DATASET ANALYSIS OF

# ZACHARY'S KARATE CLUB

# MOTIVATION

- **Understanding social interactions** is vital for comprehending human behavior, as demonstrated by **Zachary's Karate Club** dataset.
- The analysis aligns with the principles of **Network Science**, emphasizing intricate relationships among community **members** and the **dynamics** shaping social structures.

## OBJECTIVES

- **Extracting properties** and **insights** from the dataset.
- Computing various **Centrality** measures and **Visualizing** network structures.
- **Analyzing communities** and **user** roles within network.
- **Deriving insights** into user behavior and the **influential nodes** in network.

# DATA

- The Zachary's Karate Club features an **undirected social network** with **34 nodes** and **78 edges**, illustrating **social dynamics** and **interactions** within the club.
- Each **member** is represented as a **node**, and an **edge** exists between two nodes if they are **friends**, illustrating **relationships** and **interactions** among members of the club.

## GRAPH CREATION

- Loading Zachary's Karate Club dataset via NetworkX, creating a clean undirected graph for precise analysis, using Python libraries like Matplotlib, NumPy, and Seaborn.
- Applying Community detection algorithms (Girvan-Newman, Louvain, Label Propagation) and Centrality measures to further understand user interactions, revealing insights into network structure and influential nodes.

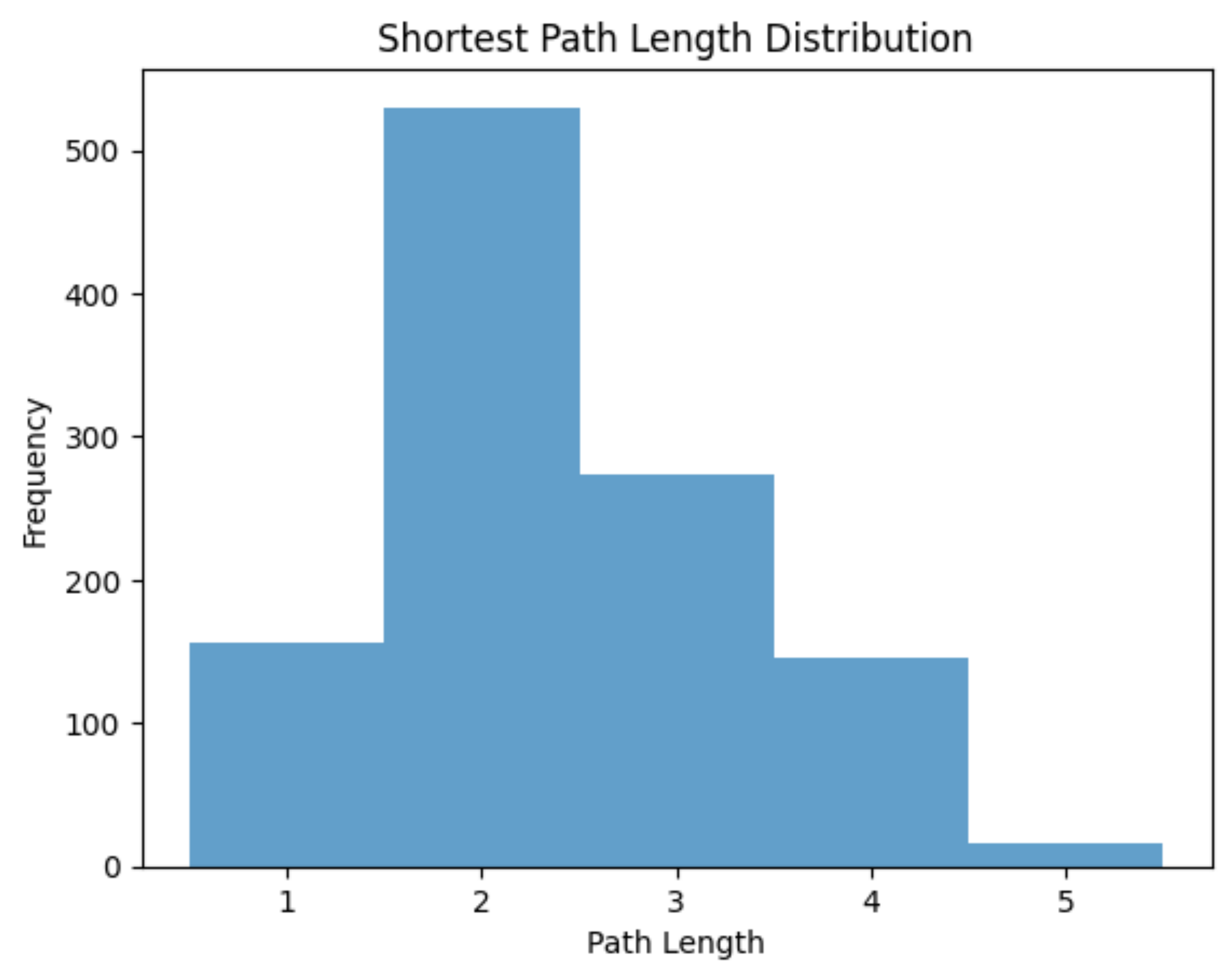
# INITIAL FINDINGS

- A Graph density of 0.1390 indicates well-connectedness in the network with a diameter of 5 and average path length of 2.41, facilitating efficient flow of information and quick reachability.
- A Radius of 3 demonstrates most users connect within three steps, exemplifying small-world properties enhancing communication and collaboration.

