FACEBOOK DATASET ANALYSIS

Graphs And Social Network Project

MOTIVATION

- **Social networks** like Facebook play a significant role in **understanding human behavior** and **relationships**. Analyzing social networks can reveal **patterns** that enhance **efficiency** and minimize weaknesses in social interactions. As these **networks grow**, there is an increasing **need** for proper **analytical methods** to understand their **complexities**.

OBJECTIVE

- **Extracting properties** and **insights** from Facebook dataset.
- **Analyzing** user **connections** and **subgraphs** representing social circles.
- **Computing** various centrality **measures**, **visualisations**, and network **parameters**.
- **Deriving inferences** about social behaviors and **influential nodes** in the network.

DATASET

- Dataset features an undirected social network with **4,039 users (nodes)** and **88,234 friendships (edges)**, reflecting Facebook's social circles and demonstrating **strong connectivity** between nodes.

- Network is strongly connected, all nodes and edges form both the largest Weakly Connected Component (**WCC**) and Strongly Connected Component (**SCC**).

- Each user is treated as a node, and an edge exists between two nodes if they are friends. The edges reflect the **number of connections, and mutual friends**.

GRAPH CREATION

- Importing essential Python libraries, including **NETWORKX** for graph analysis, **MATPLOTLIB** for visualisations, and community detection algorithms like **Louvain, Label Propagation, and Girvan Newman**.

- Utilizing these libraries and algorithms to calculate **centrality measures**, identify **communities**, and analyze other graph properties.

INITIAL FINDINGS

- The graph density of 0.0108 shows a **sparse network** with **limited connections**
- A diameter of 8 indicates **maximum user distance**, impacting information spread, while a radius of 4 allows **connections within four steps**, reflecting **small-world theory**.
- A global clustering coefficient of 0.605 suggests **cluster formation**, with k-core decomposition revealing **influential subgroups** and varying local clustering coefficients indicating **diverse connectivity patterns**.

GRAPH TRAVERSAL TECHNIQUES

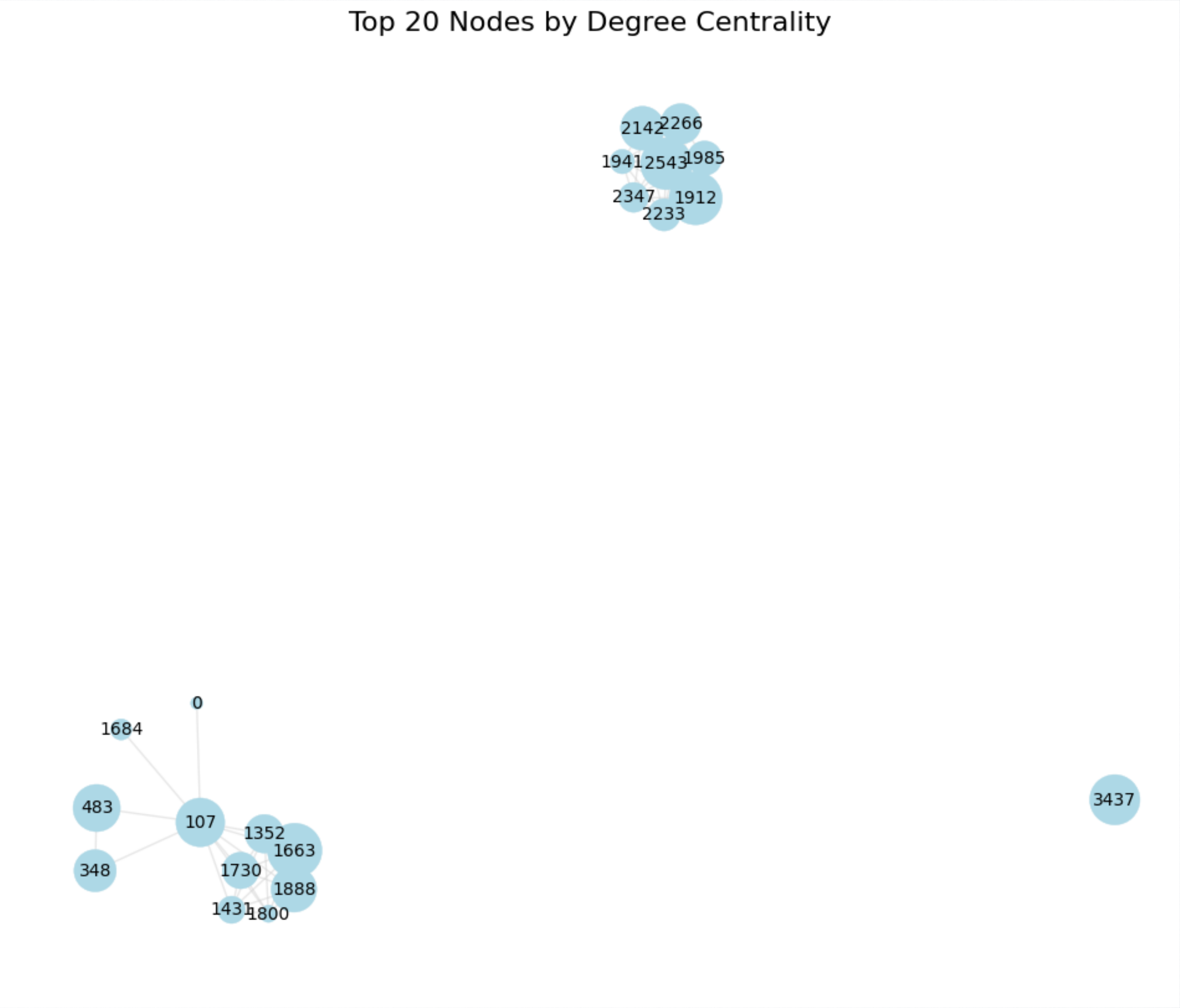
- Allows users to find the **shortest path** between two nodes by **inputting** Node IDs, enhancing **interactivity** and usability.
- Dijkstra's Algorithm: Calculates **shortest path** using **priority queue** and handling scenarios where no path exists between nodes.
- Breadth-First Search (BFS): Finds shortest path by **exploring all neighbors** before moving deeper.

