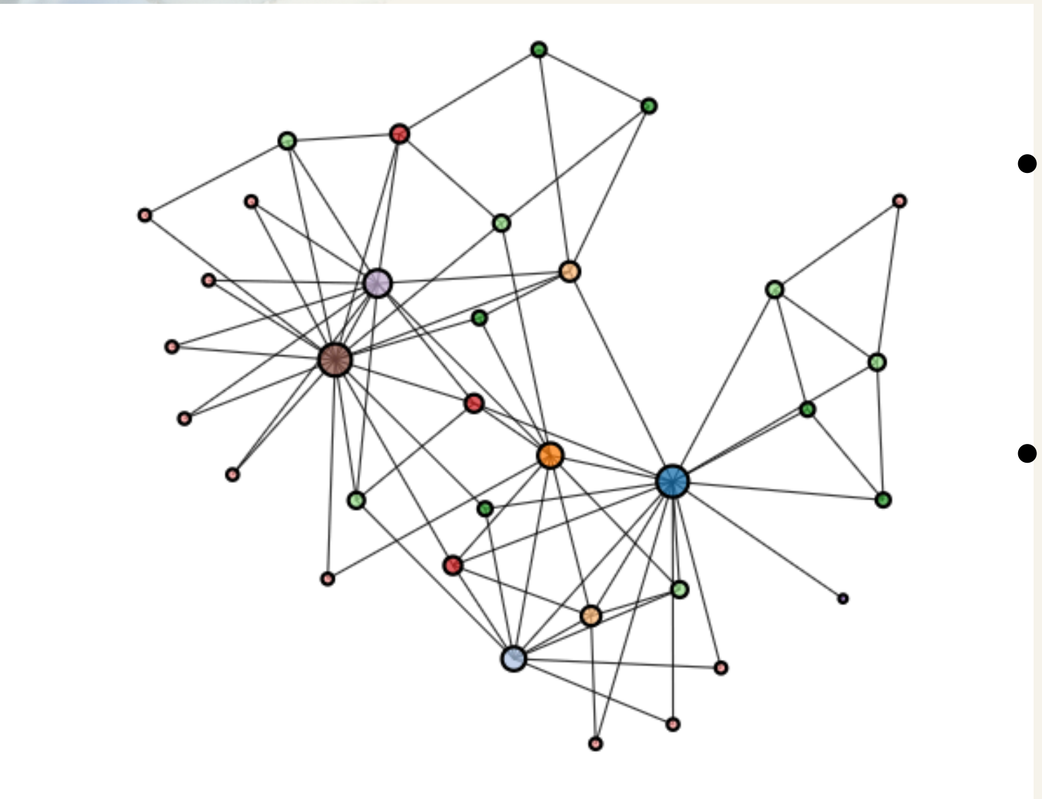
GRAPH AND SOCIAL NETWORK PROJECT

GROUP MEMBERS

- D SHANMUKHA 22BDS019
- G LEELADITYA 22BDS024
- RAJDEEP MANIK 22BDS048
- VANSH LAL TOLANI 22BDS061

ABOUT THE DATASET



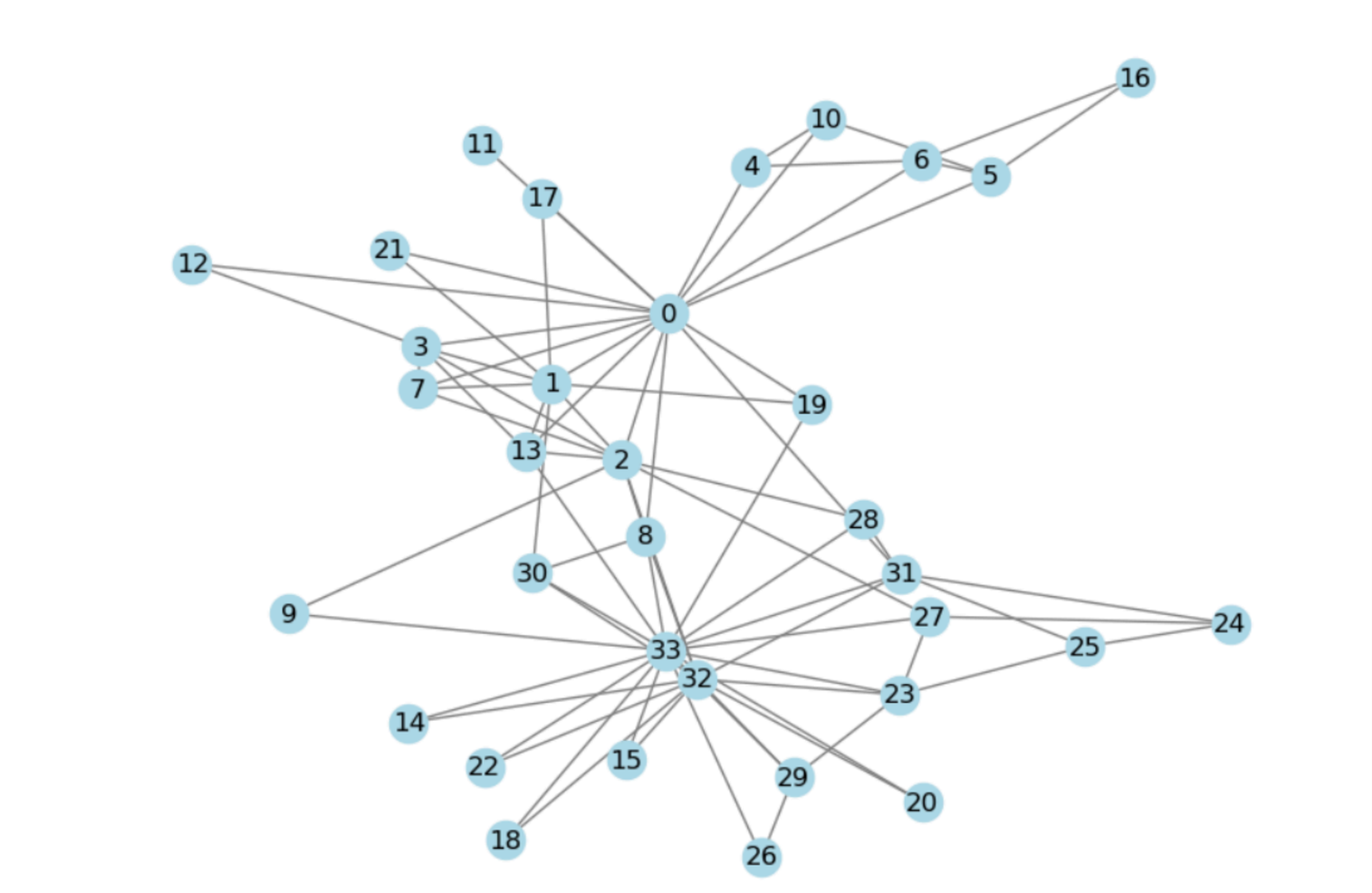
The Karate dataset

- The Karate dataset is a well-known social network dataset that represents the friendships among members of a karate club at a US university in the 1970s.
- It was originally studied by Wayne W. Zachary, who analyzed the social interactions among the members to understand the dynamics of the group Nodes

