

ITP20002-01 Discrete Mathematics

Programming Assignment 3 (revised version)

24 Nov 2018

PA3. Twitter Network Analysis

- Tasks
 - Find interesting properties of the Twitter Follower relations by writing C programs
 - <https://github.com/hongshin/DiscreteMath/blob/master/assignments/twitter-sampled.txt>
 - answer to 5 given questions
 - find your own observations of the given data
- Setting
 - Due date: **11:59PM, 4 Dec (Tue)**
 - Evaluation
 - report: 60%
 - program artifact: 40%
 - No late submission will be accepted

Data



- Use Twitter Social Circle Dataset tby SNAP@Stanford
 - The original is on 81306 users having 1,768,149 Follows
 - <https://snap.stanford.edu/data>
 - The data is extracted obtained to understand which features characterize user's circles in Twitter, Facebook and Google+
 - Use **a sampled data with 5128 users and 21270 Follows cases**
 - <https://github.com/hongshin/DiscreteMath/blob/master/assignments/twitter-sampled.txt>
- Data format
 - IDs are positive integers and they are deidentified
 - "X Y" means that user X follows user Y in Twitter
 - Not always, XY implies Y X

Questions (1/2)

1. How many other users a user follows? How many followers a user has?
2. User X's twitter page shows links to other twitter accounts that X follows (i.e., Following). You can click one of these links to make a transition.
The distance from user X to user Y is the minimum number of clicks from X to Y. What is the maximum distance between two reachable users? And who are they?
3. Let's call two users X and Y are *connected* iff X follows Y, or Y follows X, or there is another user Z who is connected with X and Y at the same time.

Is every connected with another one? If it is not, how they are?
4. When X and Y mutually follow, we call them Friends. Or, X and Z are friends when X and Y are friends and Y and Z are friends. And, obviously, X is a friend of itself.
How many friend partitions in the given data? Would you describe them?

Submission

- Submit a report together with the program artifact
- Your report must be a Power Point Slide up to 10 pages (slide)
 - visualize your answer if it's possible
 - for each question, write an answer and explain how you find the answer with the programming
 - try best to write precisely and concisely
- Your program artifacts should be something executable and the result must be reproducible by TA