Title - Social Media Network Analysis Application

Description - Networks Analysis of the connection of users on social media can give us great insight and have a decisive influence on events. We developed a Full-stack application for network analysis and data Visualization that enables us to predict results & analyze networks.

Backend Details and Features

1. Data processing and Scraping - Our analysis is centered around cluster analysis and we were required to store and scrap the social media data using python scrapping tools, the data was stored in MongoDB atlas cloud database.

2. Clustering & closeness inspection - We also implemented the Hierarchical k-means clustering algorithm to detect sub-groups or communities within the network. Moving a step ahead I also analyzed the closeness between clusters i.e determine if they can be merged into one social group. Clusters that are close to each other were merged, and clusters having at least one pair of users with a distance higher than some threshold are split into two.

3. Cluster closeness Analysis - Shortest path algorithms such as Floyd Warshall as written over the dynamic data to find the distance between 2 users or clusters and determine their closeness. i.e. the minimum number of profiles a user has to visit before the target user's profile can be reached. Certain utility controllers and services have developed that prepare adjacency matrix for the data and other related sub-tasks.

4. Metric & Centroidal Node Analysis - Each cluster formed will contain users ranked to a certain metric which could be a combination of the mean distance to reach other nodes (users) and minimum distance to reach any one of the nodes (users) in that cluster. This ranking will give the centroid of the cluster and will help to answer questions like who are important people/clusters in the network.

5. Backend API & Visualizations - A REST API was also required to establish communication with the database and servicing algorithm, and hence we used the Flask framework to implement the REST API, exposing endpoints for the front-end to establish communication. The plotting and visualization libraries had specific attribute requirements resulting in the development of a more complex service that served cluster.json files as a response to incoming AJAX requests.

6. DevOps & Hosting - Version control systems - GitHub, Cloud Database, and Heroku CI & CD played a crucial role in rapid project development and shipping. Automation build systems were set based upon commit branch was used for the production release of a specific feature.

Sachin duhan
2K17/MC/087

Tanay Singhania
2K17/MC/107

Adarsh Agarwal
2K17/MC/008