Async Learning Task: Networks

Patrick Fritzvoyle Cruz Jeanne Paulene De Leon
Charles Rey Diomampo Le Thi Minh Tam

November 5, 2024

1 Wikipedia Admin Election Network Visualization

1.1 Network Visualization

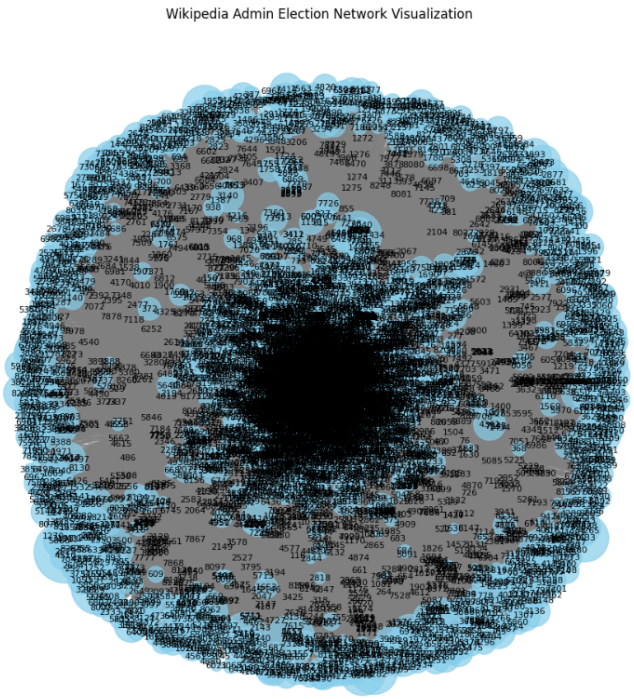

Figure 1: Visualization of the Network

Wikipedia is a free encyclopedia written collaboratively by volunteers around the world. A small part of Wikipedia contributors are administrators, who are users with access to additional technical features that aid in maintenance. In order for a user to become an administrator a Request for adminship (RfA) is issued and the Wikipedia community via a public discussion or a vote decides who to promote to adminship. Using the latest complete dump of Wikipedia page edit history (from January 3 2008) we extracted all administrator elections and vote history data. This gave us 2,794 elections with 103,663 total votes and 7,066 users participating in the elections (either casting a vote or being voted on). Out of these 1,235 elections resulted in a successful promotion, while 1,559 elections did not result in the promotion. About half of the votes in the dataset are by existing admins, while the other half comes from ordinary Wikipedia users.

The network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes in the network represent wikipedia users and a directed edge from node i to node j represents that user i voted on user j.

1.2 Visualization of subset data

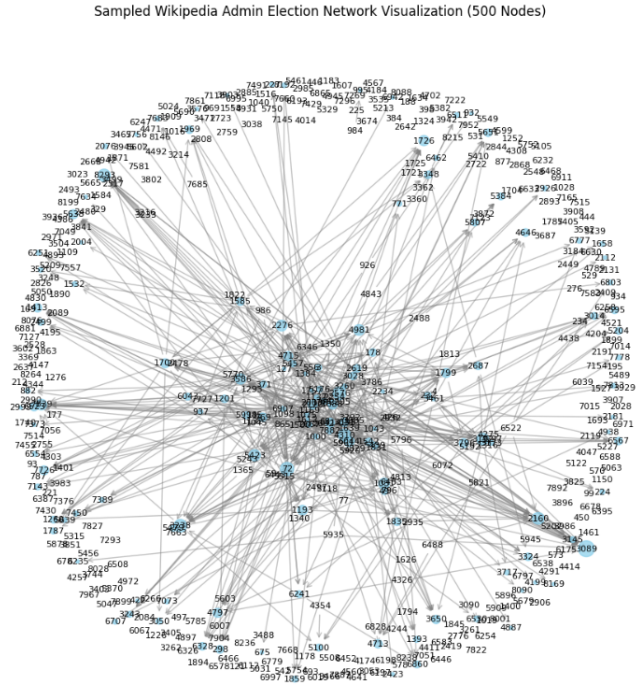

Figure 2: Subset Visualization of the Network - 500 nodes

The dataset contains 7,115 nodes and 103,689 edges, representing votes for Wikipedia administratorship. Analyzing and visualizing the full network may be computationally intensive and hard to interpret visually due to its density. A sample of 500 nodes will allow us to:

- Focus on a manageable subset for meaningful visualization.
- Capture essential patterns without overwhelming computational resources

1.3 Network Characterization

1.3.1 Node Degrees

Degree Distribution

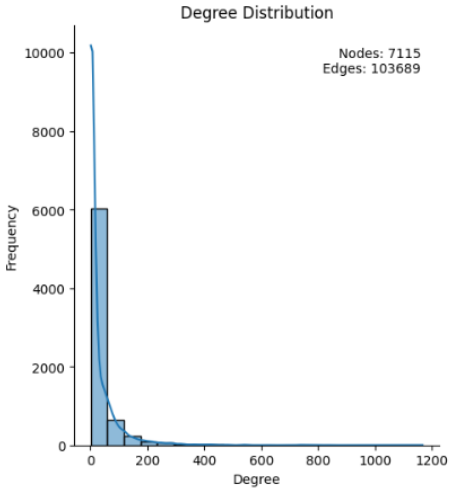

Figure 3: Degree Distribution Plot

Nodes: 7115
Edges: 103689

Hubs

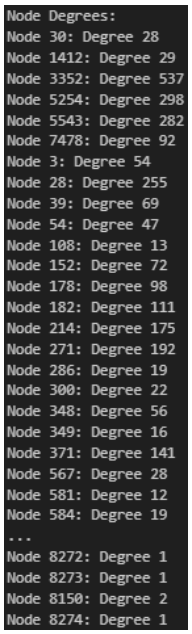

Figure 4: Sample Node Degrees Plot

1.3.2 Average Degree

Average Degree: 29.146591707659873
Average In-Degree: 14.573295853829936
Average Out-Degree: 14.573295853829936

1.3.3 Average Path Length

Average Path Length (Largest WCC, Undirected): 3.247509990226615

1.3.4 Diameter

Diameter (Largest WCC, Undirected): 7

1.3.5 Clustering Coefficient

Average Clustering Coefficient (Largest WCC, Undirected): 0.14187491841626332

1.3.6 Centralities

Degree Centrality

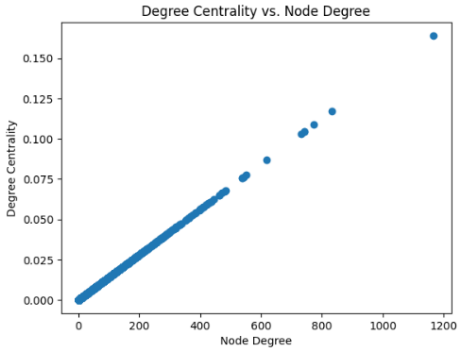

Figure 5: Degree Centrality Plot

Interpretation: The scatter plot shows a positive correlation between node degree and degree centrality. This means that nodes with higher degrees (more connections) tend to have higher centrality scores. In simpler terms, nodes that are more connected to other nodes are generally more important or influential in the network.

The plot suggests a linear relationship between the two variables, indicating that as the degree of a node increases, its centrality also increases proportionally.

Top 10 nodes with highest in-degree (most votes received): Node 4037: In-degree 457— Node 15: In-degree 361— Node 2398: In-degree 340— Node 2625: In-degree 331— Node 1297: In-degree 309— Node 2565: In-degree 274— Node 762: In-degree 272— Node 2328: In-degree 266— Node 5254: In-degree 265— Node 3352: In-degree 264

Betweenness Centrality

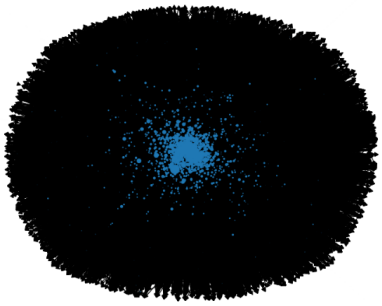

Figure 6: Betweenness Centrality Plot

Interpretation: Node size (visibly color blue nodes) corresponds to betweenness centrality. Larger nodes have higher centrality, indicating they play a crucial role in information flow or influence within the network. The specific meaning of the nodes and edges depends on the context of the network being analyzed.