# VISUALIZATIONS

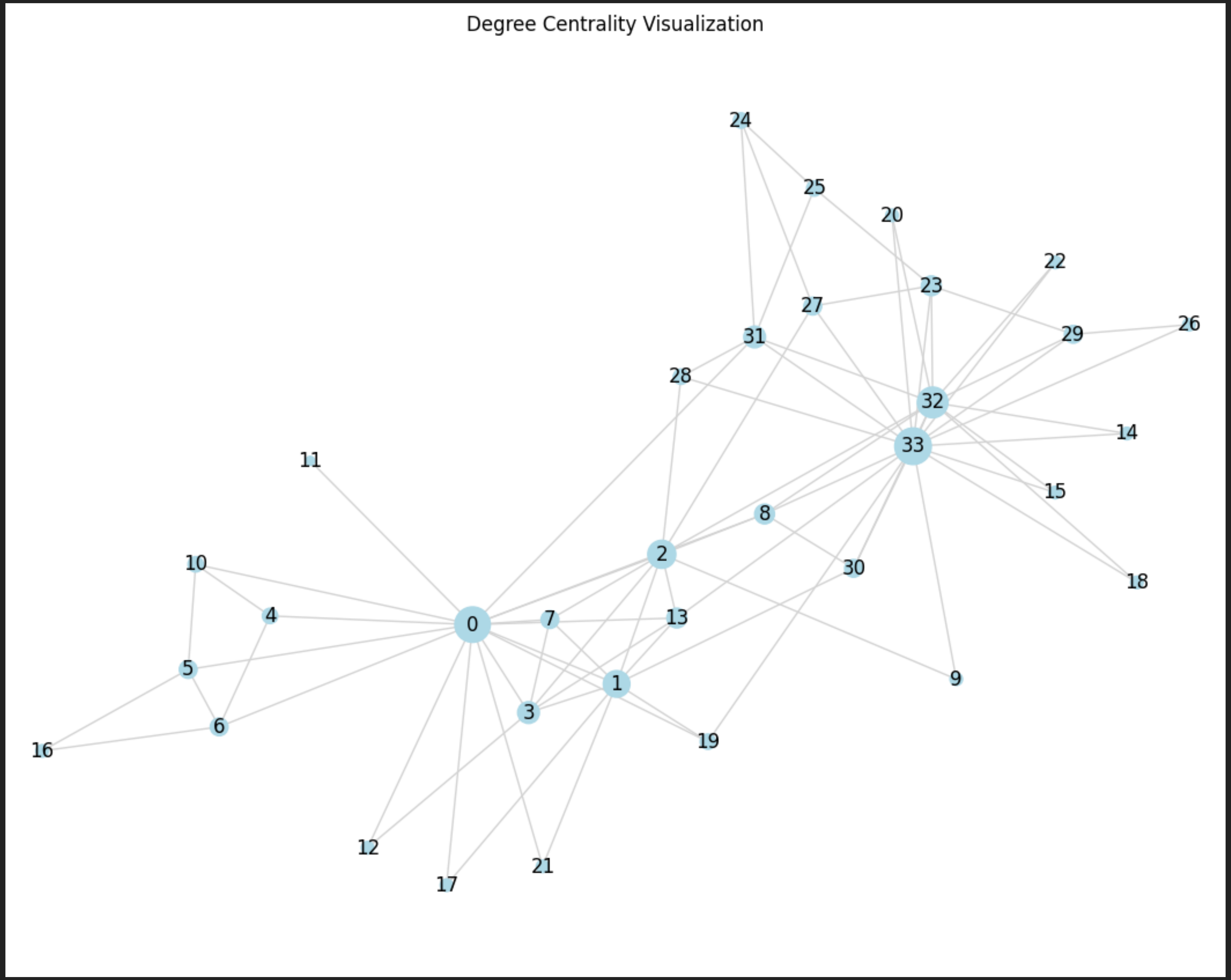
# SHORTEST PATH LENGTH DISTRIBUTION

- Illustrates connectivity of individuals within network and ease of information flow from one user to another. Lower values indicate closely linked users, while a higher signifies more isolated individuals.



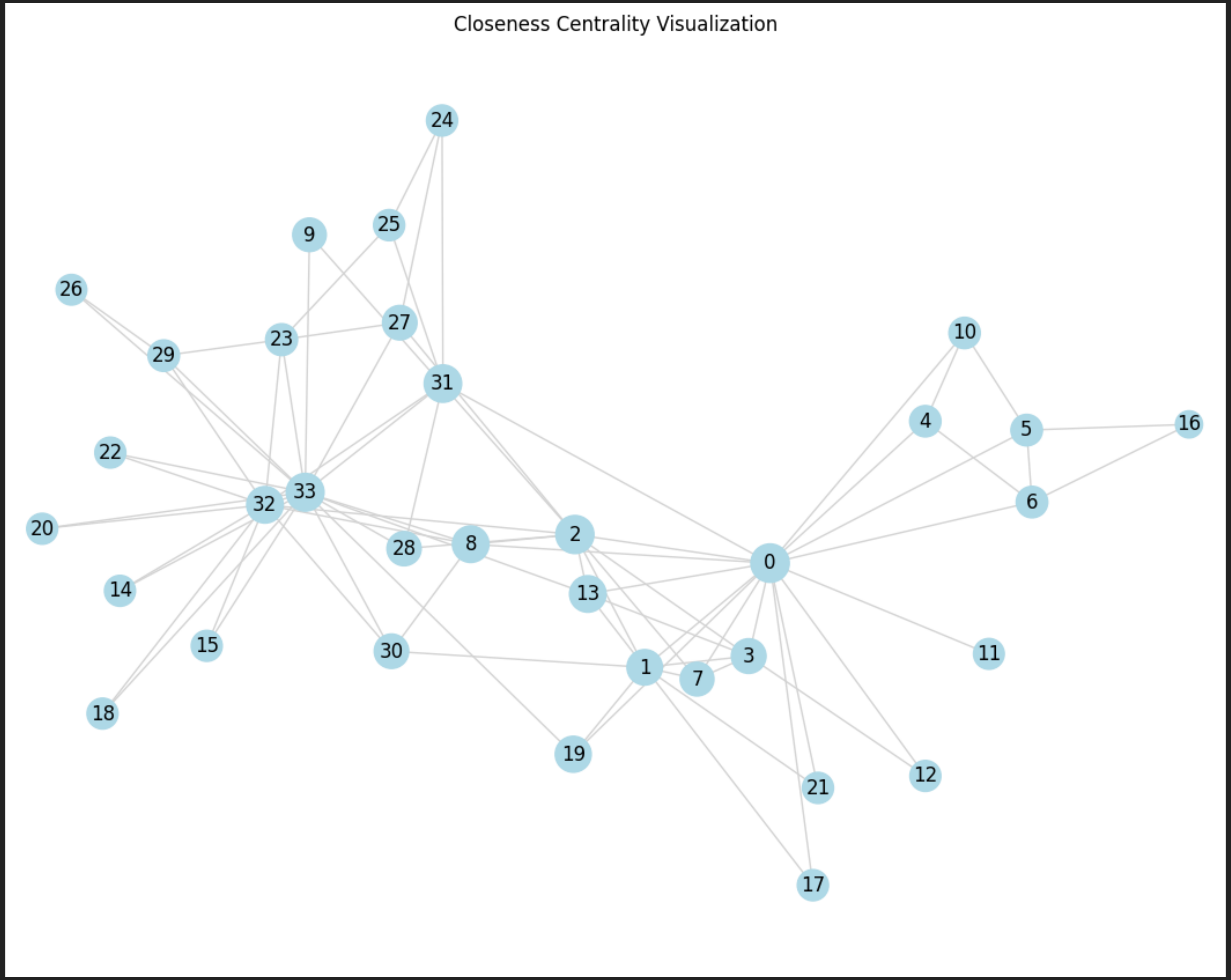
# DEGREE CENTRALITY

- Reveals number of direct links each node has within a network, highlighting influential nodes. Helps us understand who can quickly communicate and disseminate information to others in their immediate circle.