ABOUT THE DATASET

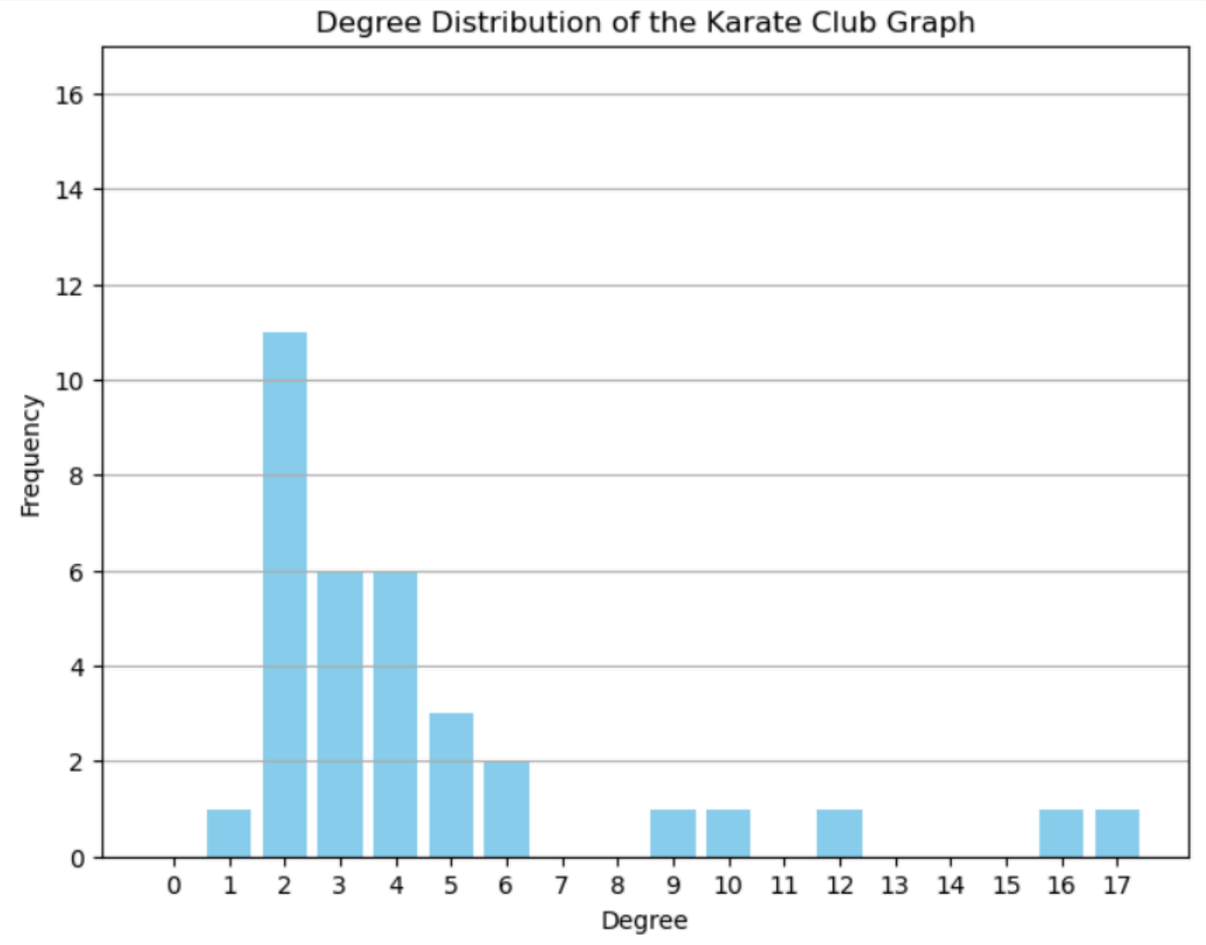
- Each node represents a member of the karate club. The dataset consists of 34 members.
- Edges represent friendships between the members. If two members are friends, there is an edge connecting them.
- Graph Representation:
- The dataset can be represented as an undirected graph, where nodes represent members and edges represent the friendships.
- Graph has 34 nodes and 78 edges.

Graph Representation



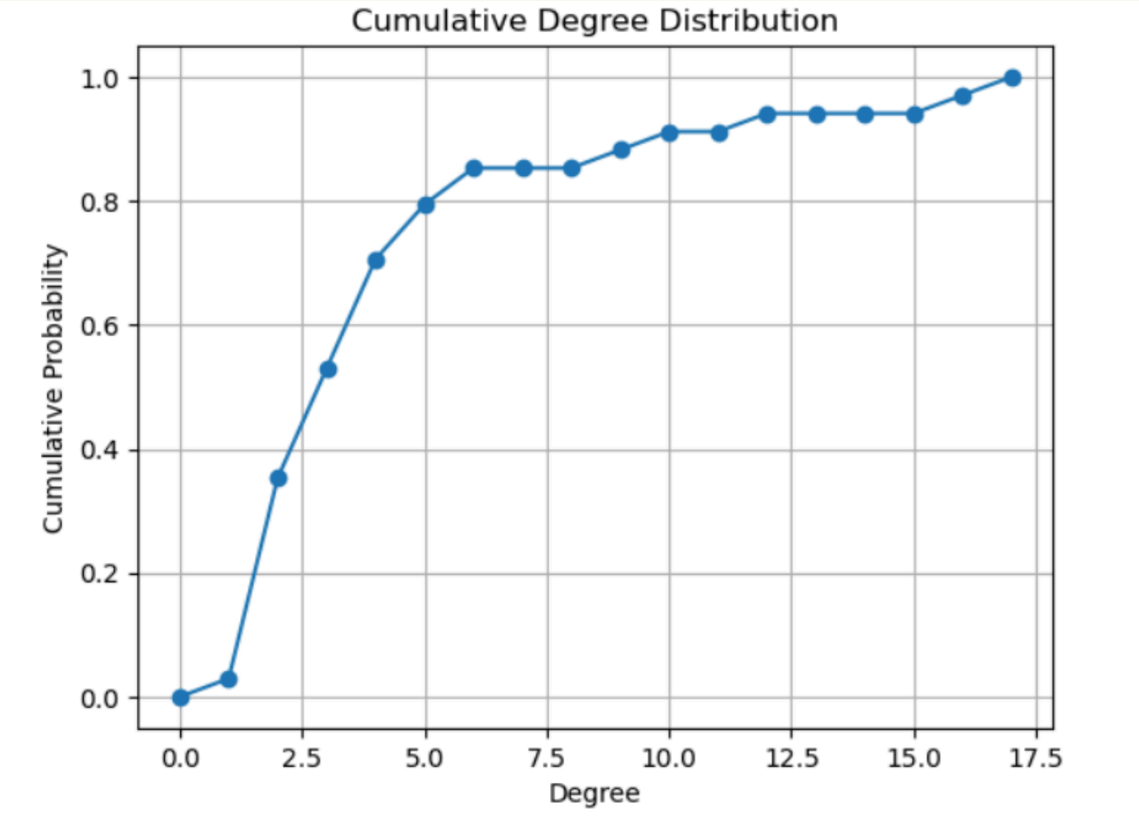
Degree Distribution

- Node 33 has highest degree(17)
- Node 11 has lowest degree(1)



cumulative degree distribution

- Connectivity Pattern: The cumulative degree distribution indicates how well-connected nodes are within the network. $P(X \geq 3)=0.65$
- Influential Nodes (Hubs): The distribution helps highlight highly connected nodes . For example, only around 2.9% of nodes have a degree of at least 17, indicating that just one or two people are exceptionally well-connected, likely representing leaders or influential members.

