Homework 4

Name: Jenna Goodrich
Student ID: 5603326

Name: Fahim Ahmed
Student ID: 4223466

- Reasoning and work must be shown to gain partial/full credit
- Please include the cover-page on your homework PDF with your name and student ID. Failure of doing so is considered bad citizenship.

1. (1–4 points) **Final Project Planning** In this question we are going to evaluate your organization skills in planning for the Final Project. These plans are preliminary and may change later, and more questions will follow that will help you sharpen your plan progressively as the course progresses.

   Instructions on the final project

   - The project will be a team effort, and each team will consist of 2 students.
   - Each project will focus on one particular kind of network and of network data which must be chosen by the team.
   - The team must formulate the goal of analyzing the network data chosen.
   - The team must choose an approach for their analyzing the data.

   To answer this question you are asked to:

   (a) Identify your team.
   Fahim Ahmed and Jenna Goodrich.

   (b) Write a plan on how to divide the tasks among team members.
   Both Fahim and Jenna will participate in the selection of the dataset. Jenna will handle the initial creation of the network/converting the data into a graphical format, and do any associated data preprocessing. Fahim and Jenna will split the analysis portion, each implementing and reporting on specific metrics. Assuming there will be a final report, the team will collaborate using Overleaf to fill in sections each person worked on. Both individuals will check the other persons work. The team will meet regularly to ensure prompt completion of the project and to work through any issues either person runs into. Specific task allocation will depend on what types of analysis we decide on.

   (c) Identify the network and analysis that you are interested in carrying out. Real-world networks cases are strongly recommended and will boost your grade. In this context, describe the type of data you will need.
   The network we would like to work with is a specialized social network (also a bipartite network) built from a dataset consisting of book reviews by users, in which nodes are books and users, and edges are ratings from users to books. Additional links between users could be added based on similarities, but the initial network will only have edges between users and books. For analysis, there are a number of approaches we could pursue:

   - Centrality Measures - Identify influential books and users (degree, betweenness, closeness, and/or eigenvector centrality).
   - Clustering - group similar users and books.
   - Collaborative Filtering/Recommendation - recommend books to users (similarity metrics/correlation coefficients to measure similarity between users or between books, PageRank).
   - Link Prediction - predict future connections (common neighbor, Jaccard similarity).

   A possible end goal is to create a recommendation system to suggest books to users with a GNN.
   For this project, the type of data we will need is a collection of book reviews and users. Some possible datasets include: Book Rec Dataset, Book Crossing Reviews, and Amazon Book Reviews

At a later point each team will be asked to use models and inference methods introduced in Graph-based AI tools to analyze network signals.