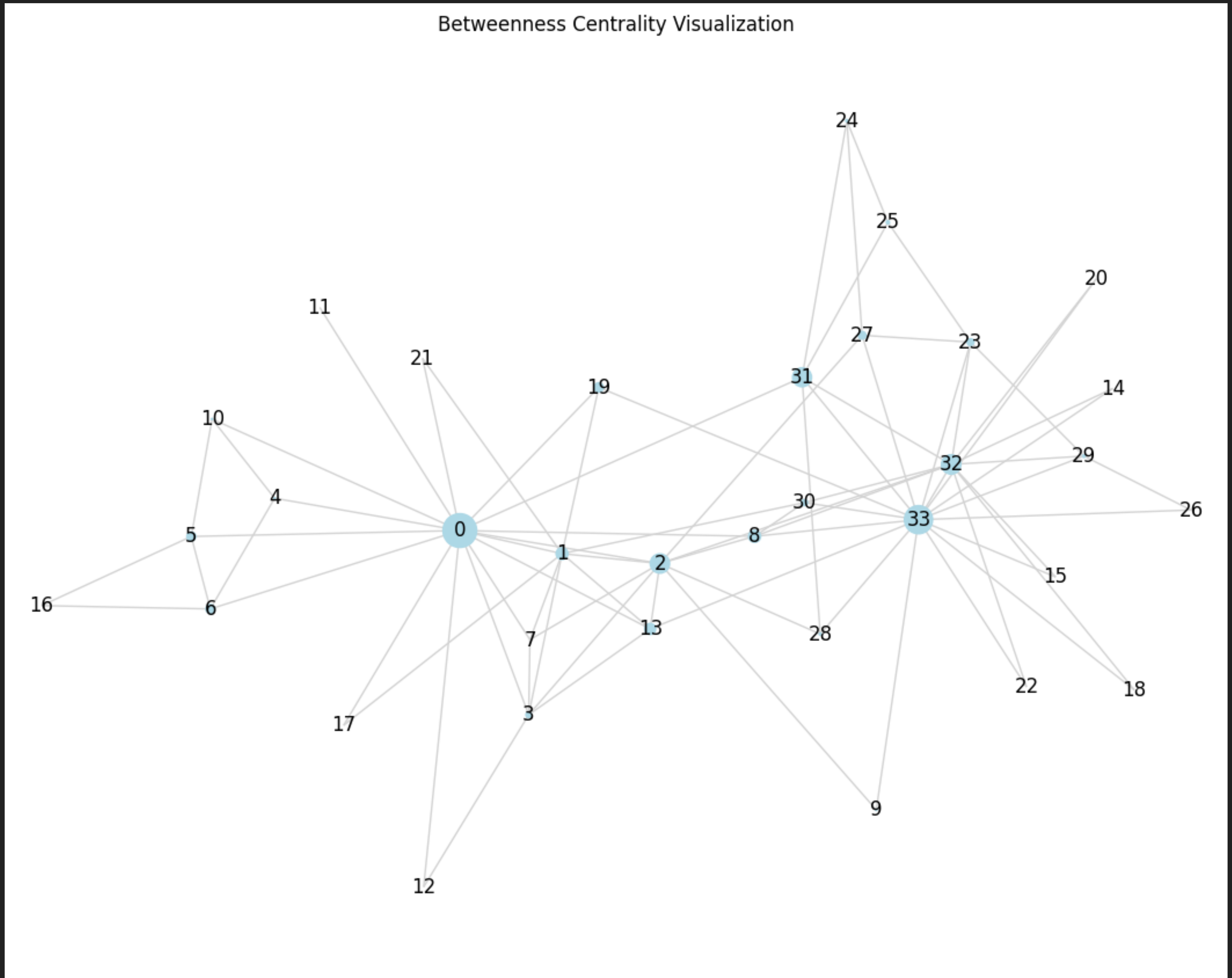
# CLOSENESS CENTRALITY

- Measures a node's speed in reaching others within the network, emphasizing effectiveness in communication. High closeness centrality indicates rapid spread of information within group, making these nodes vital for efficient information flow.



