Report of mini project 3

Nazmus Sakib

Introduction

This project did an extensive network analysis of a Facebook-like Social Network, encompassing 1,899 users and 59,835 messages exchanged. The main goals were to identify active users, top message recipients, influential users, calculate node distances for recommendations, and detect communities. These analyses are pivotal for gaining insights into user engagement, understanding influence dynamics, and uncovering intricate network structures in diverse social and online communities. Here is the project link:
https://colab.research.google.com/drive/1y3DLvlSMopFtwRxi9oK_KBMW8bdzPTWw?usp=sharing

Data processing

I have done the following preprocessing:

- Calculated the sum of edges of each unique pair. This sum means the count of message transfers between the pair of nodes
- Stored the message count in a dedicated 'count' column, replacing the Unix timestamp.
- Finally I have verified that the sum of the count column equals the number of edges in the original dataset (59,835), ensuring data integrity.

Graph Creation and Data Reduction:

- A Directed Graph was created using the NetworkX library.
- For demonstration purposes, the initial graph size was reduced to the first 100 nodes and their corresponding 348 edges. This subset ensures a manageable scale while retaining key characteristics of the larger network.

Degrees Analysis:

The analysis of degrees, in-degrees, and out-degrees for each node provided crucial insights. They shed light on the fundamental connectivity and communication patterns exhibited by individual users.

Top Message Recipients and Visualization:

- Identified users who received messages from the most people, emphasizing the highest in-degree.
- Visualized the node with the highest incoming messages, providing a context-rich view along with its neighbors.

Top Outgoing Messages Sender and Visualization:

- Identified users who sent messages to the most people, highlighting the highest out-degree.
- Visualized the node with the highest outgoing messages, offering a comprehensive understanding of its network interactions.