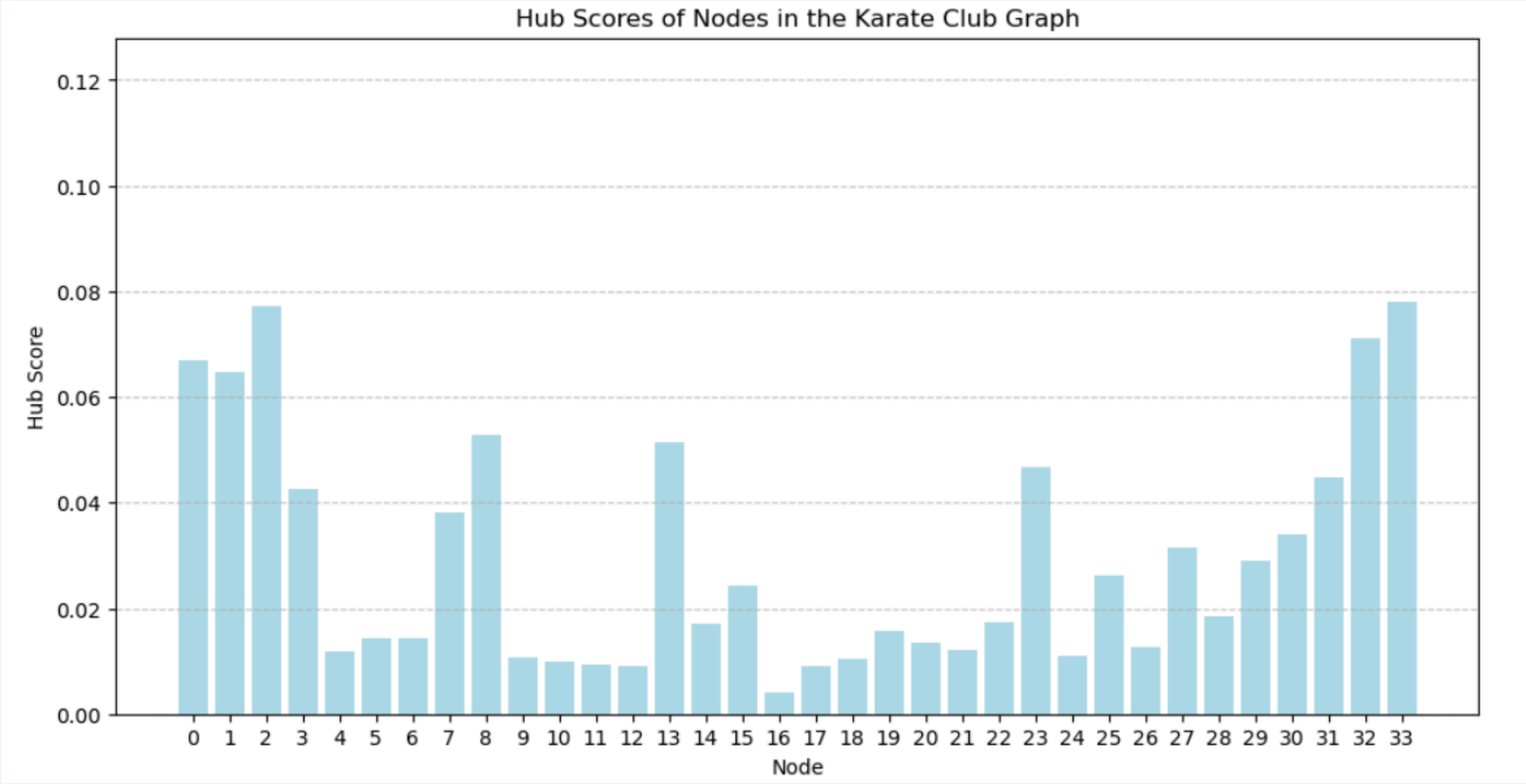


HITS Algorithm

- Authority Scores: In the karate club context, a member with a high authority score is likely to be considered an expert or a central figure within the club, receiving recognition and respect from others. They may be the go-to person for knowledge or advice.
- Hub Scores: In this dataset, a member with a high hub score actively connects to many others, potentially promoting interaction, communication, and information sharing.