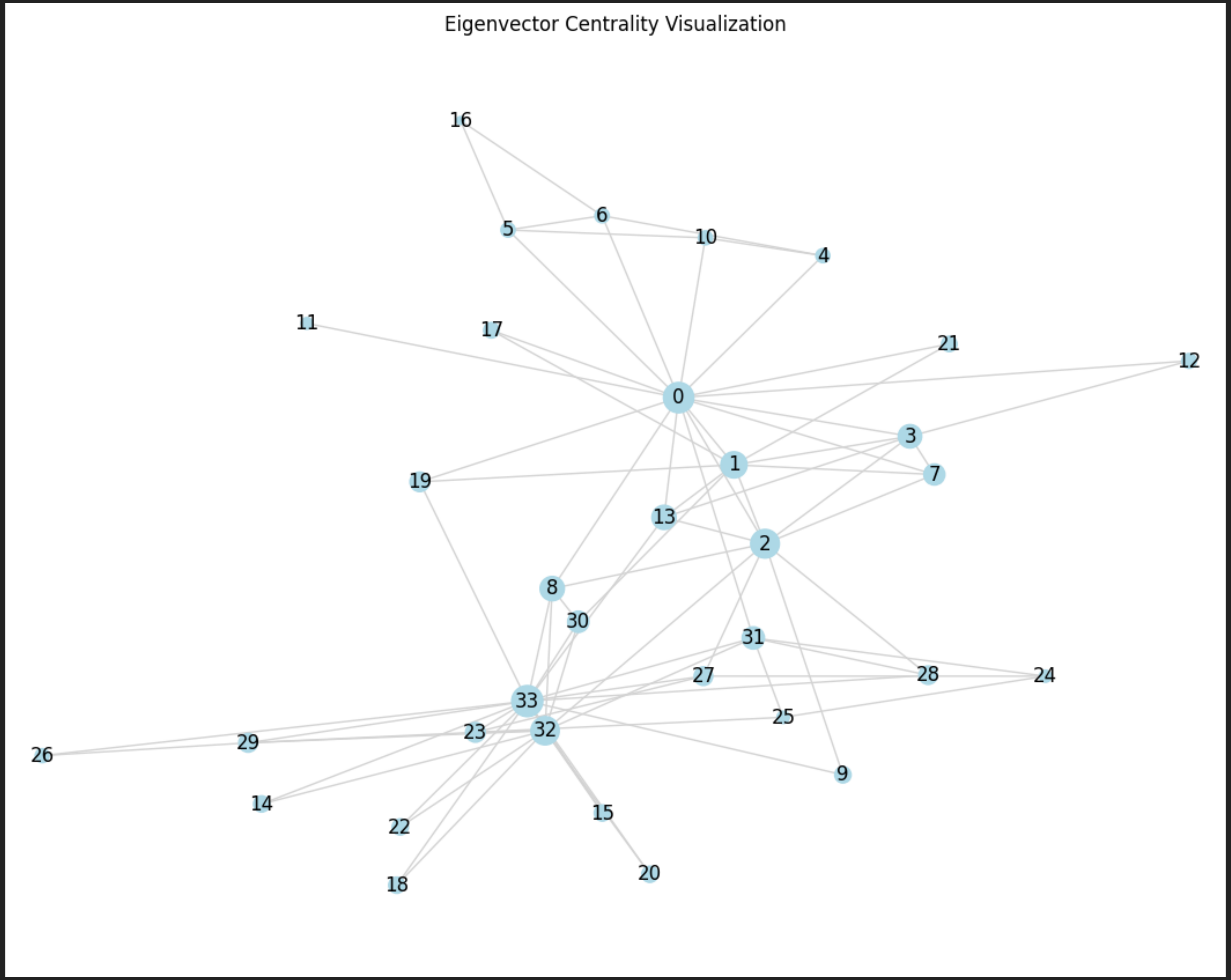
# BETWEENNESS CENTRALITY

- Measures importance of individuals in connecting others within network. Higher values indicate individuals playing vital roles in information flow, serving as bridges facilitating communication between different groups.



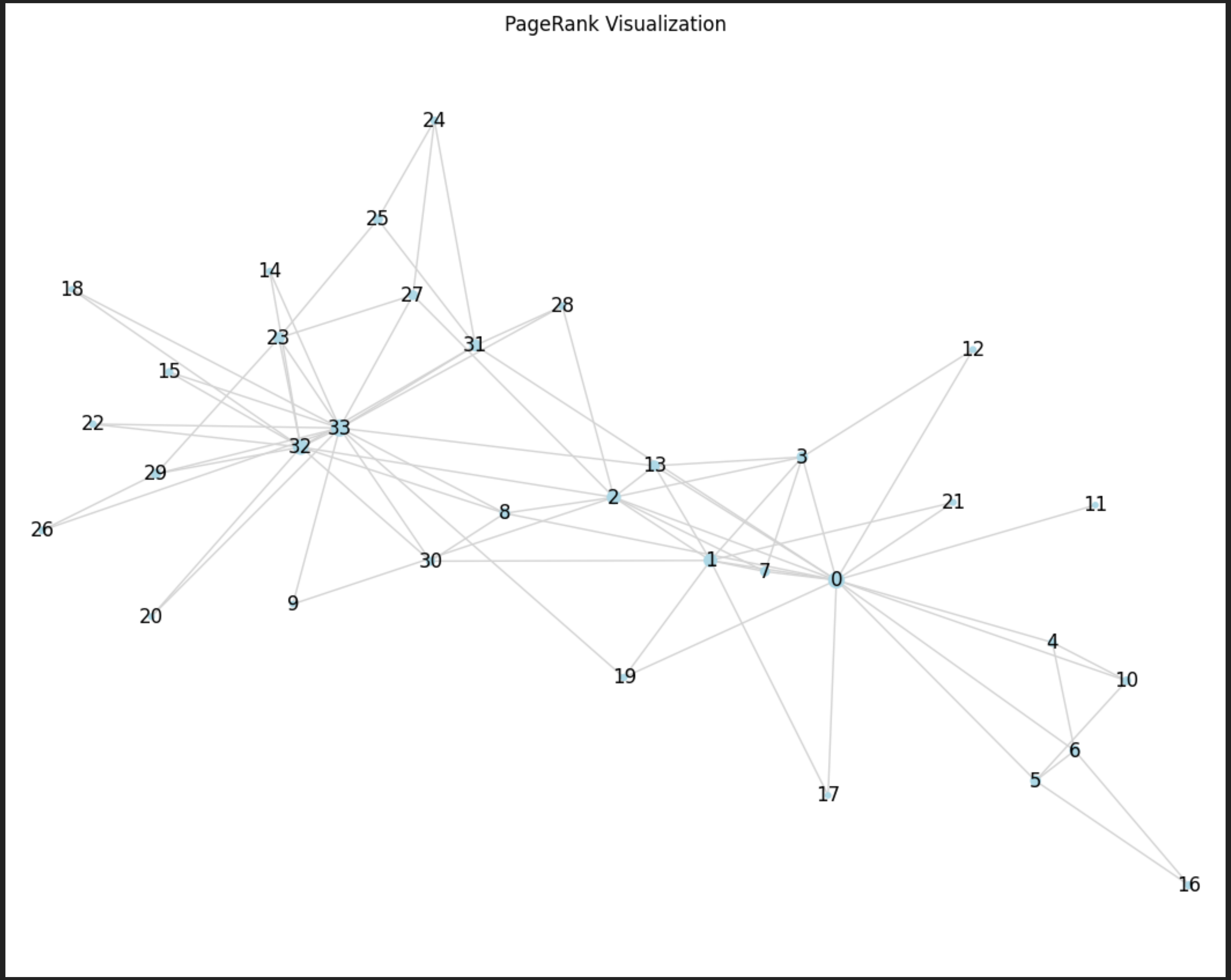
# EIGENVECTOR CENTRALITY

- Assesses **node's influence** based on **quality** of its **connections** rather than merely quantity. Identifies nodes that are not only well-connected but also **linked** to other **influential nodes**, highlighting **significance** within the network.