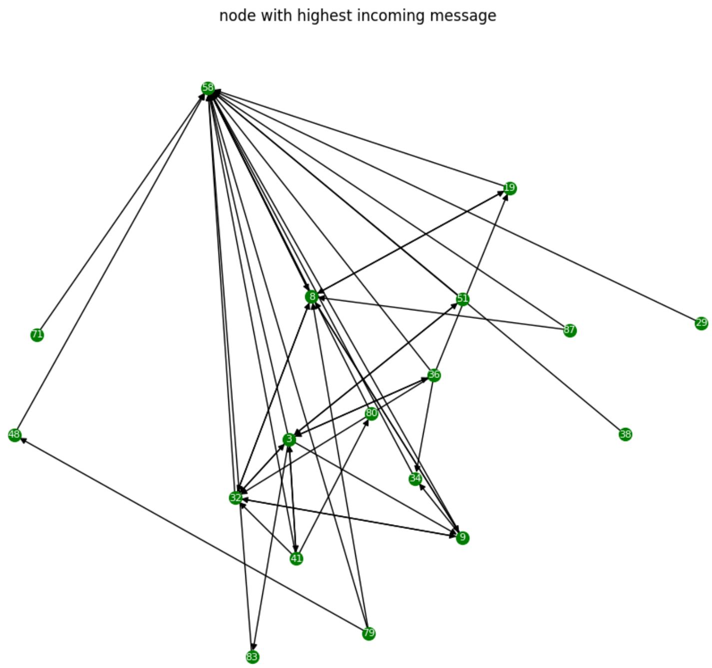

Fig: node 58 received messages from the most number of people

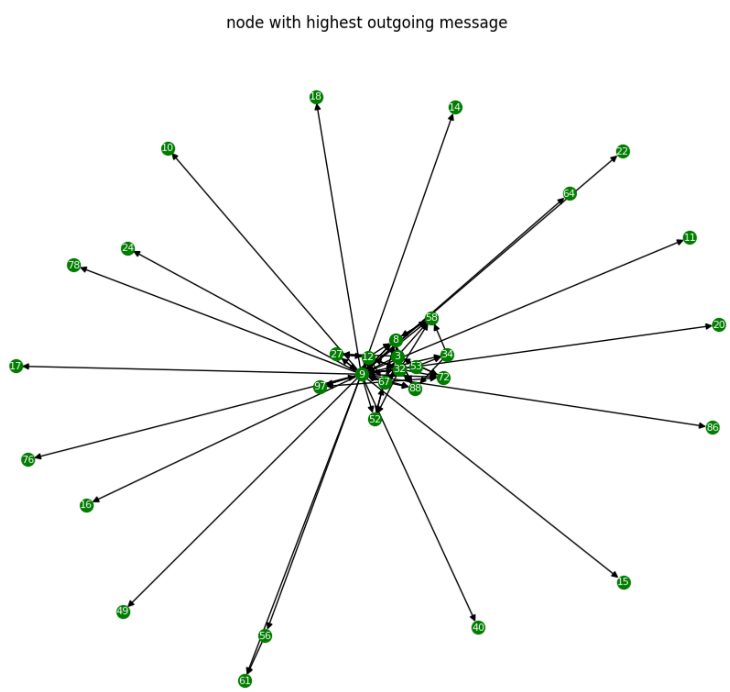

Fig: node 9 sent messages to the most number of people

Nodes with Highest Edge Weight / Active Users and Visualization:

In this phase of the analysis, my focus was on identifying and understanding the most active users within the network. I took a comprehensive approach, considering both the quantity and intensity of connections each user has. The results provide valuable insights into the network's dynamics, emphasizing the users who play a crucial role in shaping interactions.

Weighted Degrees Analysis:

- Weighted degrees were calculated, taking into account both the number of connections (degree) and the intensity of those connections (edge weight).
- Edge weight, determined by the 'count' attribute representing message exchanges, offers a more clear measure of the strength of connections.

Top Active Users:

- A list of users was compiled based on their weighted degrees, showcasing those with the highest overall impact on the network.
- The weighted degrees provide a holistic view of user activity, highlighting individuals who not only have numerous connections but also engage in substantial message exchanges.

Strategic Insights:

- This analysis aids in identifying key players within the network, offering strategic insights for user engagement, content distribution, and community building.
- By understanding the most active users, we can tailor our strategies to maximize impact and enhance overall network interactions.