VISUALISATIONS

DEGREE CENTRALITY

- Indicates **most connected users**, identifying influential individuals within social circles, impacting information dissemination.

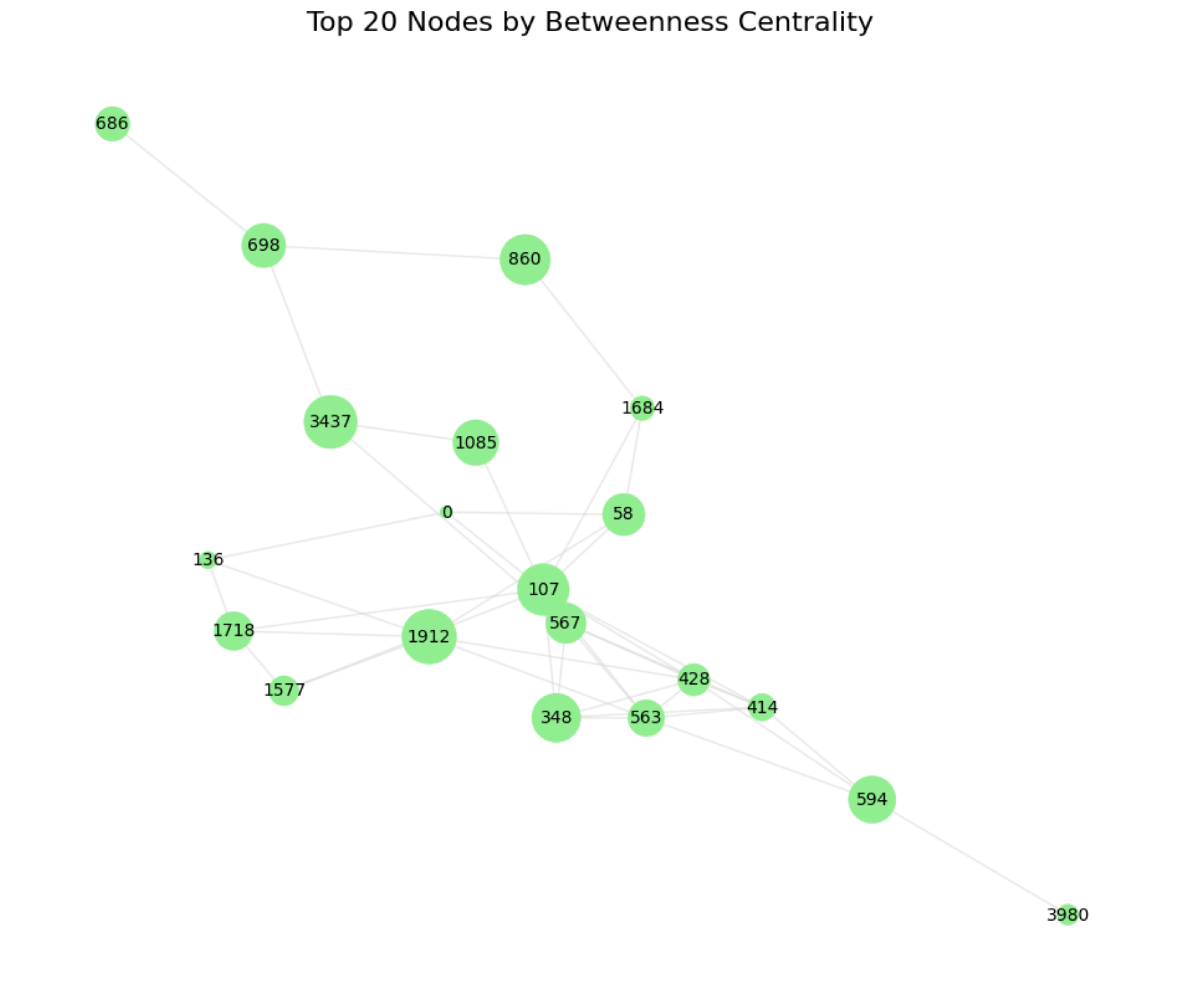
- Nodes **107** has the highest Degree Centrality of 0.2588.



BETWEENNESS CENTRALITY

- Highlights **intermediaries** in the network, crucial for **understanding information flow** and potential influencers in shaping opinions or trends.