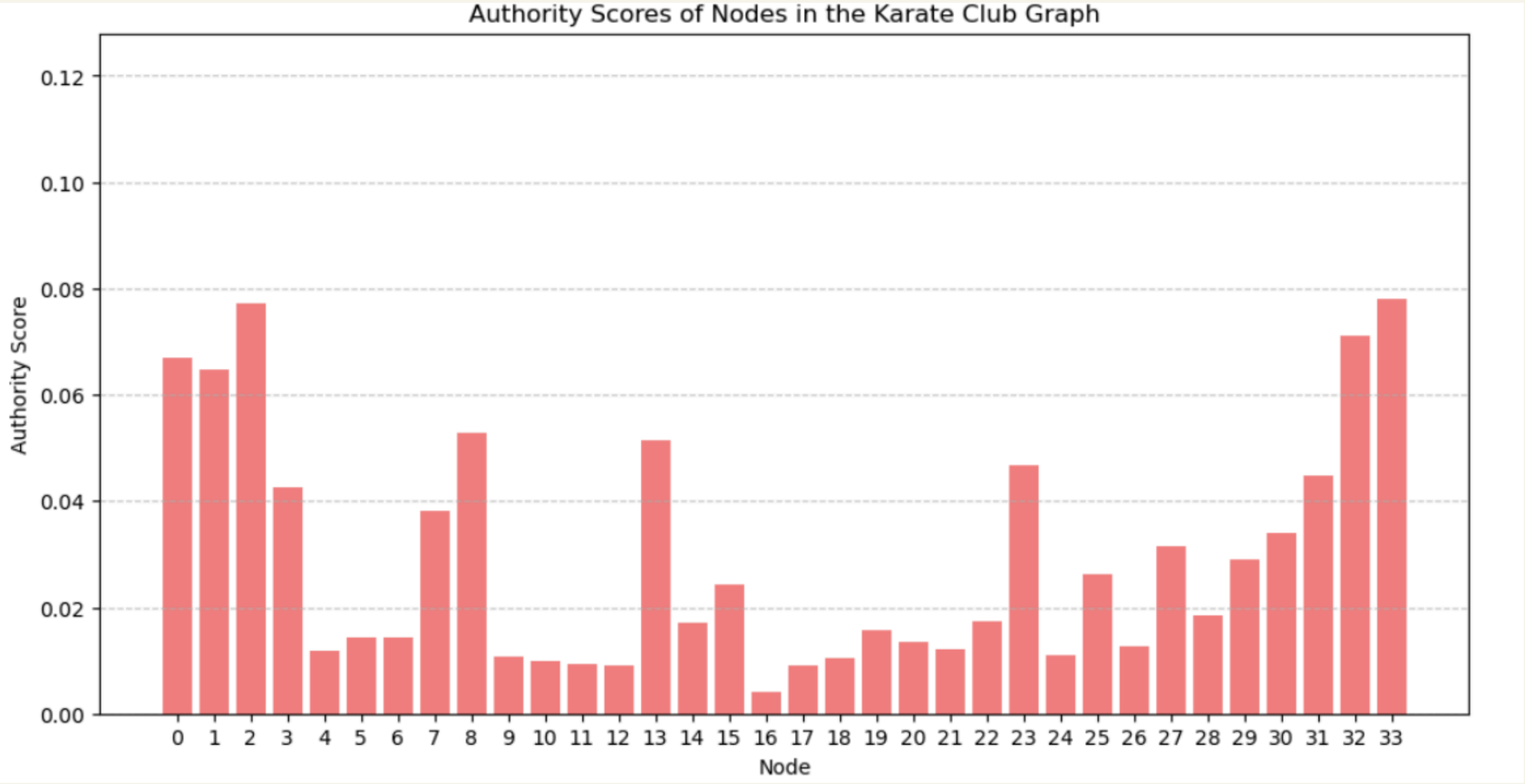
Hub score

- Node 33 has higher hub score
- Node 16 has lower hub score



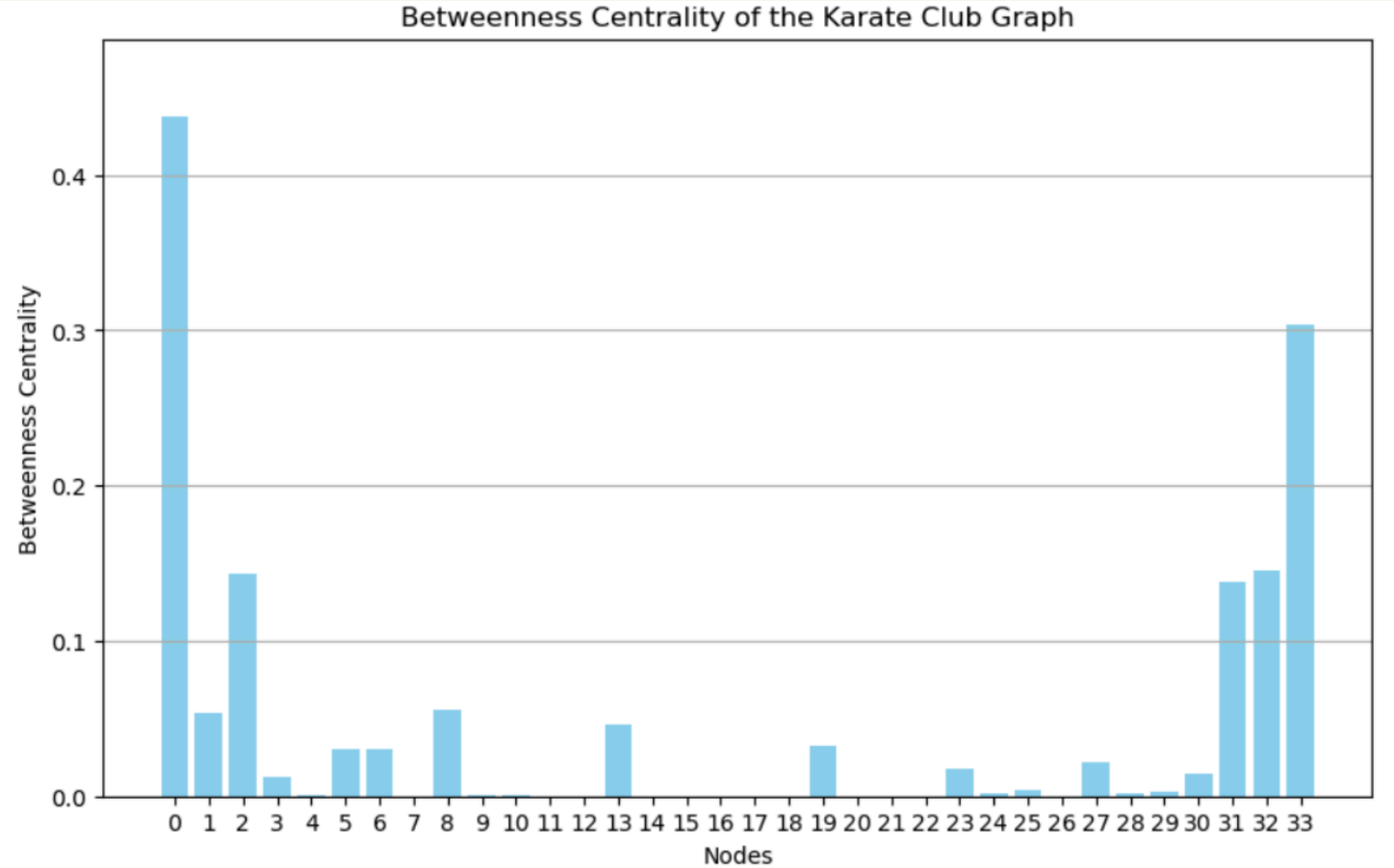
Authority score

- Node 33 has higher authority score
- Node 16 has lower authority score



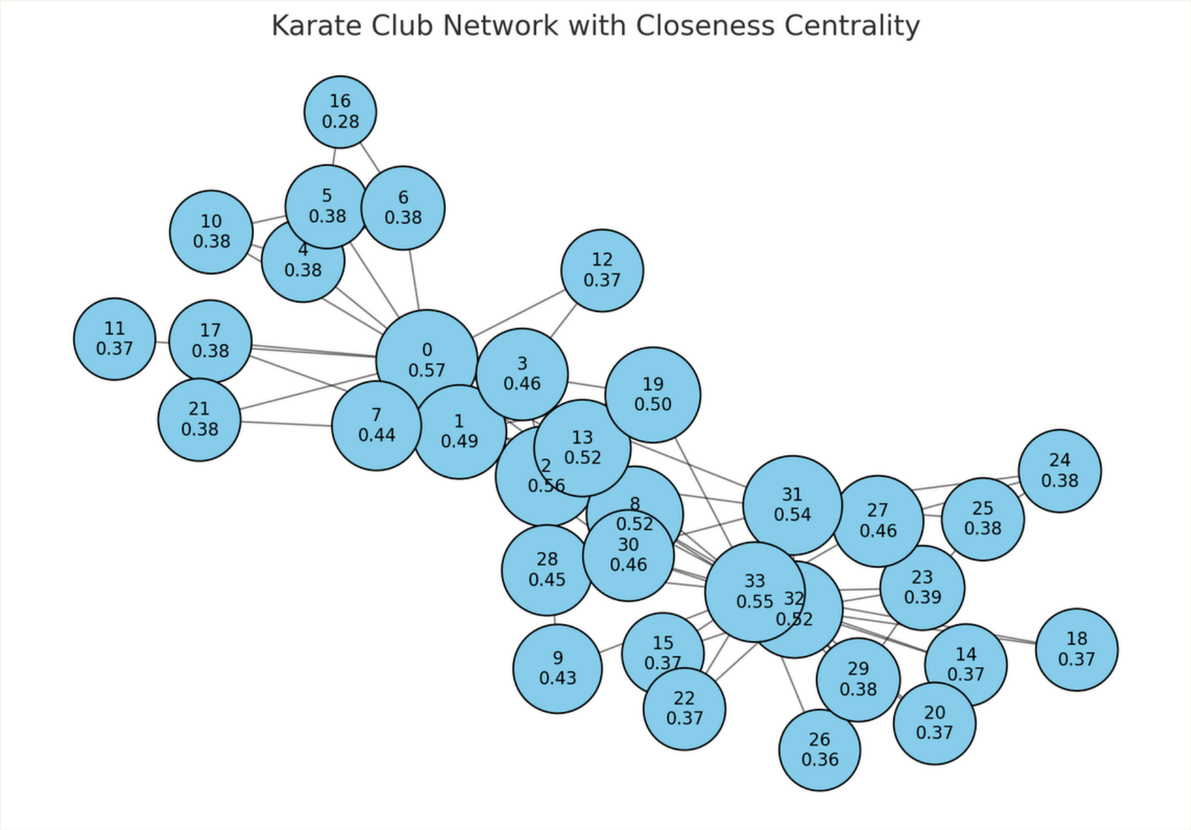
Betweenness centrality

Nodes 0 and 33 have the highest betweenness centrality, indicating that they are the primary "bridge" nodes in this network. This aligns with their known roles as influential figures in the club, likely serving as connectors and facilitators between other members.



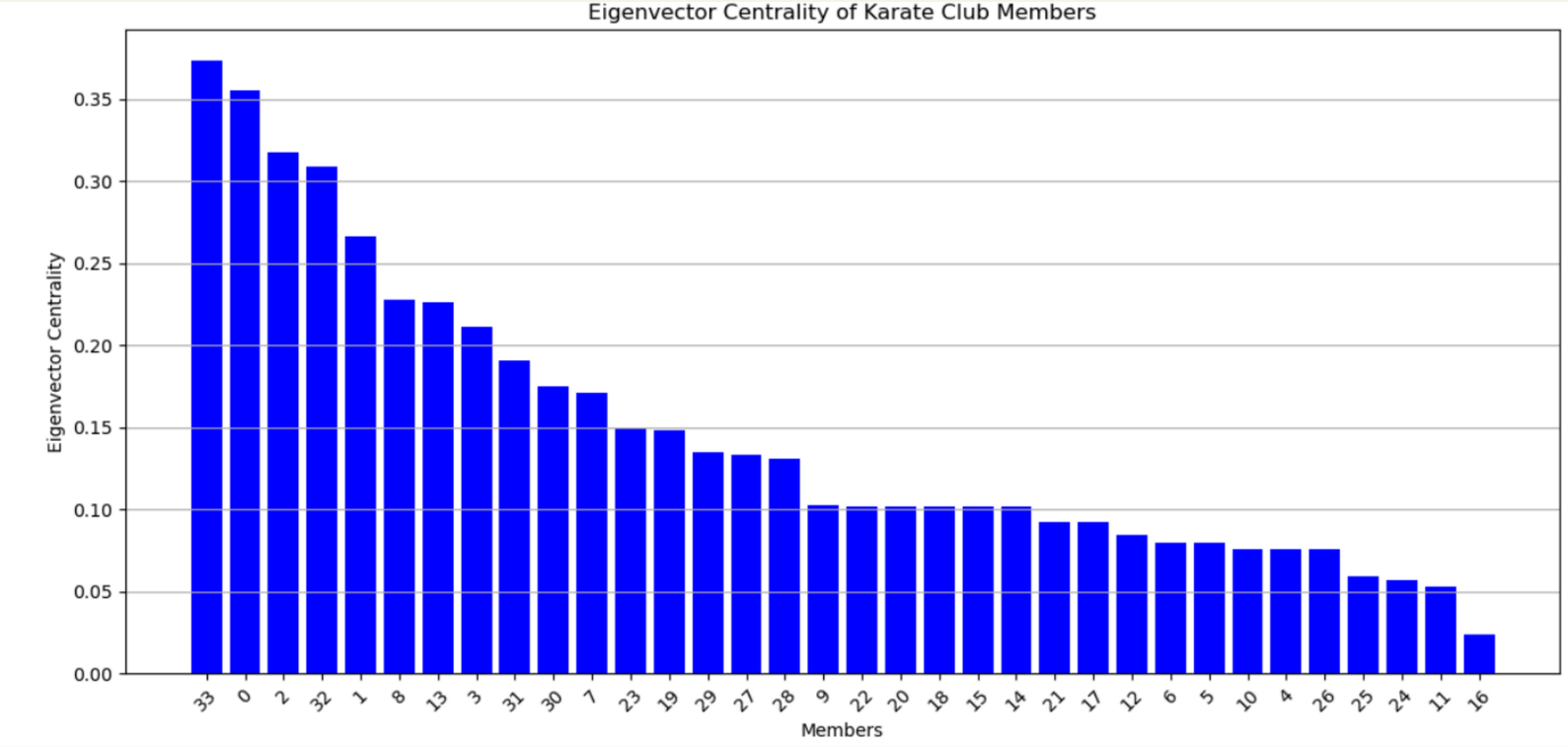
Closeness Centrality

Node 0 has the highest closeness centrality, suggesting that it is well-connected and can reach others in the network quickly, followed closely by nodes like 2 and 33.



