대연결요소 존재
 - 99.9%의 정점이 하나의 거대 연결 요소에 포함됨
- 랜덤 그래프에서도 거대 연결 요소 존재
 - 정점들의 평균 연결성이 1보다 클 때 존재

- 군집

- 집합에 속하는 정점 사이 많은 간선 존재
- 속하지 않은 정점 사이에는 적은 간선

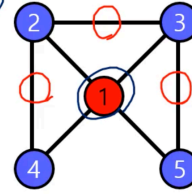
- 지역적 군집 계수

- 정점에 연결된 이웃 쌍 중 간선으로 연결된 것의 비율

지역적 군집 계수(Local Clustering Coefficient)는 한 정점에서 군집의 형성 정도를 측정합니다

정점 i 의 지역적 군집 계수는 정점 i 의 이웃 쌍 중 간선으로 직접 연결된 것의 비율을 의미합니다
정점 i 의 지역적 군집 계수를 C_i 로 표현합니다

예시 그래프를 살펴봅시다
정점 1의 이웃은 4개이며, 총 6개의 이웃 쌍이 존재합니다
그 중 3개의 쌍이 간선으로 직접 연결 되어 있습니다
따라서, $C_1 = \frac{3}{6} = 0.5$



- 전역 군집 계수
- 지역적 군집 계수의 평균
-

2. 피어세션

1. 캐글 리뷰

1. 앙상블 기법
2. 훈련 데이터 수정

2. 강의 리뷰

3. 금요일 퍼더 퀘스천

1. BERT의 Masked Language Model의 단점은 무엇이 있을까요? 사람이 실제로 언어를 배우는 방식과의 차이를 생각해보며 떠올려봅시다.
 1. 이 전의 단어로 부터

4. 알고리즘 문제 풀이

1. <https://leetcode.com/problems/132-pattern/>