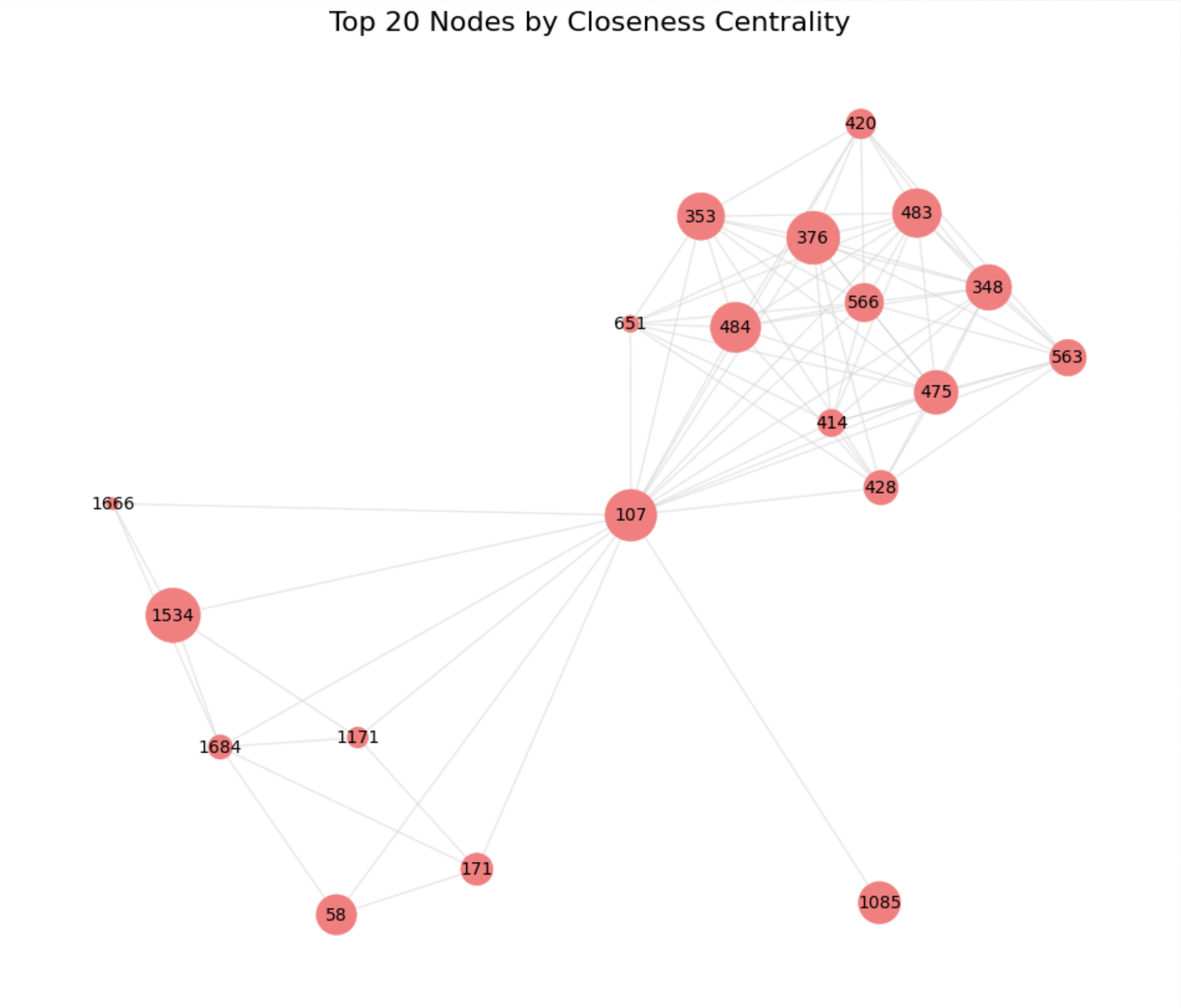
- Node **107** has the highest Betweenness Centrality of 0.4805.



CLOSENESS CENTRALITY

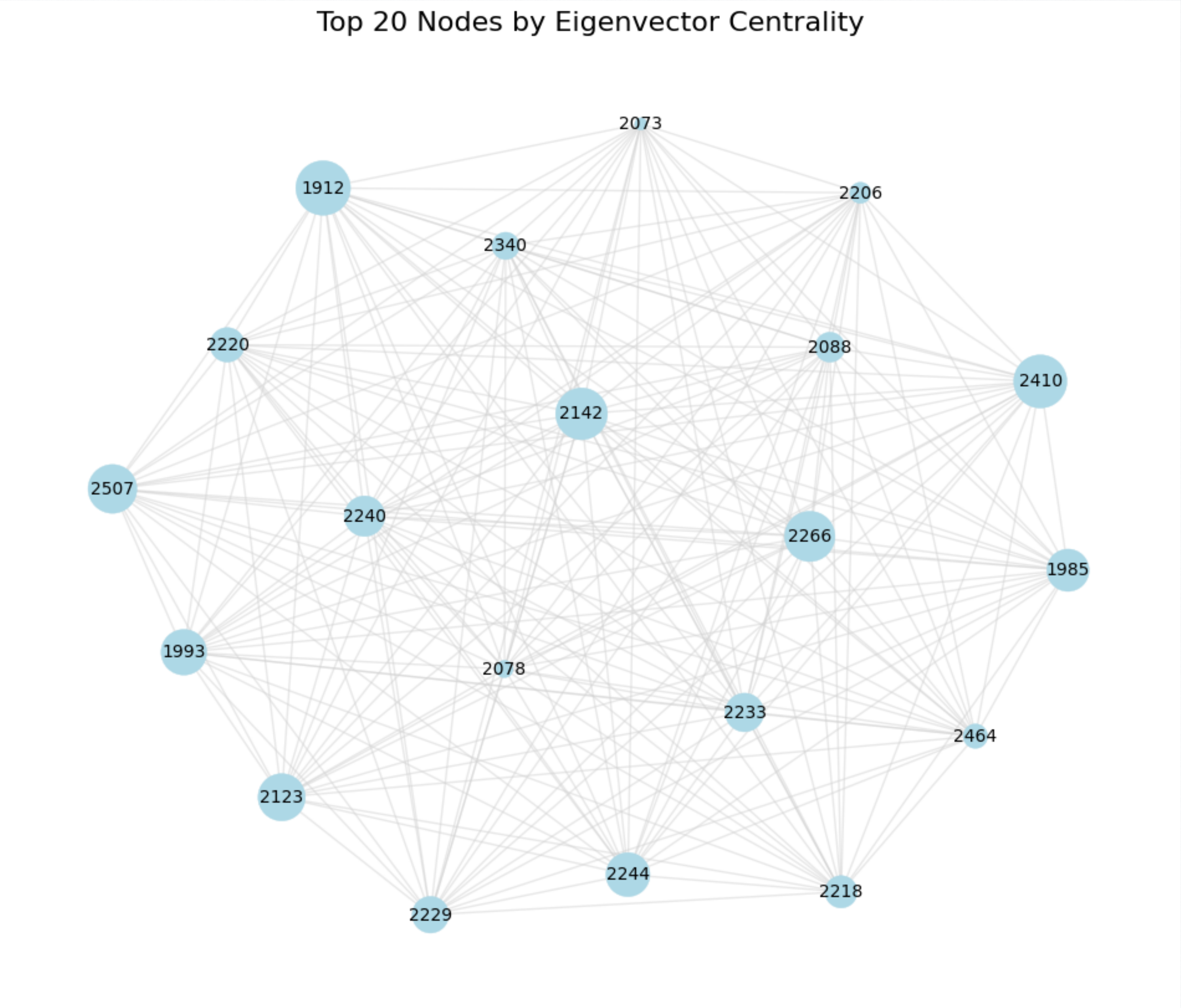
- Identifies users who can **quickly** reach others, suggesting importance in **efficient** information spread within social circles.