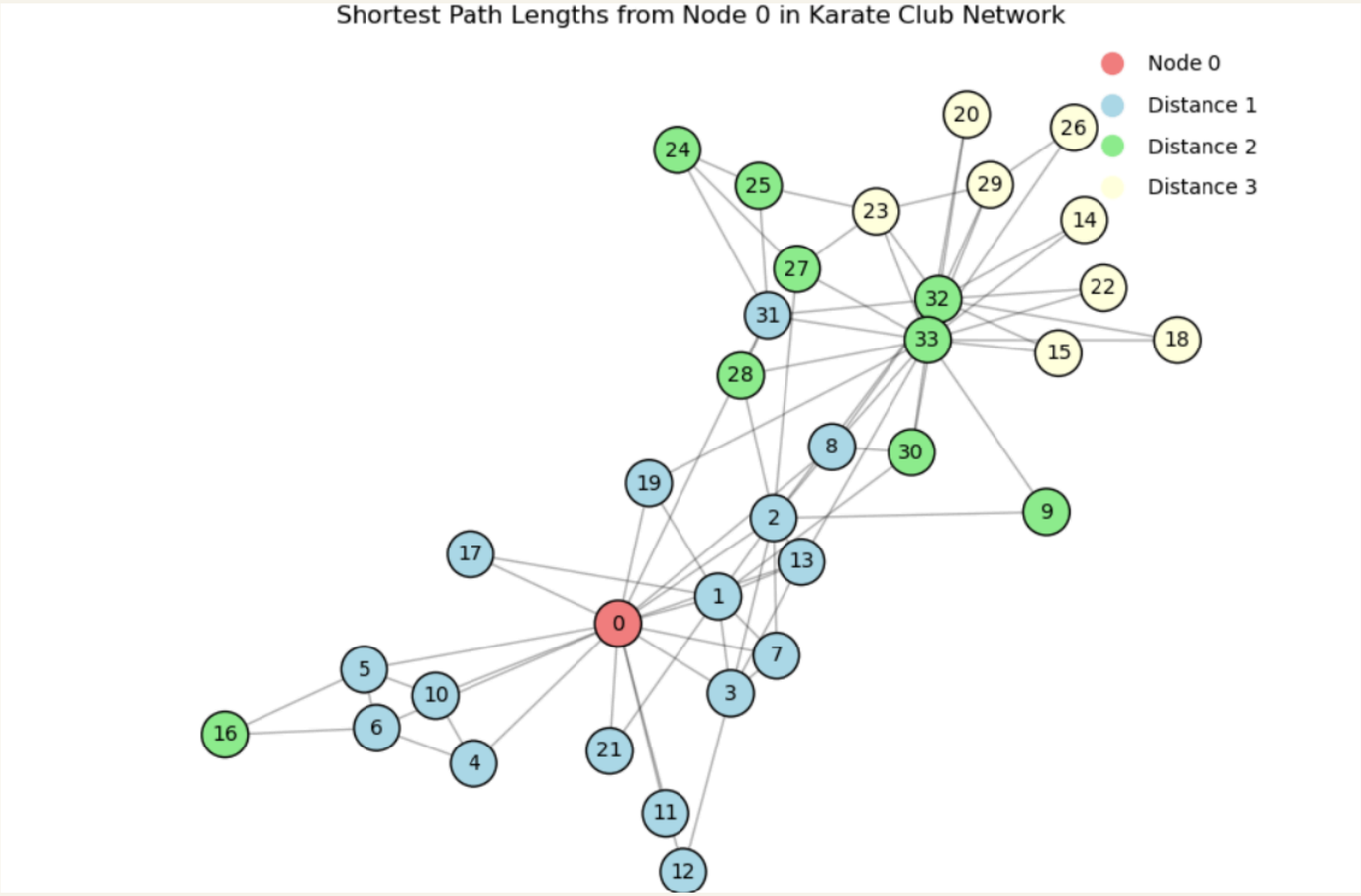
Eigen vector centrality

- In the karate club context, higher eigenvector centrality might indicate members who are key leaders or influential figures within the group
- Nodes 0 and 1 have the highest eigenvector centrality scores, indicating they are the most influential within the club.



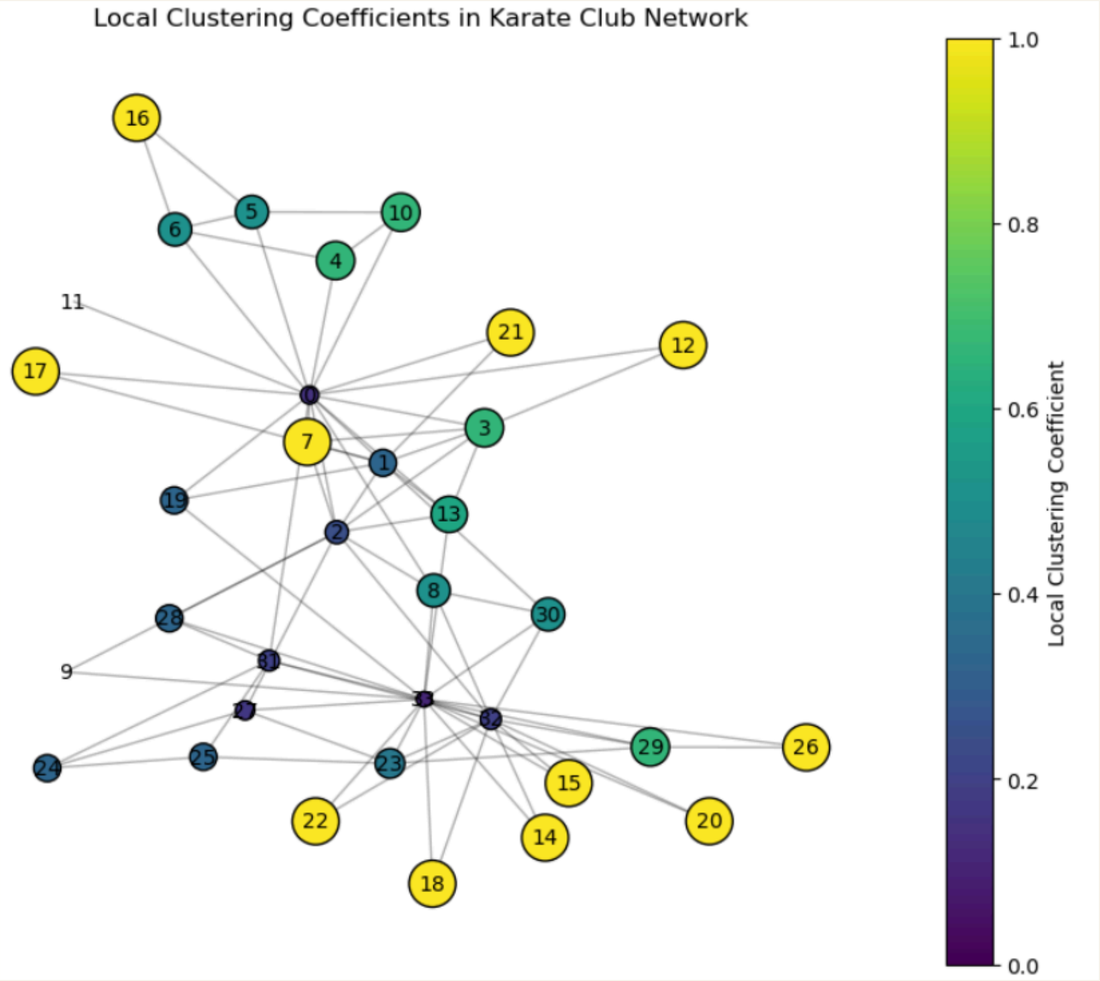
Shortest path length(From 0)

