Due Date: October 16, 2022 (11:59 PM)

Total points: -10% (i.e., if you don't submit it, you will lose 10% of the project)

Purpose

The project helps you to 1) practice what you have learned and what you'll learn in the course, 2) engage in a small-scale research project, and 3) shine your CV.

1 Topic

You are free to pick up any topic. However, it should meet the following criteria

- **Relevancy.** The topic should be related to (social) network analysis. That is, it should involve a network. Other types of networks other than social networks are acceptable, e.g., biological networks, web networks, etc. For example, a product recommendation that utilizes the customer-product bipartite network is considered relevant. However, a topic like sentiment analysis of tweets is unacceptable unless you approach it from a graph perspective.

- **Scale.** The project should be proportionate to a class project and involves doing research. For instance, simply reproducing assignment questions (e.g., calculating centrality in a network) is not acceptable. The number of members in a team will be considered for the scale.

1.1 Type of Project

- **Dataset.** You can construct a new (social) network dataset. For instance, you can construct a Twitter network starting from your Twitter account (i.e., expanding followers/followees). For the database project: 1) you have to do the collection yourself (obviously), 2) collect a considerable number of users (i.e., a relatively large dataset), and 3) perform a deep exploratory data analysis of the dataset e.g., full statistics, network measures, etc–See Dataset Papers in this conference for example datasets and performed analyses.

- **Implementation.** For this project, you need to specify a topic and implement it on a network(s). The expectation is not to propose a novel methodology (if you do, though, it would be great!). However, for a selected methodology (ies), you need to understand it well, implement it correctly, get the results and analyze them deeply.

1.2 Resources

The following conferences and journals are great resources for selecting your project topic. Read recent their recent proceedings.

- Cyberpsychology, Behavior, and Social Networking
- International Conference on Web and Social Media (ICWSM)
- International Conference on Web Search and Data Mining (WSDM)
- The Web Conference (formerly WWW)
- Social Networks (An International Journal of Structural Analysis)
- International Conference on Advances in Social Networks Analysis and Mining
- Social Network Analysis and Mining
- International Conference on Social Media and Society

- Computational Social Networks

- Other AI-related conferences: KDD, ICDM, CIKM, NeurIPS, ICML, ICLR, JMLR, TKDE, AAAI, IJCAI, and SDM.

1.3   Datasets

You can always construct your own dataset (for implementation projects). In addition, Stanford Large Network Dataset Collection (SNAP) include a large set of networks.

1.4   Suggested topics

- **Community Analysis.** Communities are always interesting! SNAP website contains several datasets with ground truth communities, which you can use to evaluate your community detection algorithm(s)– See these surveys here and here.

- **Homophily.** The presence of homophily in social networks is an exciting topic (the tendency of "similar" people to connect to one another). In this project, try to discover on what basis users in a social network create homophilous links, e.g., politics, interests, race, gender, etc. One interesting project is to construct a Twitter network, manually or automatically determine users' attributes (e.g., job, interests, hashtags, etc.), and then determine and analyze homophily.

- **Evolving Social Networks.** Social networks are not constant due to, e.g., users joining, users leaving, added links, broken links, etc. It is interesting to analyze how these networks *evolve*. For instance, how do network measures (e.g., centrality) change? Do two users become more similar? Why links are broken (e.g., unfollow on Twitter)? How communities change? Again, you can use temporal networks in SNAP or build your own dataset (e.g., perform Twitter data collection in a few snapshots).

- **Fundamental Graph Problems.** In the course, we have introduced many fundamental graph properties and algorithms. Theoretical and/or empirical assessments of fundamental graph problems are welcome. For instance, finding a specific subgraph of size k (known as *graphlet*) in large networks is an interesting and useful topic–See this paper.

- **Network Visualization.** As you might have realized by now, existing libraries such as Networks are inadequate to visualize networks specially large networks. Visualizing (social) networks is an active line of research. For this project, you visualization should offer new insights and interpretable explanations about the network.

2   Workflow

Below is the suggested workflow for completing your project.

- **Planning.** Make sure you understand the problem and plan it well. Don't be ambitious or too modest. Consider the scale.

- **Execution.** If you work in a team, coordinate your efforts and split the tasks. Try to approach the execution in a modular and step-by-step way.

- **Analyze and revise.** Perform the project in a recursive manner where you obtain the results, analyze them, and, if needed, revise/improve your methods/code/plan.

- **Writing.** Consider this project as a research project that you want to publish. So, spend enough time preparing the report. After all, the project report is the main source of grading!

Deliverable

Use the template provided on Canvas and submit the project proposal. Each project can have up to 3 members. Individual projects are fine.