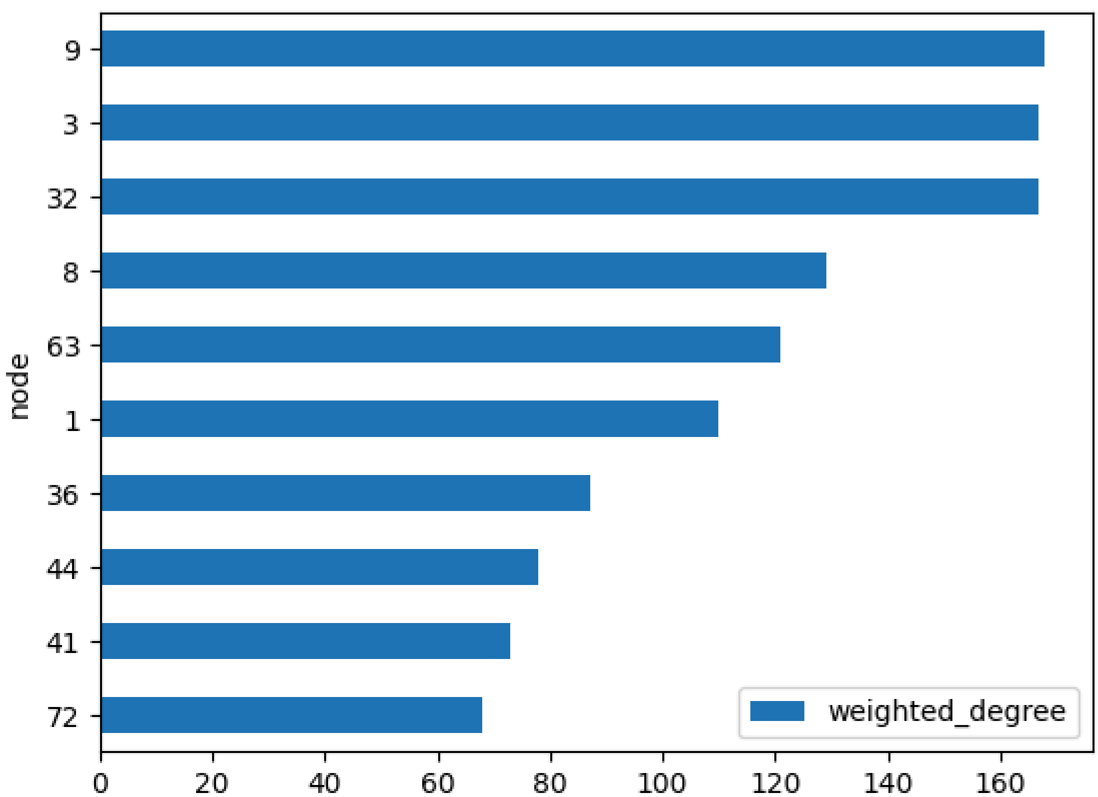

Fig: Nodes that have most connections and most message exchanges

Influential Users Analysis: Eigenvector Centrality and Betweenness Centrality

This segment of the analysis focused on identifying influential users within the network based on Eigenvector Centrality and Betweenness Centrality metrics. The goal was to uncover users who play pivotal roles in connecting others and, in turn, influencing the overall network dynamics.

Eigenvector Centrality:

Top Influential Nodes:

- The Eigenvector Centrality calculation identified the top 10 most influential nodes in the network.
- Users were ranked based on their Eigenvector Centrality scores, which take into account both the user's connections and the connections of their connections.

Subgraph Visualization:

- A subgraph was created to visually represent the top influential nodes, showcasing their interconnectedness and prominence within the network.
- The green nodes in the visualization indicate the most influential users based on Eigenvector Centrality.

Betweenness Centrality:

Top Nodes with High Betweenness Centrality:

- The Betweenness Centrality metric was utilized to identify users who act as crucial bridges or intermediaries within the network.
- The top 10 nodes with the highest Betweenness Centrality scores were determined.

