Analysis and Modeling of Social and Information Networks

CIS 4524/5524, Spring 2023

Assignment 1, due January 26 by 5:00 pm on Canvas

*Please write your name and TUID at the top of your **CANVAS** submission.

Homework Policies (applicable for all assignments):

1. You are required to do the homework problems in order to pass.
2. Understandability of the solution is as desired as correctness.
3. Penalty for late homework assignments submissions is 20% per day. So, do it on time.
4. Solutions are expected to be your own work. Group work is **not allowed** unless explicitly approved for a particular problem. If you obtained a hint with help (e.g., through library work, etc.) acknowledge your source, and write up the solution on your own. Plagiarism and other anti-intellectual behavior will be dealt with severely.

Download the Homework 1 from the class homework folder. This table is generated according to the background survey for this course. In this table each raw represents a response of one of you to a survey asked as the Homework 0, the column values are 1 if you have some background experience on (1) Data Mining (CIS 4523/5523) or Machine Learning); (2) Python or R programming; (3) Graphs or Statistics.

Problem 1. [Visualizing a Multilayer Network] In this exercise your job is to visualize a 3-layer course network, where nodes represent students, and each layer corresponds to a single topic network representing one type of relationship among students taking this course (e.g. in layer 1 nodes representing two students should be connected if both students took a Data Mining/Machine Learning course). Note: you need to convert row data to graph format edge list or matrix, .csv or .txt format should be fine. Make a screenshot of your visualized networks.

Problem 2. [Visualizing a Weighted Network] Visualize a network obtained by projecting the 3-layer network from Problem 1 to a single-layer weighted network, where two students are linked by a ***weighted*** edge representing the number of topics both students took.

Problem 3. [Visualization of a Bipartite Network] Visualize a bipartite students-topics network where an edge between a student node and a topic node exists if and only if this student has taken a course on that topic.

Problem 4. [Computing Global Network Properties] For each layer of the 3-layer network constructed in Problem 1 compute and explain the following global network properties:

a) the size and diameter of the network largest connected component
b) degree distribution (you can report average degree distribution or plot degree distribution histogram)
c) average path length
d) average clustering coefficient

Repeat this for the network constructed in Problem 2. Compare between properties of each layer and aggregated network in Problem 2.

Impotent Notes:

- Before importing the student network, you must transform the row data into a supported graph format. Check how to convert row data to graph input in the library/software you plan to use.
- Visualizing a multilayer network is different from a single-layer network. To visualize a Multilayer network, make sure to use a supporting library for example, Pymnet and Multinetx (check Syllabus-Software section for more libraries)
- It is highly recommended to use Python library NetworkX. Check sections 'Creating a graph' and 'Drawing graphs' to see how to visualize a graph. You can also use Gephi, a platform-free software for graph visualization and analysis, which you can download from https://gephi.org/ (follow the quick start guide) to import the student network and visualize it.
- All metrics can be calculated using the Python library NetworkX, but you can use any freely available packages to compute these properties or develop your code.