- Node **107** has the highest Closeness Centrality of 0.4597. Thereby being a prominent member of the dataset.



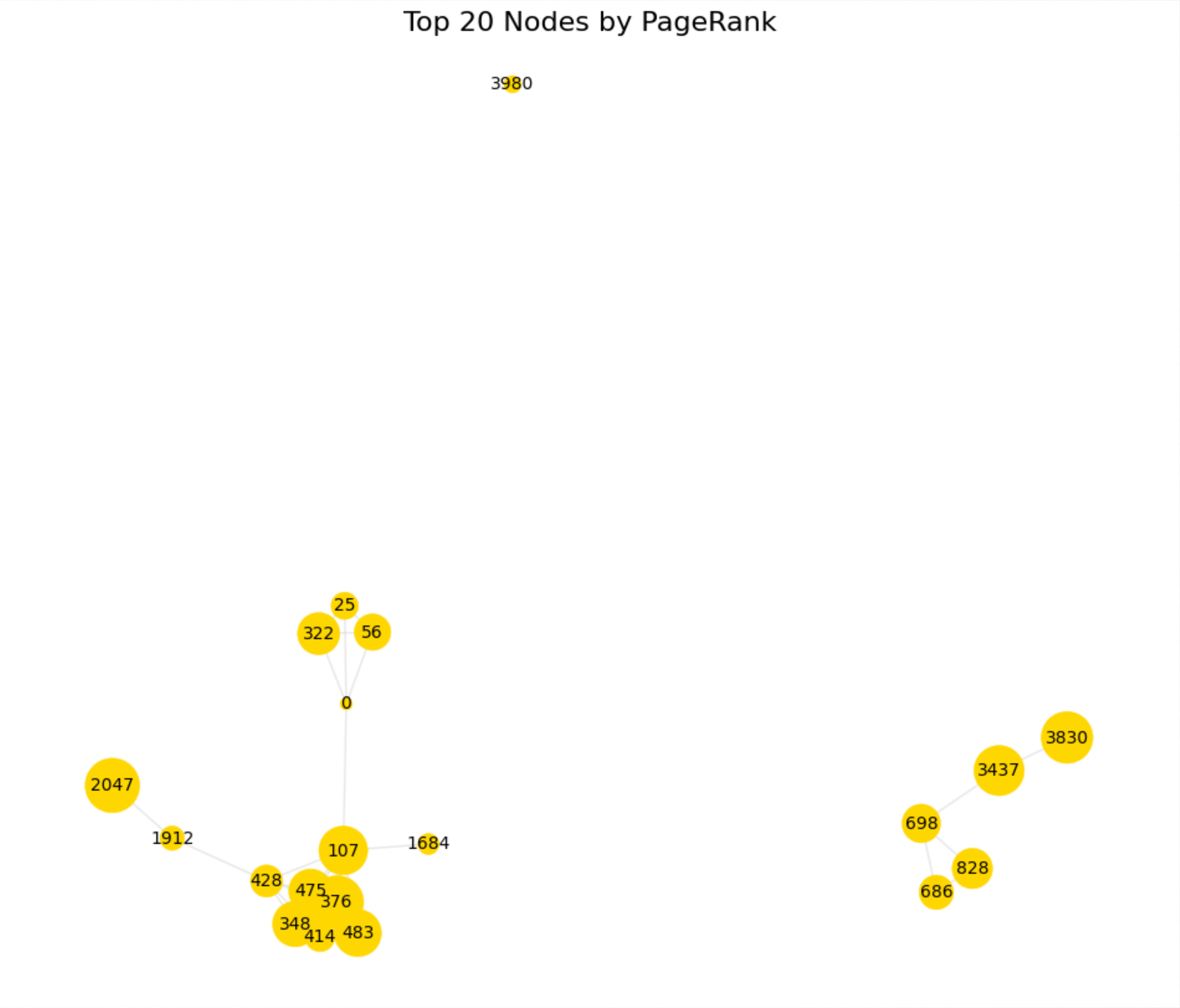
EIGENVECTOR CENTRALITY

- Recognizes nodes connected to other **highly influential** nodes, revealing key players whose connections **enhance importance** in the network.
- Node **1912** has the highest Eigenvector Centrality of 0.0954.