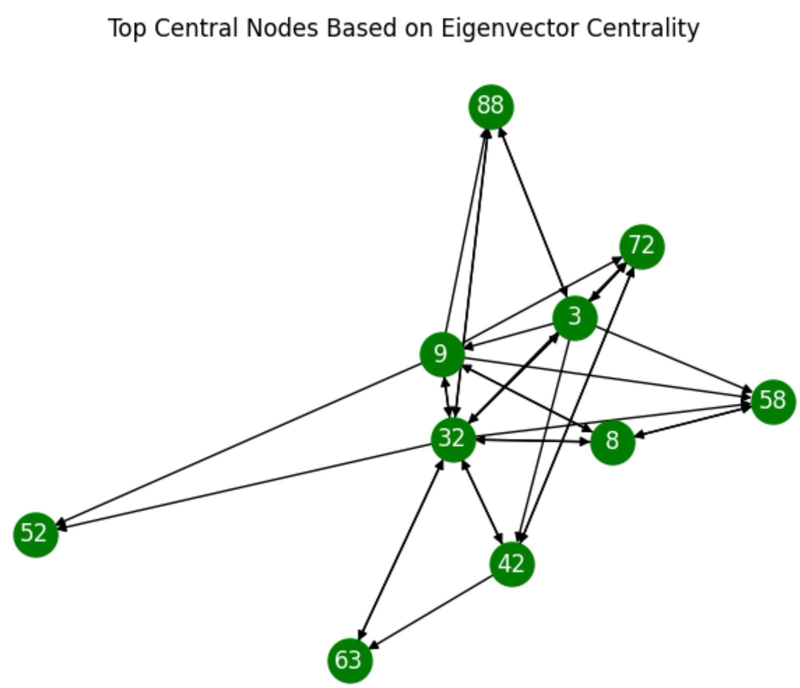

Fig: Top influential nodes based on eigenvector centrality

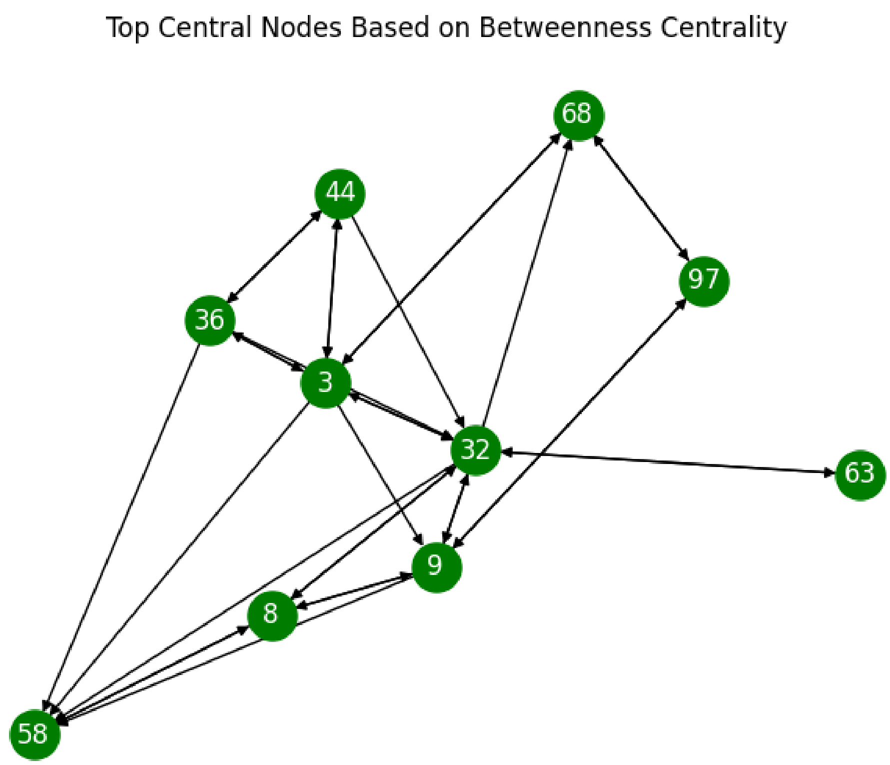

Fig: Top influential nodes based on betweenness centrality

Subgraph Visualization:

- A subgraph was generated to visually represent users with high Betweenness Centrality, emphasizing their pivotal role in connecting different parts of the network.
- The nodes in the visualization highlight users with significant influence based on Betweenness Centrality.

Key Findings:

Eigenvector Centrality:

- Identified users who not only have many connections, but are also well-connected to other highly connected users.
- These users are likely influential in shaping the overall network structure.

Betweenness Centrality:

- Uncovered users acting as critical connectors or bridges between different parts of the network.
- Highlighted individuals who play a central role in facilitating communication and connections among diverse user groups.

Strategic Insights:

- **Network Hubs:** Users with high Eigenvector Centrality are likely network hubs, influencing a broad range of connections.
- **Connectors**: Users with high Betweenness Centrality act as crucial connectors, ensuring effective communication across the network.

Node Distance Calculation and Visualization:

I calculated the distance between two given nodes and visually represented the path between them if it exists. They give us the following insights:

Distance Calculation:

Illuminated insights into the network's structure and connectivity.

Path Visualization:

Enhanced understanding through visual representation of the connection path between specified nodes.

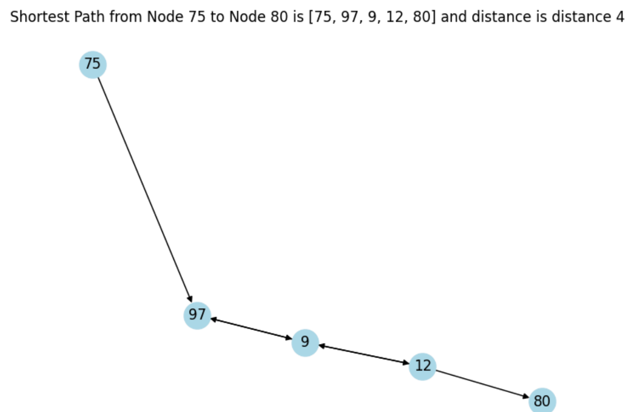

Fig: Shortest path between two nodes. For example here it's between node 75 and node 80

Community Detection and Interactive Visualization:

This phase of the analysis focused on uncovering distinct communities within the network and presenting them visually through an interactive visualization. The objective was to identify groups of nodes that exhibit a higher degree of interconnectedness, providing a more granular understanding of the network's structural patterns.