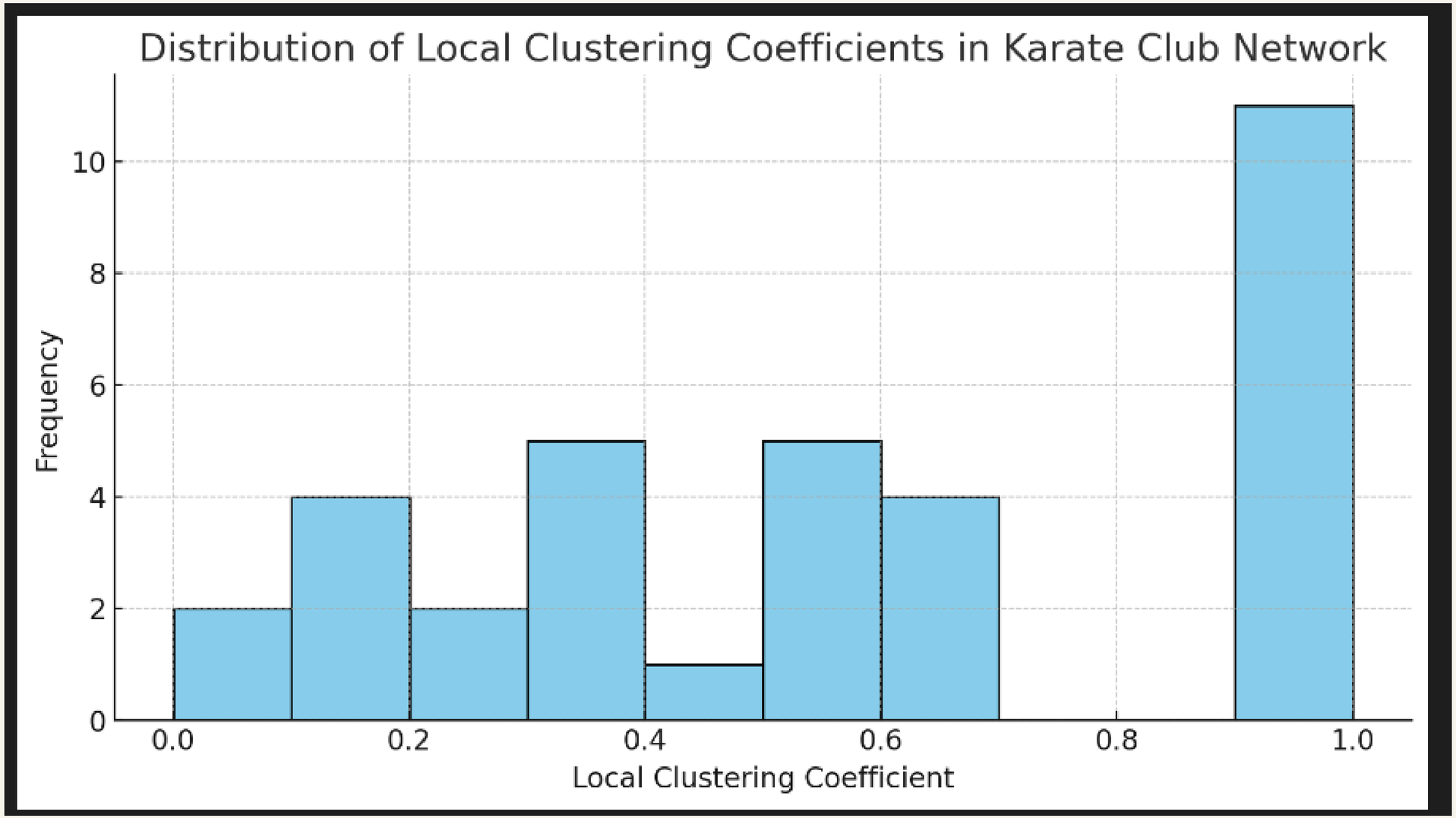
This shows that Node 0 is highly central in the network, as most other nodes are within 1 to 2 steps from it.



Local clustering coefficient

- In the karate club dataset, the local clustering coefficient of a node represents how connected that node's neighbors are to each other.
- Nodes with a High Clustering Coefficient (1.0):
  Nodes 17, 12, 26, 22, 21, 20, 18, 16, 14, 15, and 7 have a clustering coefficient of 1.0. This indicates that these nodes' neighbors are fully connected with each other, forming a tightly-knit group.



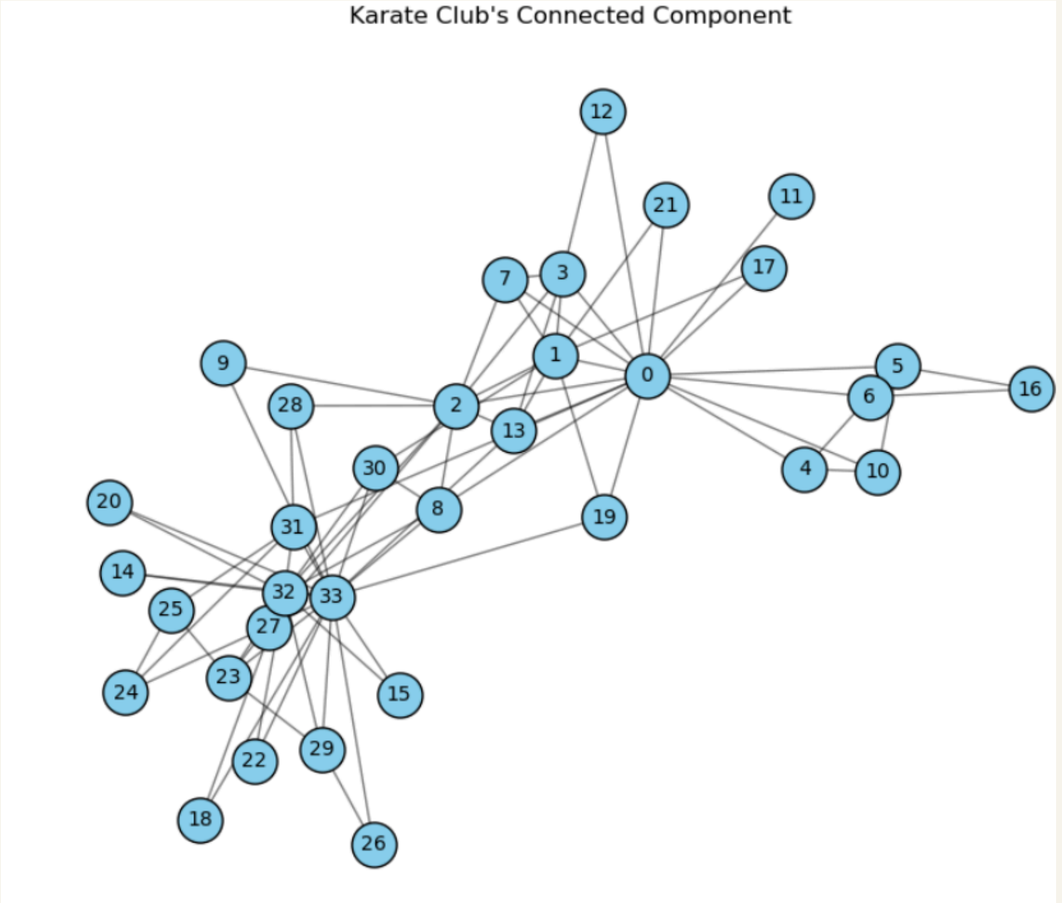

Distribution of Local Clustering Coefficients in Karate Club Network
Frequency
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
Local Clustering Coefficient

Global clustering coefficient

- The global clustering coefficient for the karate club dataset is approximately 0.256.
- The global clustering coefficient suggests that the karate club network has moderate cohesion, with a mix of tightly-knit groups and bridging nodes that connect them.