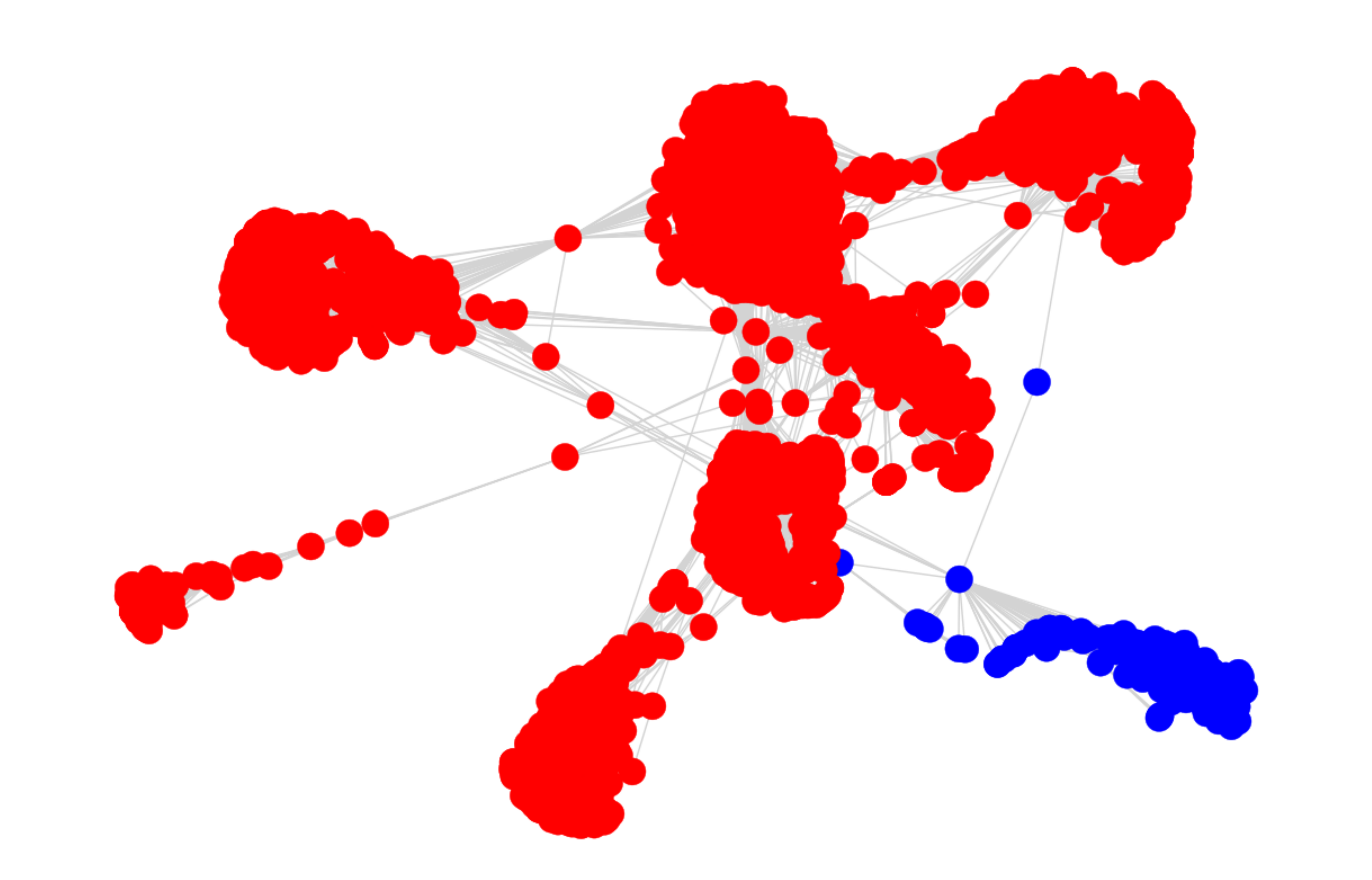
PAGERANK

- Assesses **node importance** through connections and their quality, identifying most **prominent users** who can significantly affect social dynamics.
- Node **3437** has the highest ranking PageRank of 0.0076.



COMMUNITY VISUALIZATION (GIRVAN-NEWMAN)

- This iteratively **removes edges** with the highest **betweenness** centrality, progressively splitting network into **smaller**, **densely** connected communities, revealing **divisions** within the graph.