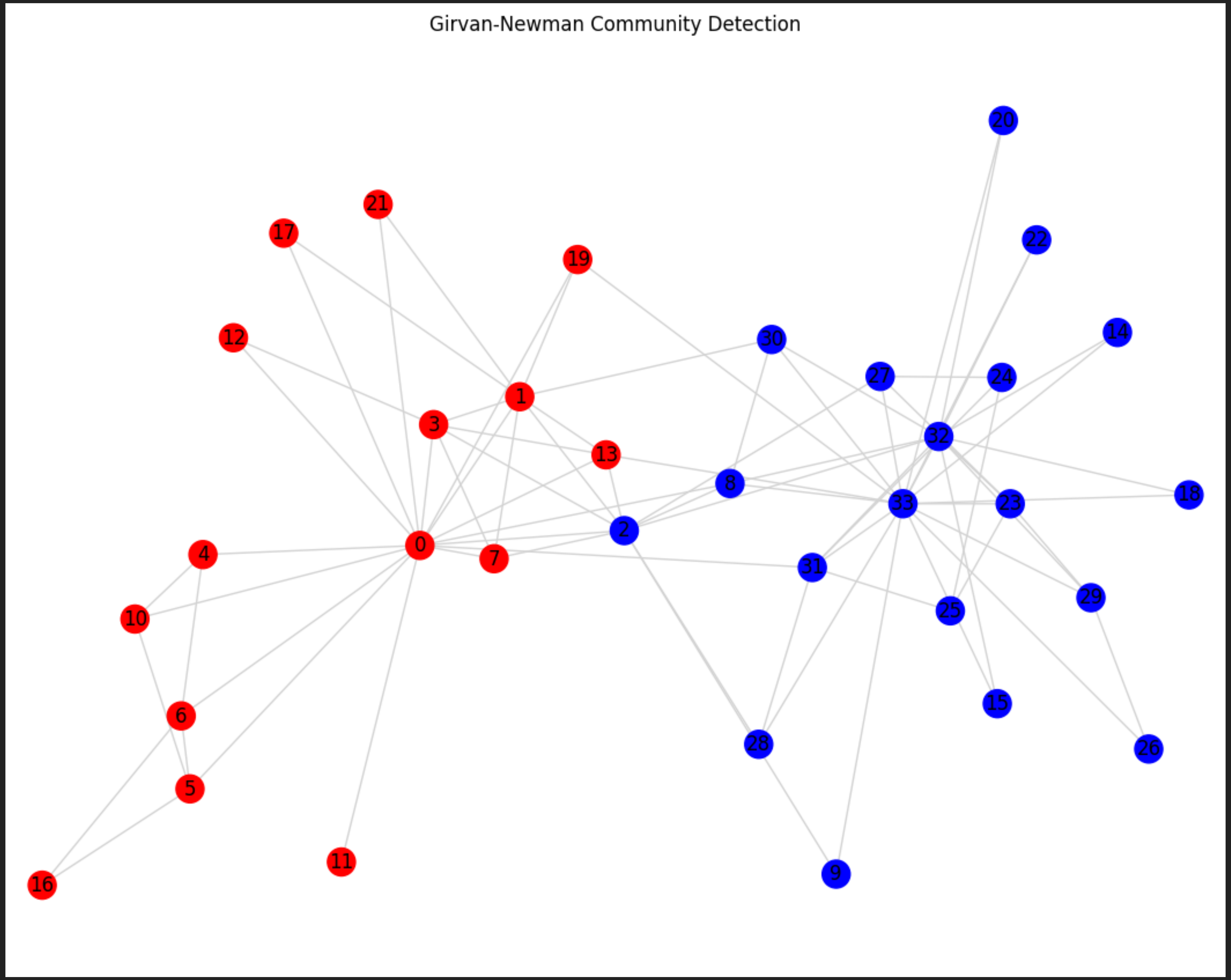
# PAGERANK

- Evaluates importance of individuals in network by analyzing number of connections received from others, similar to search engines assessing web pages. This identifies key members playing vital role in sustaining the network's connectivity and influence.



