CSE4081 – SOCIAL NETWORK ANALYSIS

Term Project – Structural & Community Analysis of a Large Social Graph Executive Summary

150123821 – Mehmet Sina Çağlar

Dataset Overview

Dataset Name: Wiki-Vote

Dataset Description: The Wiki-Vote graph contains data from Wikipedia administrator elections. In Wikipedia, some contributors are administrators who have access to additional features. For a user to become an administrator, they issue a request for adminship, and the Wikipedia community decides who to promote to adminship via votes. Nodes represent Wikipedia users. A directed edge from user A to user B represents a vote cast by A for B.

Type: Directed, unweighted

Key Network Metrics

Metric	Value	Interpretation
Nodes	7115	Total active voters
Edges	103689	Total votes cast
Density	0.002049	The graph is sparse. Only ~0.2% of possible votes were cast.
Average Path Length	2.88	Information spreads quickly (approx 3 hops)
Diameter	9	The max distance between any to connected nodes
Clustering Coefficient	0.082	Probability that two friends of a user also voted for each other.
Largest Strongly Connected Component (LSCC) Size	1300 nodes	The core “inner circle” where voting is reciprocal.
Assortativity	-0.083	Disassortative; Regular users vote for popular candidates
Reciprocity	0.0565	Low reciprocity. Voting is mostly hierarchical.
Average Conductance (Louvain)	0.085	Low conductance means well-defined and separated communities.

Modularity (Louvain)	0.42	Strong community structure.

Small-World Verdict

- The Wiki-Vote network exhibits small-world properties.
- The evidence is that the graph has a clustering coefficient of 0.08, which is much higher than a random graph ($C_{rand} \approx 0.002$). It also has a short average path length of 2.88, comparable to a random graph ($L_{rand} \approx 2.63$).
- The calculated small-world sigma (σ) is 36.01, which is greater than 1, confirming the small-world effect.

Diffusion Analysis

- We simulated the spread of information using the SIR model.
- Results show that starting diffusion from high-degree nodes (admins) leads to the fastest and widest spread of information.
- Randomly selected users failed to initiate a spread entirely. In the simulations, trends started by random users died out immediately (0 infected), whereas trends started by Admins went viral. This proves the 'Powerless Majority' effect in the network.

Top Insights

- The network is highly centralized. A small portion of “super-voters” (likely administrators) dominates the network, as seen in the heavy-tailed degree distribution.
- The Louvain algorithm detected 30 distinct communities with a modularity score of 0.42, and average conductance of 0.085. This suggests that the voting network is not random; voters form distinct clusters. It likely reflects different interest groups, languages, or tenures within Wikipedia.

- While Louvain succeeded (Modularity 0.42), Label Propagation failed completely (Modularity 0.001). This proves that in sparse, directed networks with low reciprocity, structure must be found via optimization, not information flow.
- The network has a negative assortativity of –0.08. This suggests a structure where low-degree nodes (regular users) tend to connect to high-degree nodes (candidates/admins), rather than connecting to each other.
- Degree and PageRank are strongly correlated, meaning that the most popular candidates are also the most influential.
- The lower correlation between Betweenness and Closeness suggests that some users act as bridges between different communities.