Brett Davidoff, Glen Dale Davis, Peter Gatica 2/18/2024

Project 1 proposal Data 620 – Web analytics

Dataset: <https://snap.stanford.edu/data/ego-Facebook.html>

Description of dataset: Our dataset consists of anonymized Facebook data that includes 4,039 nodes, 88,234 edges, a diameter of 8, an average clustering coefficient of 0.6055, and over 1.6 million triangles. Some of the anonymized categorical features attached to the user nodes are, education, hometown, and gender.

Our goal with this project is to analyze the Facebook dataset through centrality measures to identify influential users, understand information flow, and explore the impact of social ties on user behavior. One question we will try to answer by comparing degree centrality across categorical groups is whether certain political affiliations have “better” networks, i.e. networks through which targeted campaign ads and fundraising requests might travel faster.

**Methodology:**

* Data Preprocessing: Clean and prepare the Facebook dataset for analysis.
* Network Analysis: Use NetworkX to calculate degree, betweenness, closeness, and eigenvector centrality measures.
* Visualization: Graphically represent the network to identify key nodes and patterns of interaction.

**Expected Outcomes:** Insights into key influencers, information dissemination mechanisms, and the role of social connections in user engagement on Facebook. These findings could inform strategies for community management, targeted advertising, and enhancing user experience on social platforms.