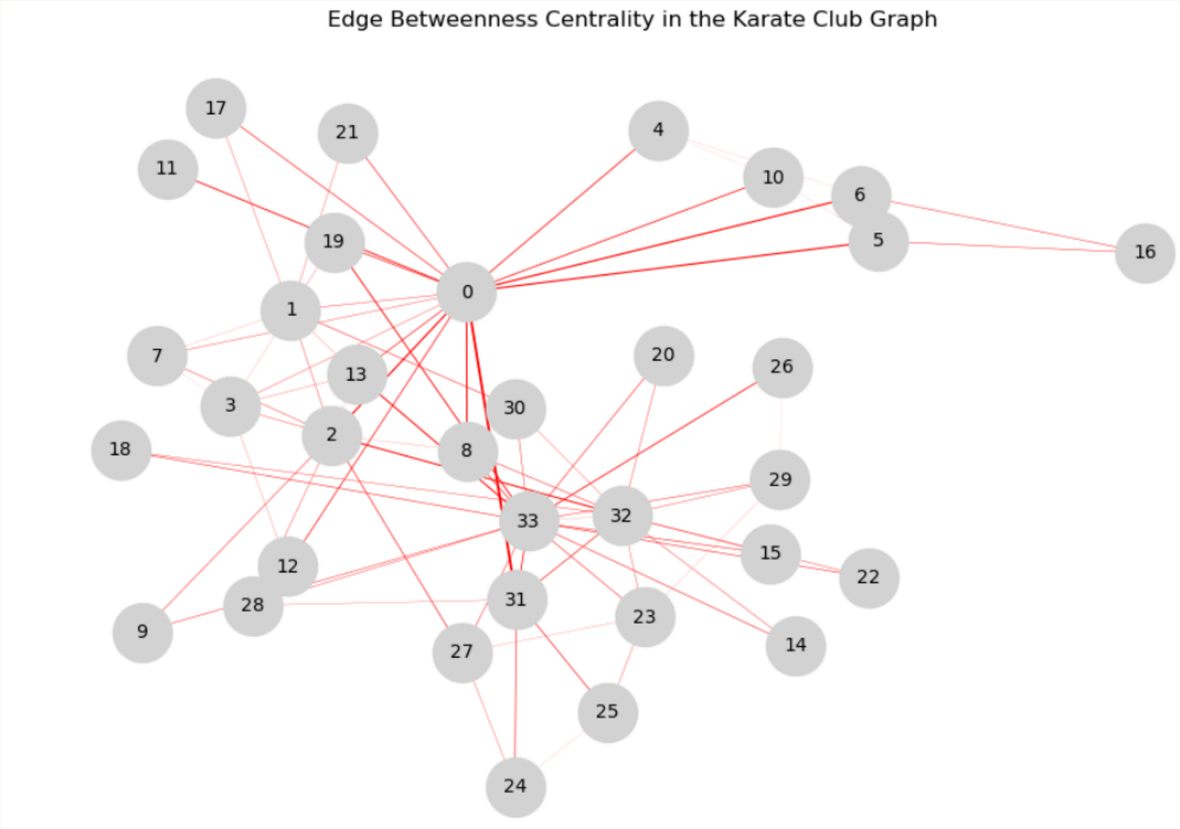
Connected Component

- One Connected Component:
  The entire network consists of a single connected component.
- Size of the Component: This component contains all 34 nodes in the dataset, meaning every member of the karate club is reachable, directly or indirectly, from every other member.



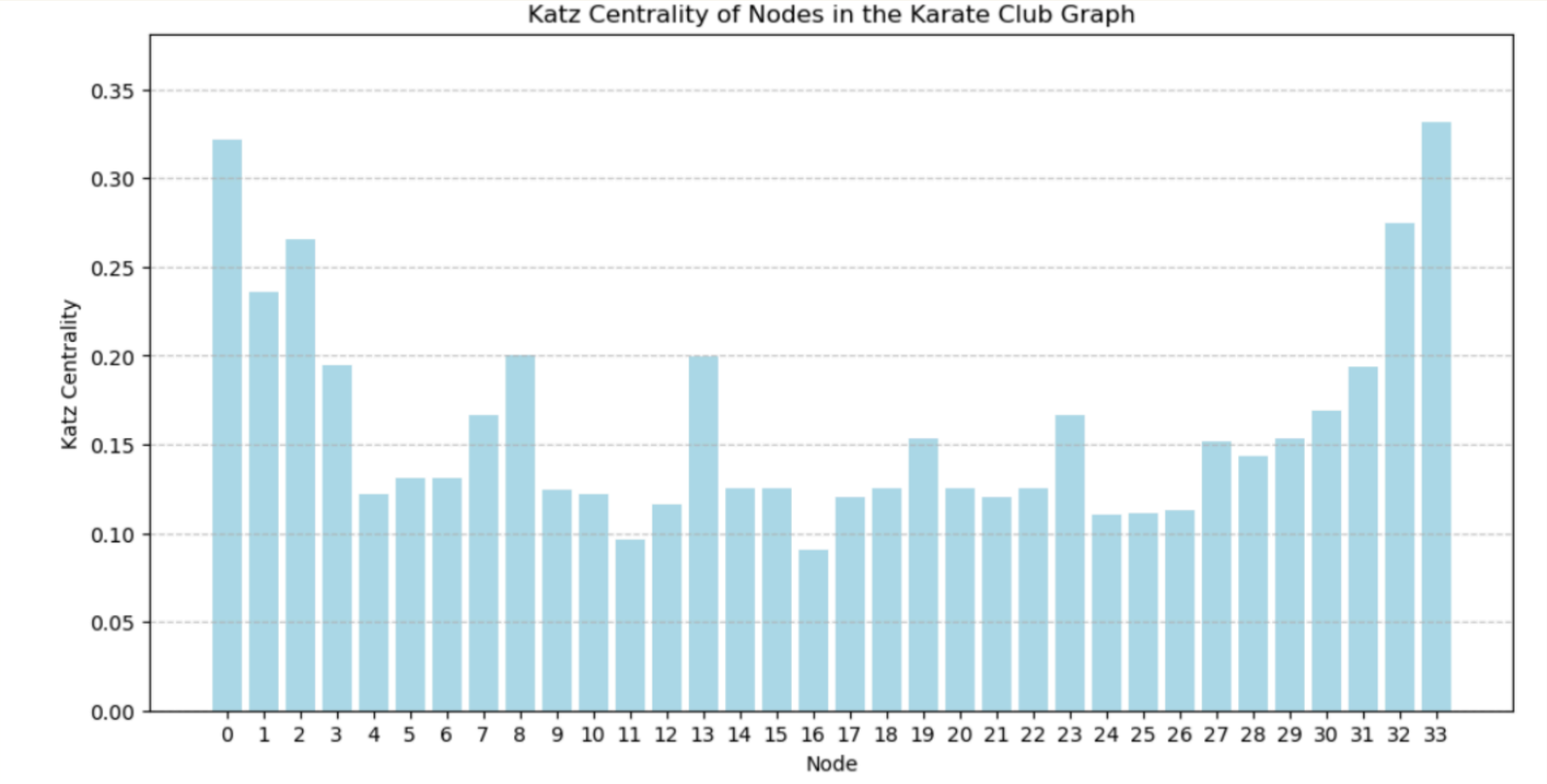
Edge Betweenness Centrality

- Thicker edges represent friendships that play a crucial role in connecting different members of the karate club.
- The visualization of edge betweenness centrality highlights certain friendships (edges) that are pivotal in connecting various members (nodes) of the karate club. Edges with high betweenness centrality serve as bridges between different groups, facilitating communication and interaction.



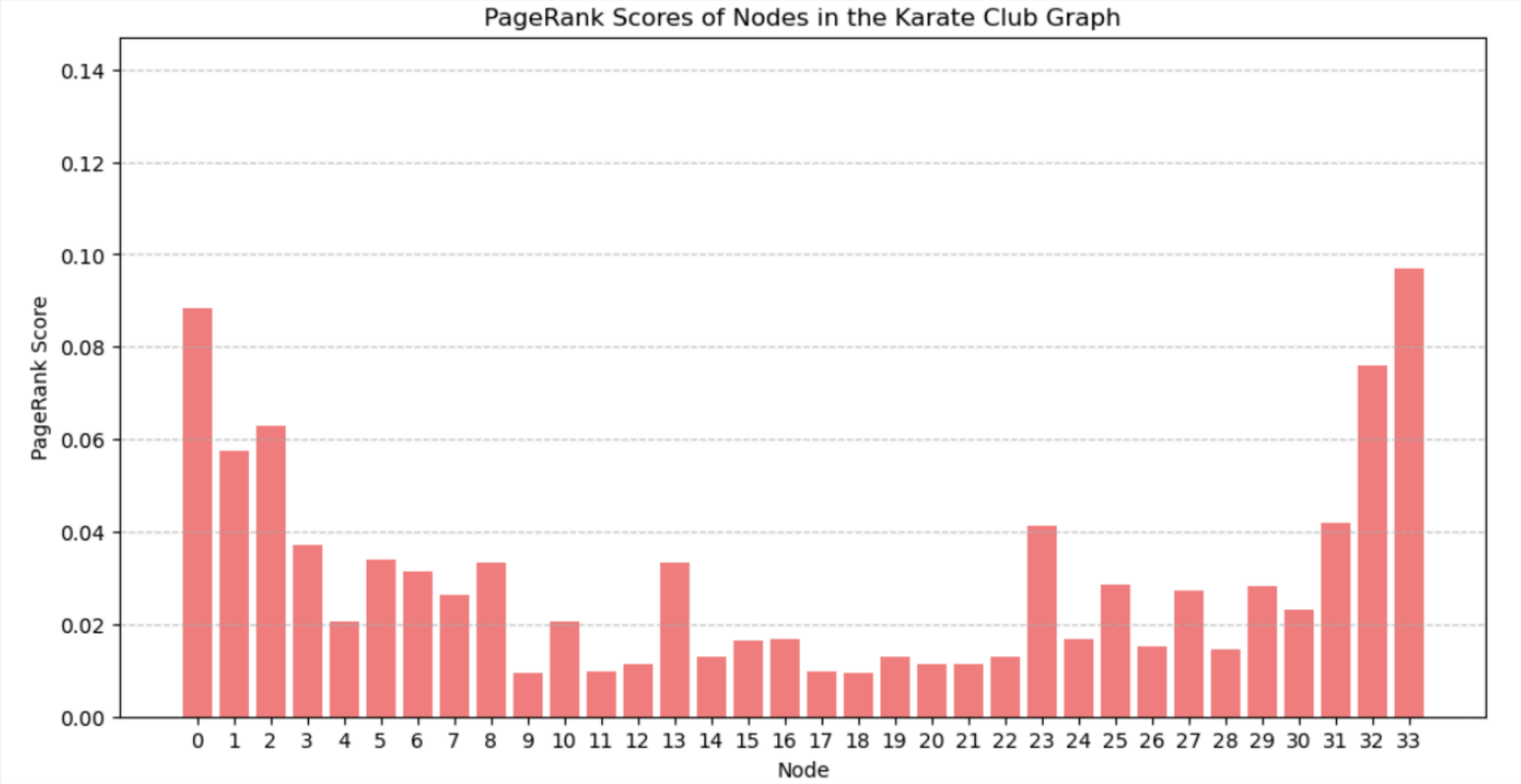
Katz cenrality

- Node 33 has higher katz centrality(0.3314) followed by node 0(0.3213)
- Node 16 has low katz centrality(0.0907)
- The Katz centrality visualization effectively illustrates the network's structure and the influence of individual members within the karate club.



