Bishoy Sokkar
DATA 620 Week Assignment: Centrality Measures
9/19/2025

I found a great dataset for studying network centrality measures: the MUSAE Facebook Page-Page Network from Stanford's SNAP collection, available from here. This undirected graph, comprising 22,470 nodes (representing verified Facebook pages such as politicians, companies, or TV shows) and 171,002 edges indicating mutual "likes" between pages, includes a categorical variable for each node: the page type (e.g., "politician," "governmental," "company," "tvshow"). This attribute enables grouping nodes for comparative analysis, making the dataset ideal for exploring how centrality measures vary across different categories of social media entities. What makes this dataset perfect is that each node comes with a categorical label for the page type, such as "politician," "government," "company," or "tvshow." These labels let us group pages to compare how centrality varies across different types of entities.

To analyze this, we could look at degree centrality, which measures how many direct connections a page has relative to the total possible connections, giving us a sense of its influence in the network. By grouping pages by their type and comparing their average degree centrality, we might uncover interesting patterns. For example, politician pages could have a higher average degree centrality. This suggests that political pages are more interconnected, potentially amplifying their ability to spread messages virally, especially during events like elections. Conversely, TV show pages may rely on more isolated fan bases, predicting lower network-driven engagement. Such findings could guide social media strategies, indicating that targeting political networks may maximize outreach and influence.