CSE416A

HOMEWORK 0

M. Neumann

Due: FRI 15 SEPT 2023 (11:59PM)

SUBMISSION INSTRUCTIONS

- **use of AI to create solutions**
  - The use of AI tools (such as (but not limited to) ChatGPT, Microsoft Bing, or Google Bard) to generate solutions (code or written answers) for labs, worksheets, or hw assignments is **not** permitted unless explicitly stated otherwise on the assignment.
  - We will use AI detection tools to identify work that may be in violation of course policies. We may request that students show evidence of their work process (such as notes or earlier versions of final work) should we suspect inappropriate use of AI tools.
- **written work**
  - needs to be submitted electronically in *pdf format* via GRADESCOPE
  - start every problem on a *new page*
  - we prefer *typed submissions*, e.g., using LaTeX (if we cannot read your handwriting we cannot give you credit)
- **code (Jupyter notebook)**
  - needs to be submitted in form of a submission to the corresponding GRADESCOPE programming assignment (instructions can be found on the course webpage)
  - make sure to change the *file name(s)* including your name(s) and follow the *formatting instructions* provided in the notebook (otherwise we cannot grade your submission)
- **group work** (up to 2 students – not for HW0)
  - make a **goup submission** via GRADESCOPE (one submission per team) for both written work and code submission

PIAZZA

We use Piazza for all course and homework related announcements. Ask **all questions** on Piazza using the appropriate tags.

GRADING RESULTS AND REGRADES

Grades will be uploaded to Canvas and detailed grading comments will be provided via GRADESCOPE . You will be notified via GRADESCOPE when the grades are published. All regrade requests need to made via GRADESCOPE **within one week** of this announcement.

PROBLEM 1: Numpy (30%)

See `hw0_YourName.ipynb` for instructions.

PROBLEM 2: Introduction to NetworkX and Simple Network Visualization (20%)

See `hw0_YourName.ipynb` for instructions.

PROBLEM 3: Construct your own Wikipedia Network (20%)

cf. **[CNA]** Chapter 5:
See `hw0_YourName.ipynb` for instructions. There is no code submission required for this problem. All answers should be included in the written/pdf submission.

3.1 Create your own network from Wikipedia using a seed term that is related to *Network Analysis* or *Network Science* (other than the one that is provided in the notebook as an example). Report the basic network statistics, number of nodes and edges (after preprocessing) and the 25 nodes with the most in-links.

3.2 What do you think number of in-links measures in the context of our network application?

3.3 Of the concepts retrieved above, which ones make sense intuitively (and why) and which ones are surprising to you (and why)? What are the two concepts/terms that you haven't heard of before or know the least about? Go to Wikipedia, read-up on them and briefly summarize them in 1-2 sentences each.

3.4 Find the `.gexf` file and load it into Gephi. You will need to install Gephi from here: `https://gephi.org`. Now, gephi is not the most intuitive software to use. You will need to play with is for a bit to figure it out. Here is a tutorial that will be helpful: `https://www.youtube.com/watch?v=371n3Ye9vVo`. And/or read up on the introduction in **[CNA] Chapter 4**. Plot the network in a "nice" way and include the figure in your written/pdf submission.

HINTS/INSTRUCTIONS FOR "NICENESS":

- The nodes should be spread apart in a meaningful (or magic !?) way. Check and play with the Layout options under `Data Laboratory -> Layout`.
- Nodes and links should have different colors. Set this under `Data Laboratory -> Appearance`.
- Since we are interested in in-degrees set the node size and/or the node color to be proportional to the in-degree. You can find set this under `Data Laboratory -> Appearance -> Nodes` and then click on the **icons** (one is for color and one for size) in the top right. Then click on `Ranking`. Then have fun playing with the settings.
- The node/wiki page names should be visible and readable. Play with the label size in `Preview -> Node Labels`.

PROBLEM 4: Networks Everywhere (30%)

cf. **[NetSci]** 1.8 Homework

4.1 List three different networks and state the nodes and links for each of them.

4.2 Tell us about the network you are personally most interest in. Address the following questions:

- What are its nodes and links?
- How large is it?
- Can it be mapped out?
- Why do you care about it?

4.3 Do you think every network relevant to network science can be mapped out? If your answer is yes, explain your reasoning. If your answer is no, describe an example of a network that cannot be mapped out.

4.4 In your view, in what area could network science have the biggest impact in the next decade? Explain your answer.