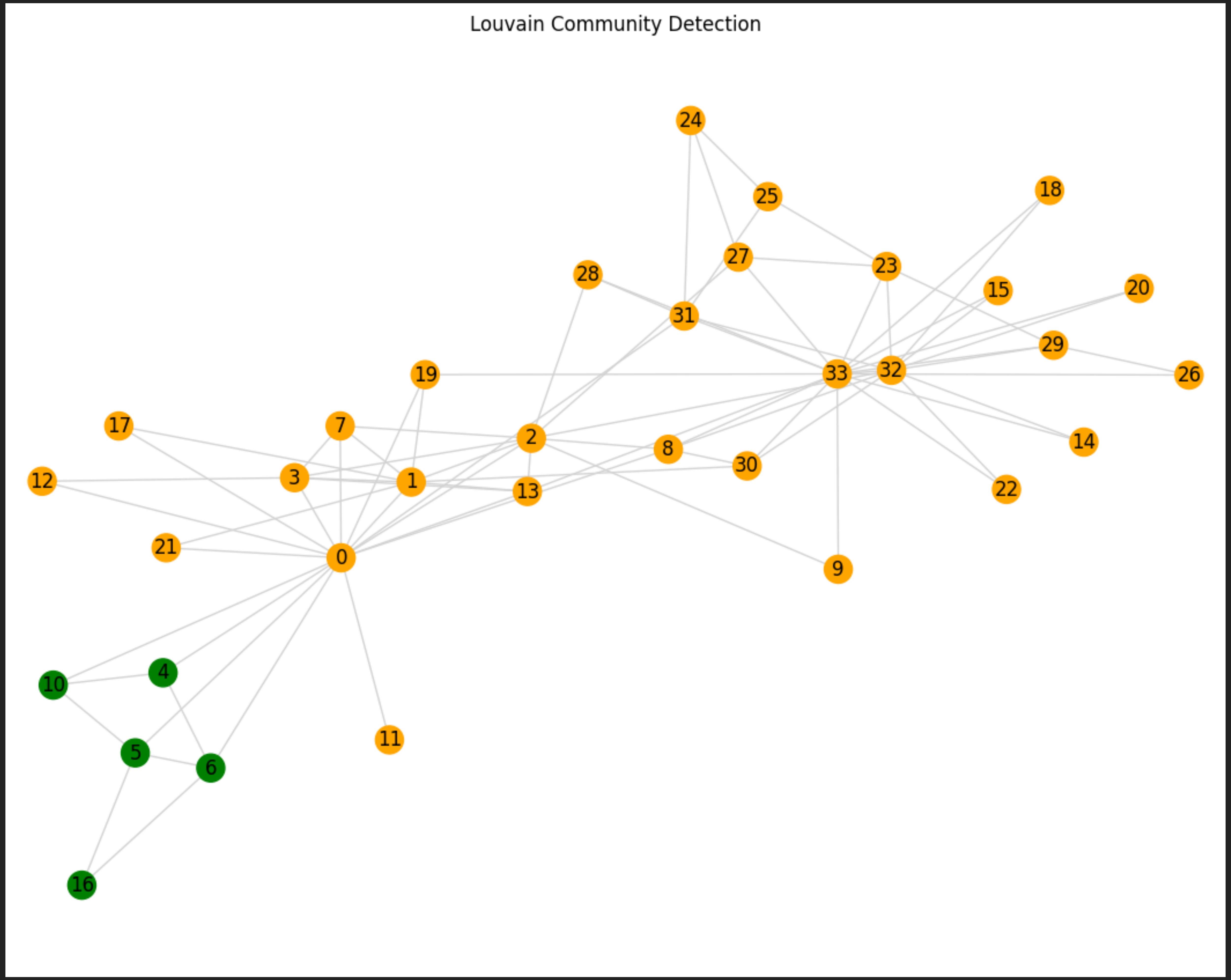
# GIRVAN-NEWMAN

- Identifies communities within a network by systematically removing edges connecting different groups, particularly those with highest betweenness centrality. This uncovers natural clustering of users, revealing how closely connected subgroups emerge within a network.



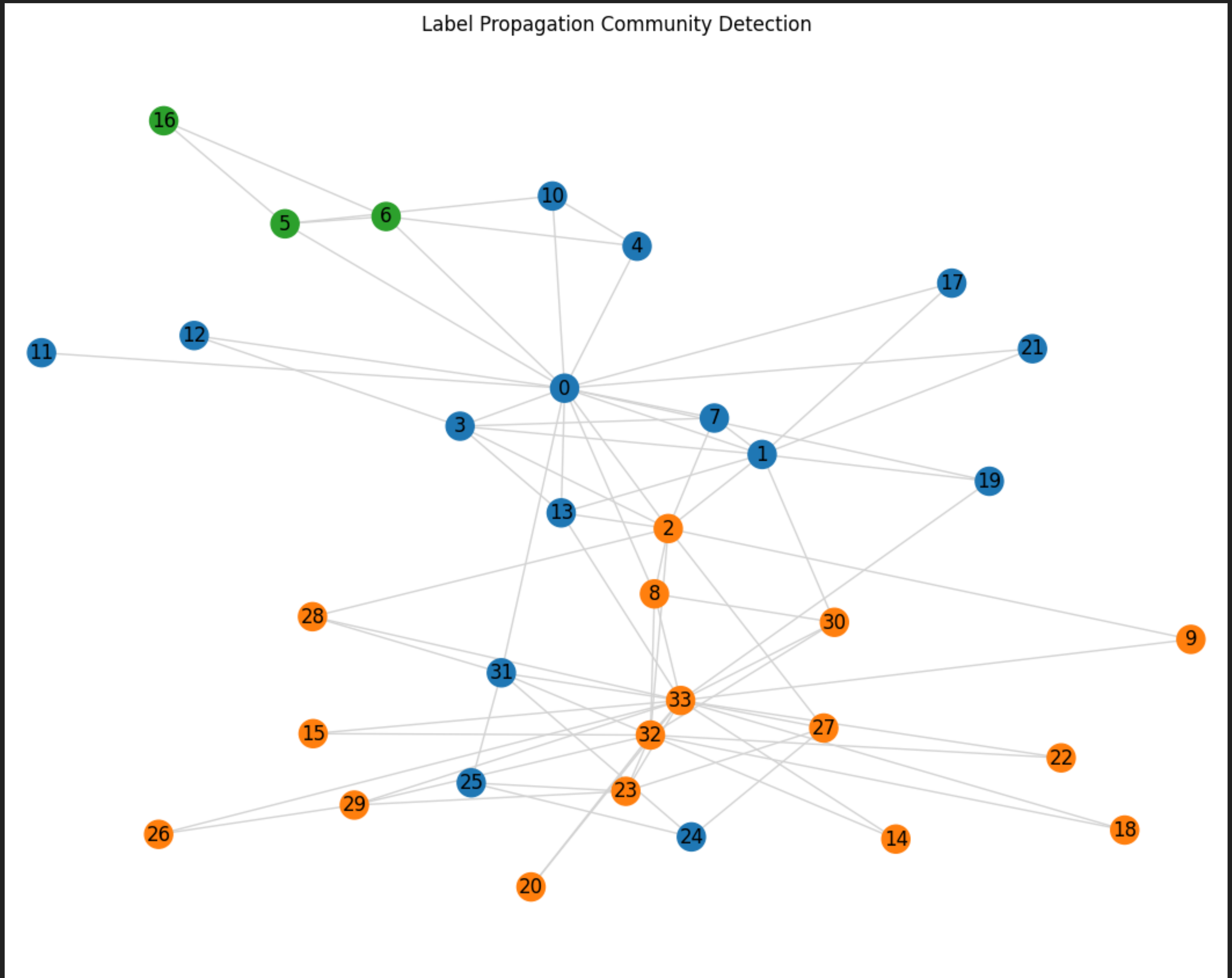
# LOUVAIN METHOD

- Identifies groups within network by detecting densely connected clusters and maximizing modularity to improve community detection. Valuable for understanding the hierarchical relationships within communities and the overall structure of the network.



