



Course Project Specifications

Graph-Based Modelling of Social Networks

Sijjil Khan
Murtaza Ali Khokhar
Taha Hunaid Ali
Ashar Abbasi

Professor Dr. Faisal Alvi

CS 201 Data Structures II

Habib University

March 2024

Our project revolves around the study of graph-based datasets of social networks, communities and the attributes that contribute to their connections and hence extracting valuable insights and analytics from them.

We will be looking into finding an appropriate dataset from multiple online repositories such as the [Konect \(The Koblenz Network Collection\)](#), [Gephi Datasets](#), & [The Stanford Large Network Dataset Collection \(SNAP\)](#). This will act as our input for the project in either text form or .csv form. The input can be manipulated to represent nodes and edges of our graph.

The data structures we will be using are as follows:

- Graphs
 - The core structure for modelling social networks, where individuals (or entities) are represented as nodes (or vertices), and the relationships or interactions between them are edges. The graph structure used will be *edge list* as our edges have properties that are valuable.
- Trees
 - We will be looking to use Binary Search Trees for searching and accessing nodes on the graph. We will also be looking towards implementing Spanning Trees to apply minimal path algorithms.
- Hash Tables
 - Used for efficiently mapping, accessing, and storing node and edge attributes, like weights for weighted edges or labels for nodes. We will be using hash tables mainly for edge property mapping.

The functionalities of the project are as follows:

- Visualisation
 - Will be able to produce visuals of how the dataset is connected indicating clusters, “popular” nodes, and major attributes contributing to connections.
- Community Detection
 - Community detection and analysis will be a key functionality of the project which will produce and deliver insights into size, density and attributes regarding communities and clusters in the network.
- Centrality Measures
 - Implement algorithms to calculate various centrality measures such as Degree Centrality to identify the most active or popular individuals and Closeness Centrality to highlight those who can spread information efficiently (are connected more to other nodes)

The output of our project, as mentioned in the functionalities, will be an analytical visualisation of the input dataset highlighting important and interesting properties of the dataset.