

# EE412 Foundation of Big Data Analytics, Fall 2021

## HW1

Name: 노현섭

Student ID: 20190220

Discussion Group (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

[TOKEN=2]

### Answer to Problem 1

Explanation about algorithm:

우선 tab 및 ','으로 나누어 실제 친구 pair 를 묶어 주었다. 이때 (친구 1, 친구 2, 0)과 같이 0 이라는 value 를 줬으며, 중복을 없애기 위해 for 문을 돌리며 친구 1 < 친구 2 인 경우만 append 해주었다. 같은 방식으로 어느 한 사람의 친구 리스트 안에서도 pair 를 묶어 주었다.(mutual friends) 이때도 앞서와 같이 (친구 1, 친구 2, 1)으로, value 1 을 줬으며 친구 1 < 친구 2 와 같이 정렬되도록 했다. 하지만 여기서는 중복을 제외하지 않았다. (다른 친구에서도 mutual friends 관계로 나타난다면, 함께 아는 친구를 2 로 계산해야 하므로). 이후 counts 라는 함수를 통해 key 값, 즉 (친구 1, 친구 2) pair 를 이용해 value 가 모두 1 인 경우만 골라서 합하며 reducing 을 해주었다. 이때 mutual friends 리스트 중, 실제 친구인 경우는 모두 사라지고 친구 사이가 아닌 pair 들의 함께 아는 친구 명수가 value 로 계산된다. 이후 value 가 0 인 경우는 모두 지우고 sorting 을 해주면 top10 리스트가 나온다.

Elapsed time: 100.76 seconds

### Answer to Problem 2

- (a) Hash table 을 위한 memory 를 무시하면, triangular matrix 에서 크기는  $N \times N$  이고 upper-triangle 이다. 따라서 pair 의 개수는  $N(N-1)/2$  이고, 각 pair 에 integer 를 하나씩 저장하기 위해 4 를 곱해야 한다. 따라서 총  $2N(N-1)$ 이 필요하다.
- item, item, count triple 을 사용하면, triple 의 수는 item pair 의 수와 같다. frequent pair 100 만개, 그리고 frequent item 으로 이루어진 M 개를 만들기 위해  $10^6 + M$  이 필요하며 각 triple 에서 integer 3 개가 저장되므로  $(10^6 + M) \times 3 \times 4 = 12(10^6 + M)$ 이 필요하다.

이때, 최종값은  $\min\{2N*(N-1), 12*(10^6+M)\}$ 이다.

Elapsed time: **43.77 seconds**

### Answer to Problem 3

- (a) 1.  $p \rightarrow p^2 \rightarrow 1-(1-p^2)^3$   
2.  $p \rightarrow 1-(1-p)^3 \rightarrow (1-(1-p)^3)^2$   
3.  $p \rightarrow p^2 \rightarrow 1-(1-p^2)^2 \rightarrow (1-(1-p^2)^2)^2$   
4.  $p \rightarrow 1-(1-p)^2 \rightarrow (1-(1-p)^2)^2 \rightarrow 1-(1-(1-(1-p)^2)^2)^2$

Elapsed time: **533.84 seconds**