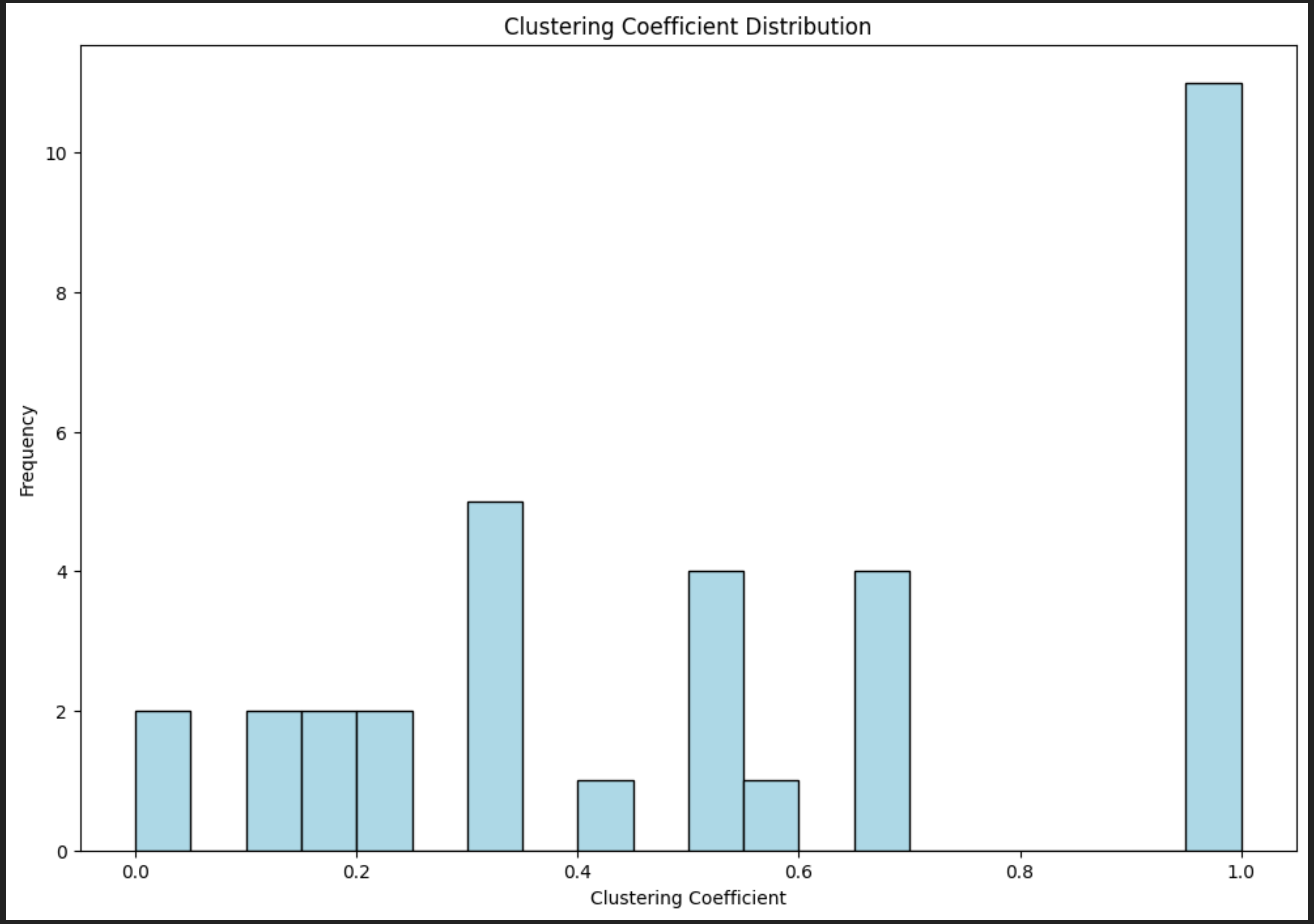
# LABEL PROPAGATION

- Identifies communities by assigning labels to nodes based on neighboring nodes. This leverages local interactions to facilitate the natural formation of communities, efficient for large networks and offering insights into the social dynamics governing relationships within graph.