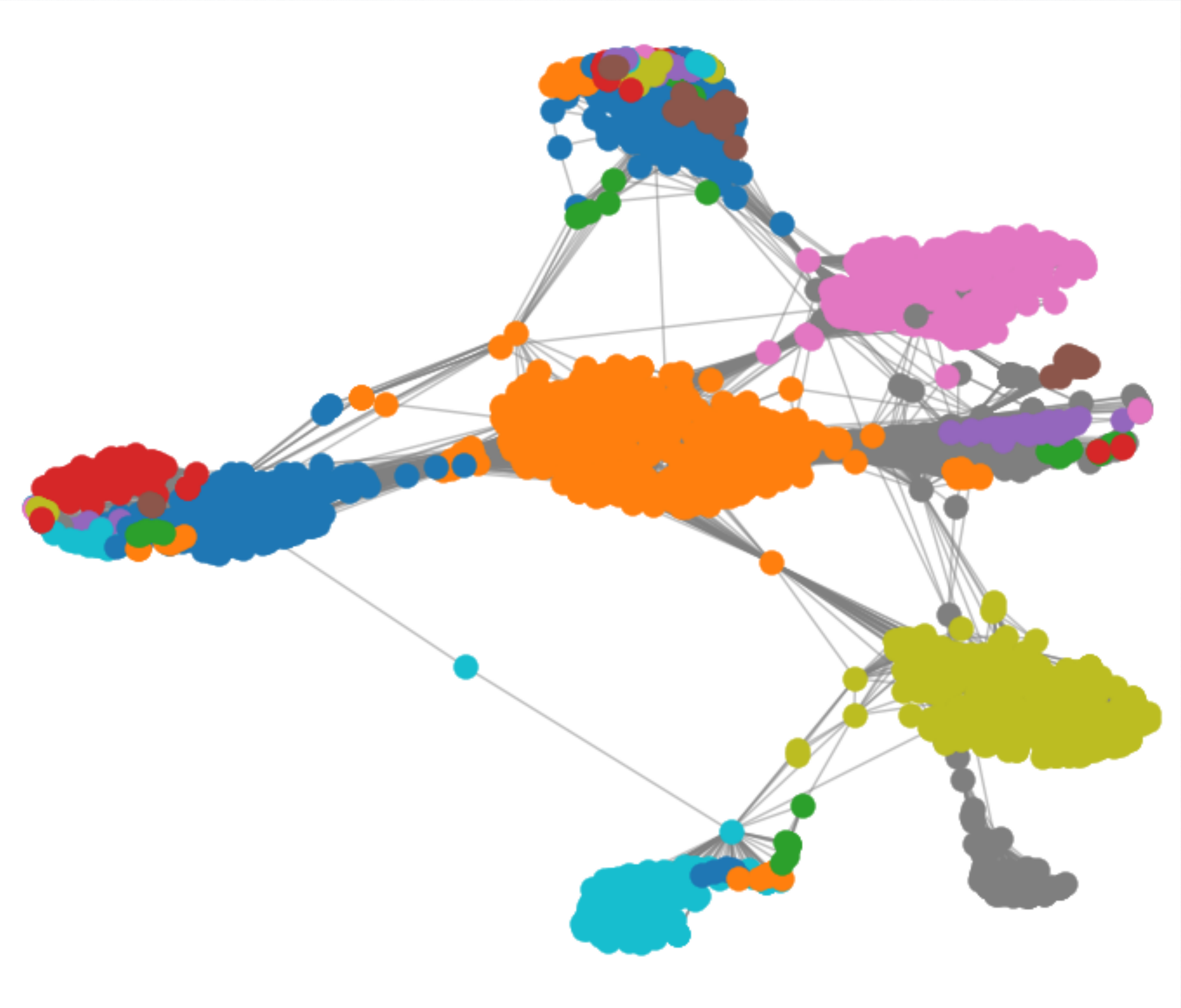


COMMUNITY VISUALIZATION (LOUVAIN)

- This algorithm optimizes modularity to find **dense clusters** in the network, revealing **tightly-knit communities** that **share behaviors** or affiliations.