Top 10 nodes by betweenness centrality: [(2565, 0.017654409558147836), (1549, 0.016564095998753692), (15, 0.01156258726064681), (72, 0.008011822532712367), (737, 0.006134997021063534), (1166, 0.0058025468349323876), (5079, 0.005438230580659507), (2328, 0.005202347375831278), (2237, 0.004714826895715726), (28, 0.004563992520641177)]

Explanation of Top 10: These top 10 nodes with high betweenness represent influential voters whose votes reflect or mediate between different factions or voting groups, impacting the outcome of elections by swaying opinions.

Closeness Centrality

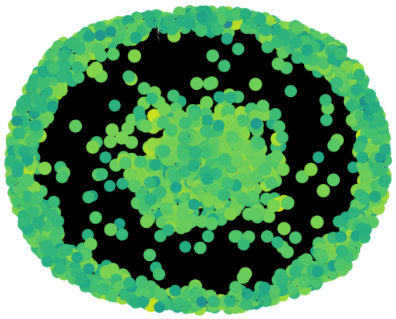

Figure 7: Closeness Centrality Plot

Interpretation: The visualization displays a network of Wikipedia users and their voting interactions. Node size corresponds to closeness centrality, indicating a user's influence in the voting network. The color variation likely represents additional user attributes, such as administrator status or voting activity.

This network can be analyzed to understand the dynamics of the Wikipedia community, identify influential users, and analyze voting patterns. By further exploring the network's structure and properties, insights can be gained into the decision-making processes and power dynamics within the Wikipedia community.

Top 10 nodes by closeness centrality: [(4037, 0.2985392599462622), (15, 0.293511230305062), (2398, 0.2929401967830677), (1549, 0.2838824682249186), (2535, 0.2818419832781087), (3089, 0.2799763130280714), (762, 0.27993467157450885), (5412, 0.2797449939744266), (2565, 0.2795396946875405), (5254, 0.2784849839951483)]

Explanation of Top 10: These top 10 nodes with high closeness represent highly integrated community members whose opinions reach and potentially affect the community efficiently.

Eigenvalue Centrality

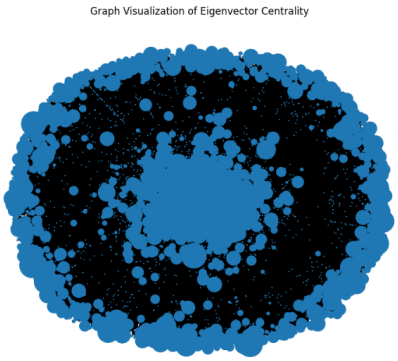

Figure 8: Eigenvector Centrality Plot

Interpretation: The visualization depicts a network of Wikipedia users, with node size representing eigenvector centrality. Eigenvector centrality highlights users who are connected to other influential users, forming a chain of importance. Larger nodes indicate users with greater influence within the Wikipedia voting network. This visualization can help identify key users and understand how influence is distributed within the community.

Top 10 nodes by eigenvector centrality: [(2398, 0.1171983832471754), (4037, 0.10895543930803386), (15, 0.09817933816669923), (4191, 0.09568989575762346), (2625, 0.09552610300572519), (1549, 0.09503186906635995), (2328, 0.09483743172846111), (3089, 0.09310705125076532), (5412, 0.09050808243358636), (2066, 0.09048358862307544)]

Explanation of Top 10: These top 10 nodes with high Eigenvector Centrality represent those well-known contributors or long-standing admins whose endorsements are especially meaningful. Their votes might carry significant weight and set trends for others.

1.4 Robustness and Vulnerability Analysis

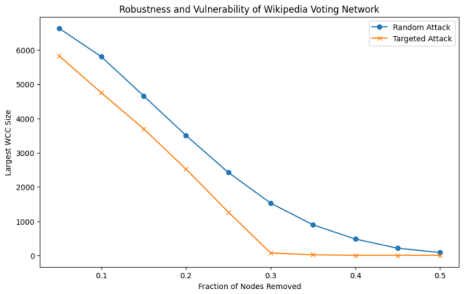

Figure 9: Robustness and Vulnerability of Wikipedia Voting Network Plot

Interpretation: As we can observe in the Random Attack, as the network's WCC size and connectivity decrease gradually which suggests resilience, implying that the community is well-distributed and not overly depended on a few central users.

However if we look at the Targeted Attack, there's a quick breakdown at around 0.3 which implies vulnerability wherein certain highly connected users(hubs) play a crucial role (influential voters or widely supported candidates).