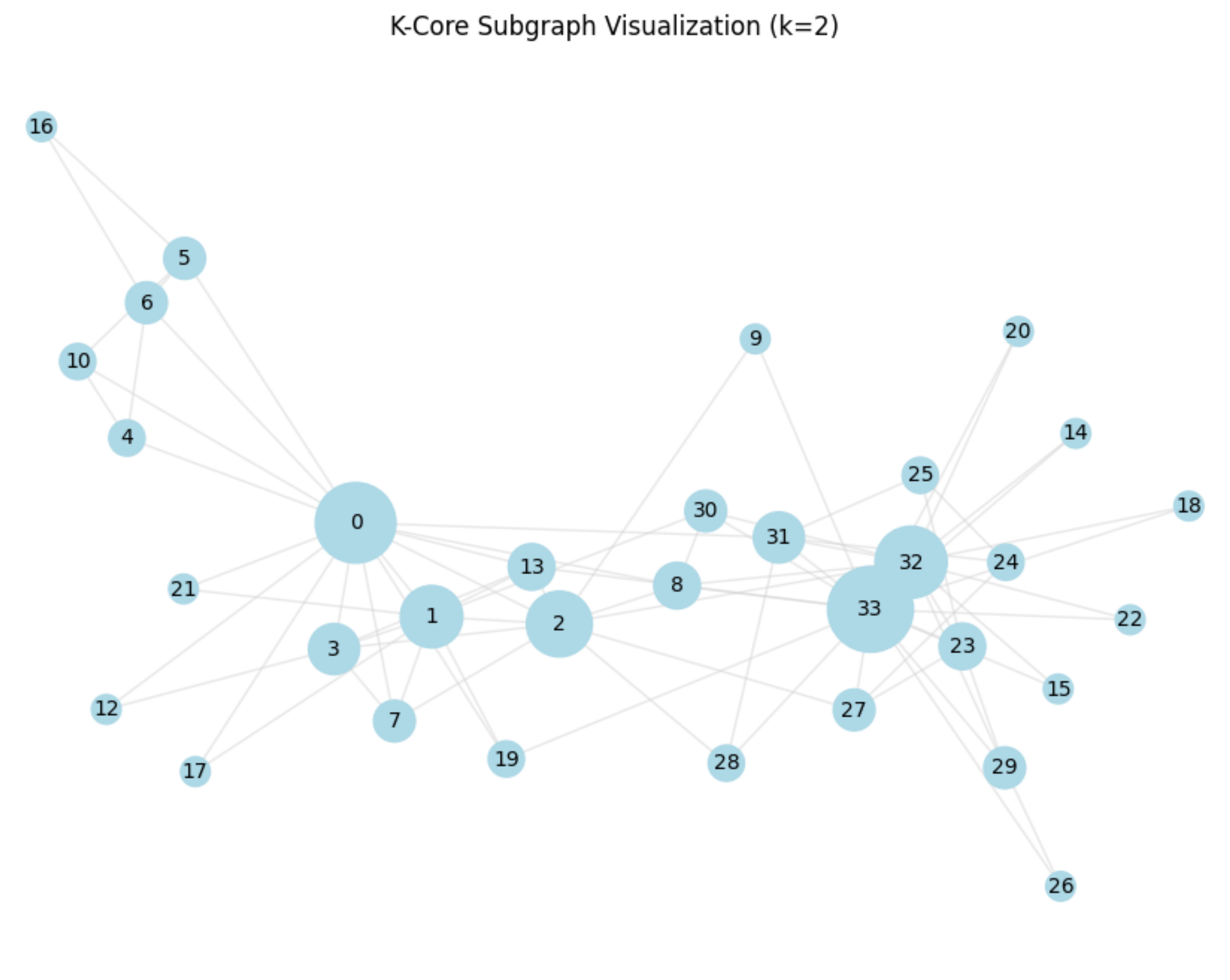
# CLUSTERING COEFFICIENT DISTRIBUTION

- Reveals extent of connectivity among groups of individuals within network. Average clustering coefficient of 0.5706 indicates many individuals formed tightly-knit groups, reflecting strong local connections.



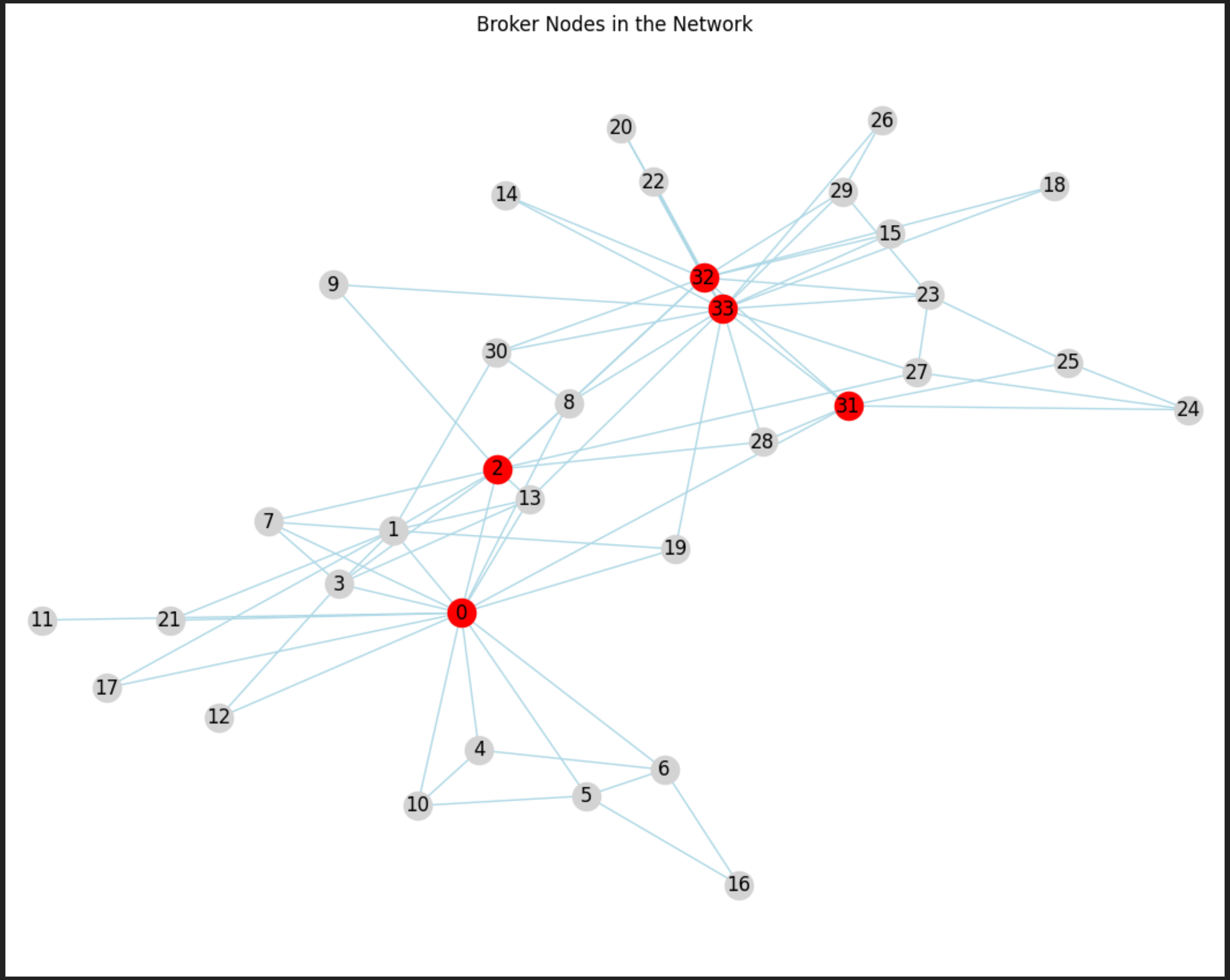
# K-CORE DECOMPOSITION (USER INPUT K)

- Breaks down network into layers where each node has at least k connections. Helping identify most connected and influential members, offering insights into the network's stability and community structure.



# BROKER NODES

- Key individuals linking separate communities, enabling interactions and information exchange between them. Identifying these helps understand influence within network, as they play vital role in bridging distinct groups and facilitating communication.



# INFERENCES

- Understand how people within a group connect, helping improve connections and solving conflicts within the group.
- Identifying influential nodes could amplify communication strategies and enhance information dissemination within network.
- Clustering observed indicates communities are tightly knit, with strong local connections facilitating quicker and more effective sharing of information among members within network.

# THANK YOU

Varang Pratap Singh