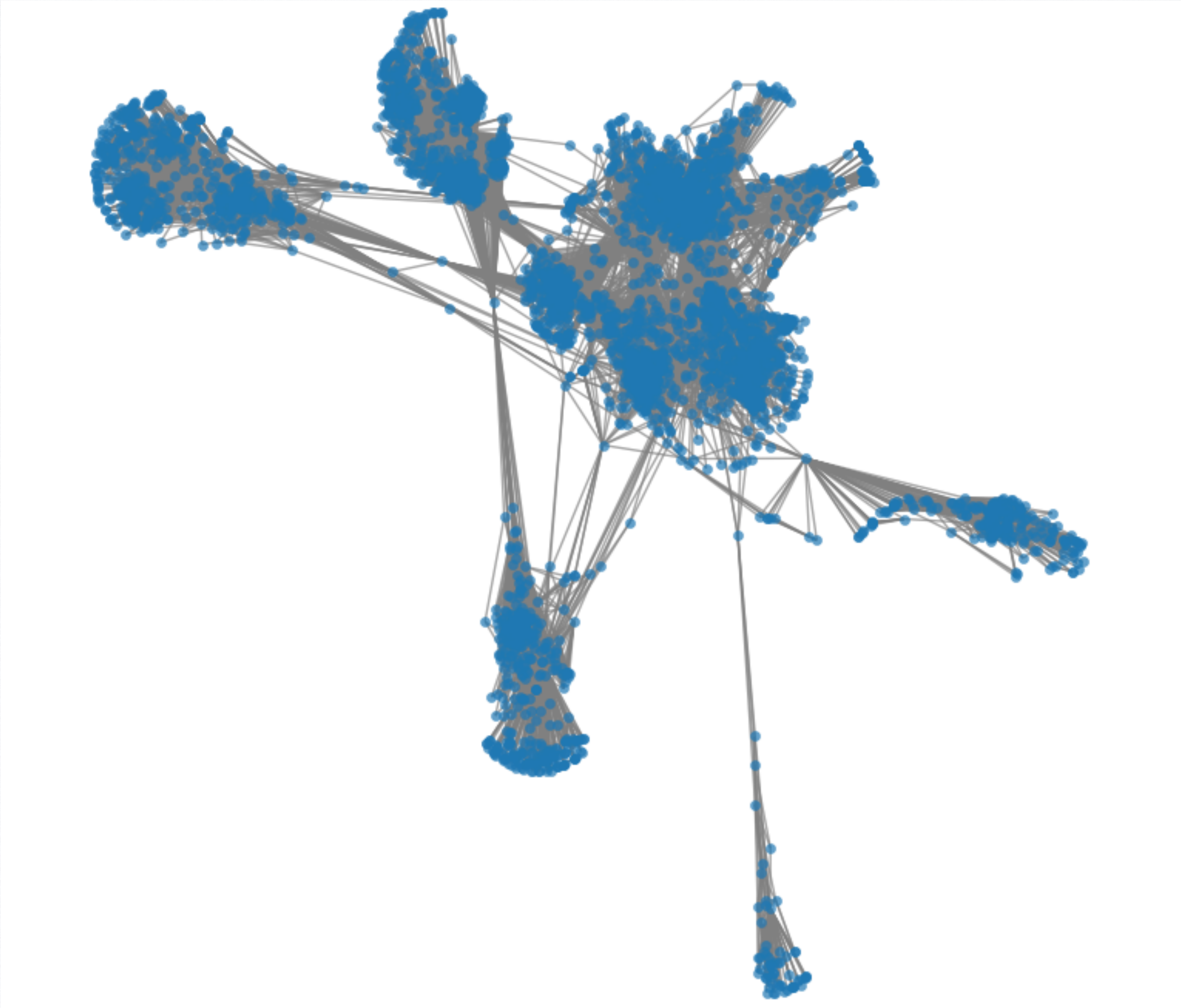
COMMUNITY VISUALIZATION (LABEL PROPAGATION)

- **Detects communities** by enabling nodes to adopt their **neighbors' labels iteratively**, revealing natural groupings reflecting social circles and shared interests.