Community Generation:

- Used community detection algorithm of greedy modularity to unveil distinct groups or communities within the graph.
- Each community is characterized by a unique number.

Node Size and Color Assignment:

- Dynamically assigned node sizes based on the degree of each node, providing a visual cue about the nodes' connectivity.
- Enriched the visual experience by employing a color palette associated with each community number, enhancing the distinction between communities.

Interactive Visualization:

- Leveraged the Bokeh library to create an interactive visualization of the network.
- Users can dynamically explore and interact with communities through panning, zooming, and tapping on nodes for detailed information.

Key Findings:

- The visualization showcases the existence of distinct communities within the network.
- Node sizes reflect their degree, indicating the number of connections.

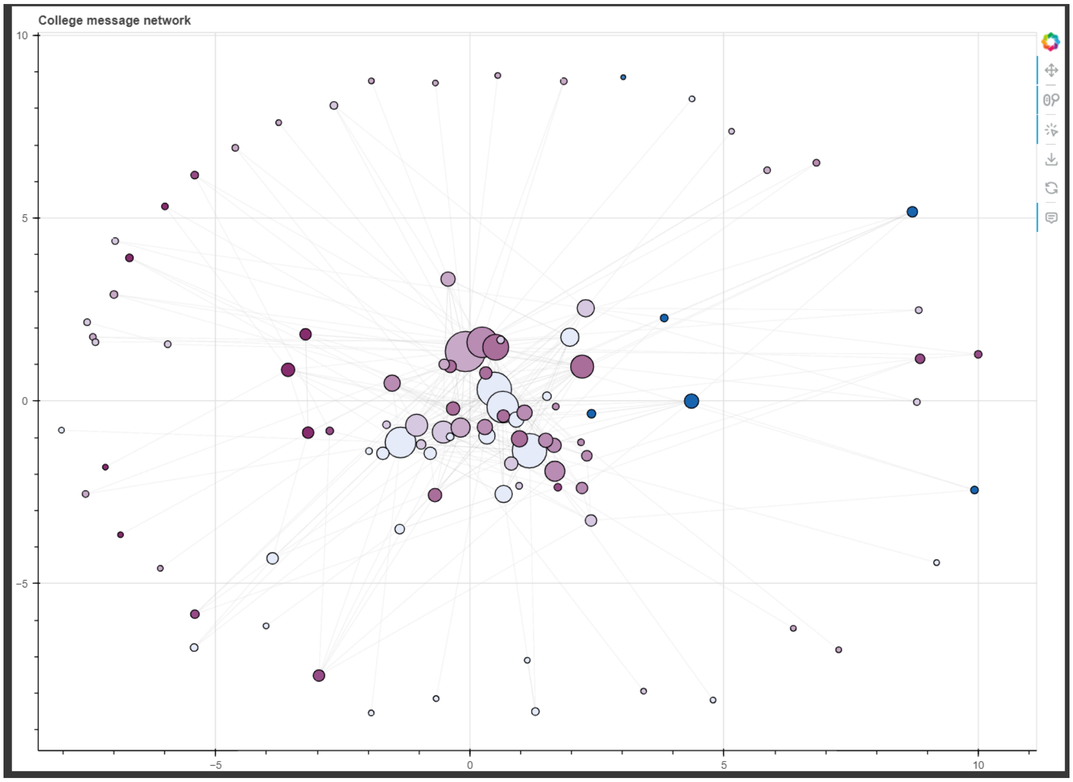

Fig: Communities in the network

Conclusion:

In conclusion, the network analysis of the Facebook-like Social Network has uncovered vital insights into user engagement, communication patterns, and influential nodes. Through data processing, graph creation, and various analyses, we identified active users, top message recipients, and influential individuals based on Eigenvector Centrality and Betweenness Centrality metrics. The visualization of distinct communities enriched our understanding of network structures. These findings can equip us with strategic insights to enhance user engagement, optimize content distribution, and foster a cohesive online community.