Big Data Systems
Project #3 – Graph Databases/Neo4j
Due Date: June 7th, 2020

In this project you will be given a dataset and you will represent it in the relational model, in the graph model, run some queries and measure performance.

1. Go to Stanford Large Network Dataset Collection (https://snap.stanford.edu/data/#socnets)
2. Use the Pokec dataset in Social Networks category (https://snap.stanford.edu/data/soc-Pokec.html)
3. For each user keep user_id, age and gender.
4. Load data in MySQL and Neo4j
5. Write queries in SQL and Cypher for the following queries and benchmark both:

    - For each user, count his/her friends
    - For each user, count his/her friends of friends
    - For each user, count his/her friends that are over 30
    - For each male user, count how many male and female friends he is having