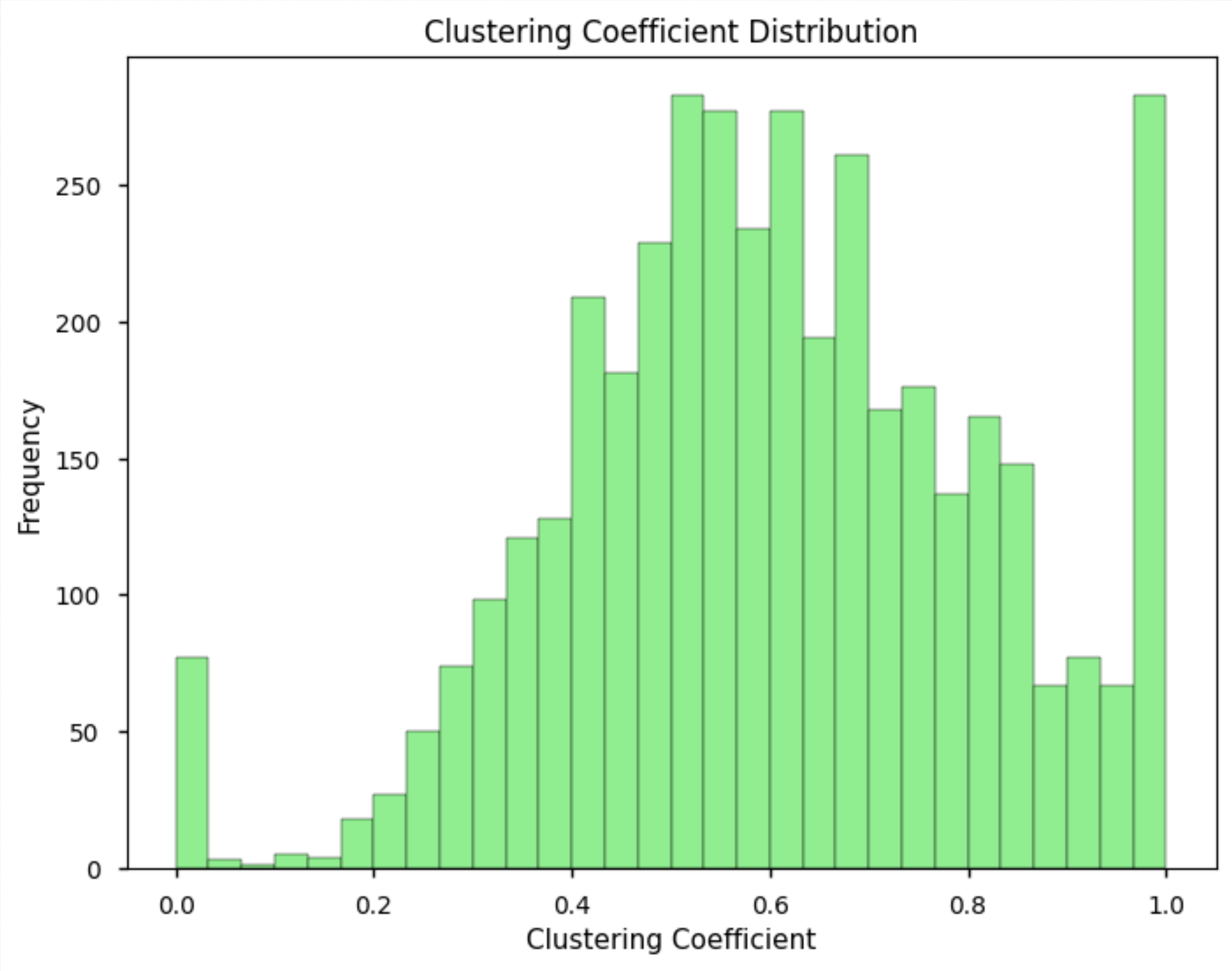
NODE LINK DIAGRAM

- Represents network's structure by displaying **nodes as points** and **edges as lines**, providing an understanding of user relationships and highlighting influential users within the dataset.



CLUSTERING COEFFICIENT DISTRIBUTION

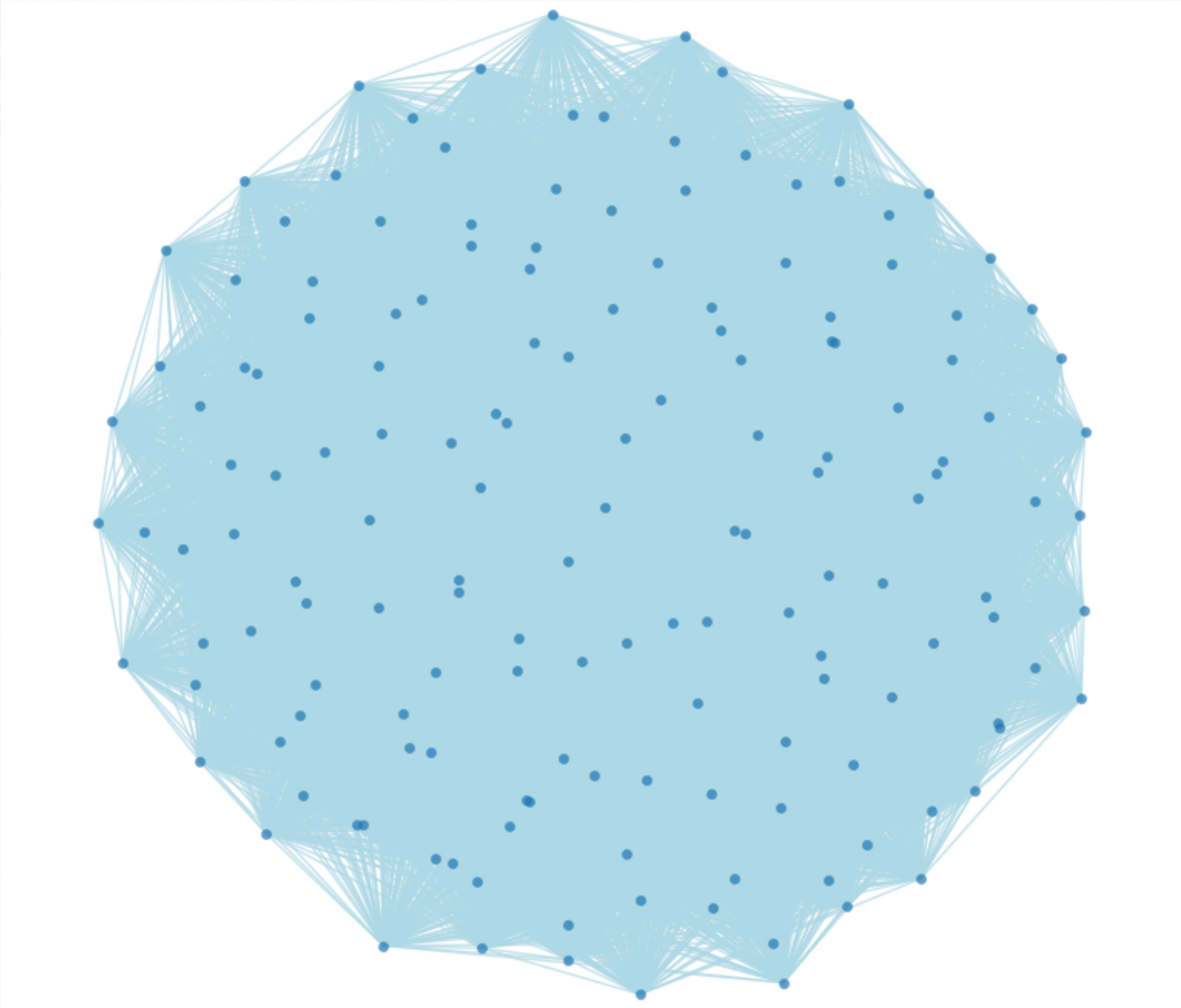
- Examines how **nodes cluster** together, a **higher clustering coefficient** suggests **strong community structures**, offering insights into the density and social dynamics of Facebook circles.



K-CORE SUBGRAPH VISUALIZATION

- Identifies core structures by removing nodes with a degree less than k, **revealing tightly connected subgroups** and helping understand robustness and resilience of community connections.

- K-Core calculated for **k=1**, meaning it retains all nodes that have at a degree of 1.



INFERENCES

- Identification of **highly influential nodes** suggests targeting these users could **amplify communication** strategies and **enhance information dissemination** within network.
- Significant clustering observed indicates that **communities are tightly knit**, facilitating **quicker** and more **effective sharing** of information among members within network.

Thank You