Page Rank

PageRank scores indicate how important a member is within the karate club network based on their connections. A high PageRank score means that a member is not only well-connected but also connected to other influential members.

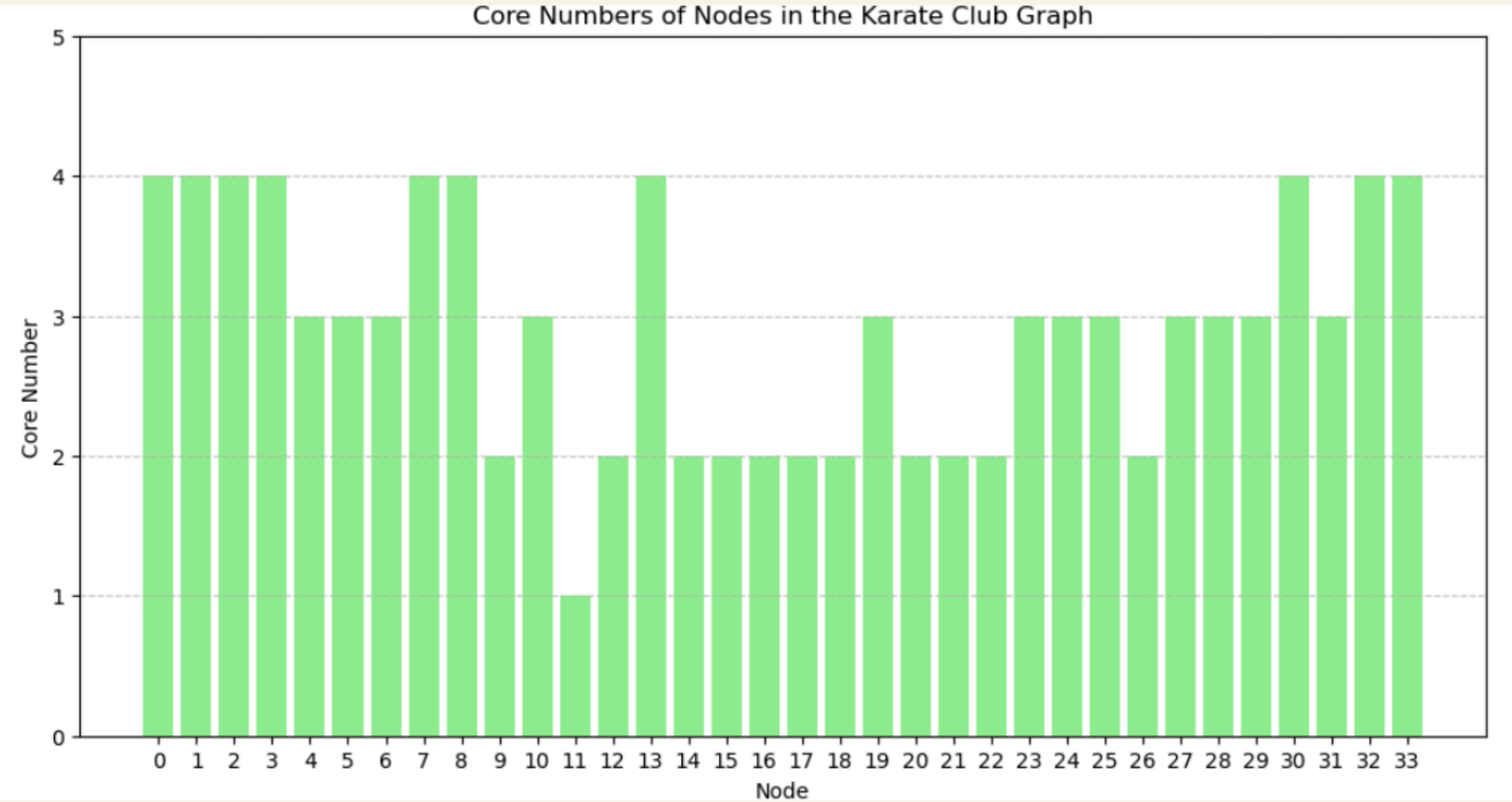


Node 33 has higher page rank(0.0970)

Node 9 and 18 has low page rank(0.0095)

Core number

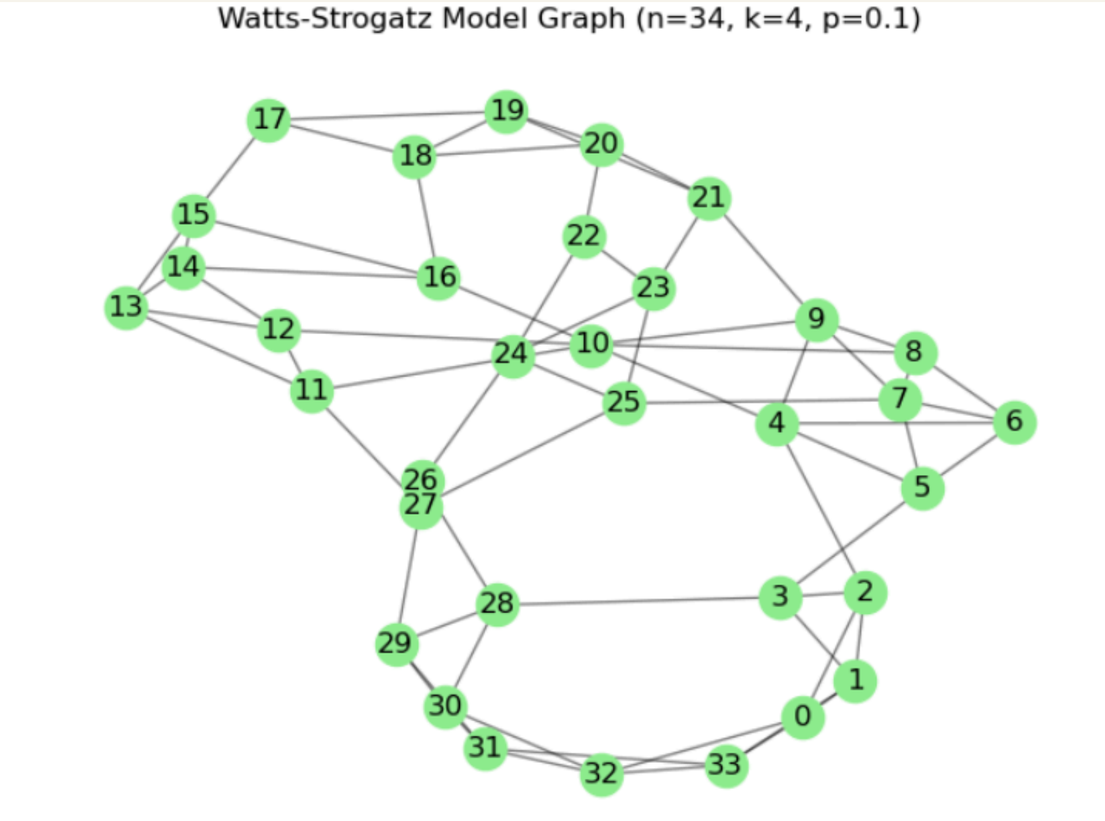
The core number indicates how many connections (edges) a member has with other members who are also relatively well-connected. A higher core number signifies that a member is part of a denser network of relationships.



Node 0,1,2,3,7,8,13,30,32,32 has highest core number of 4.

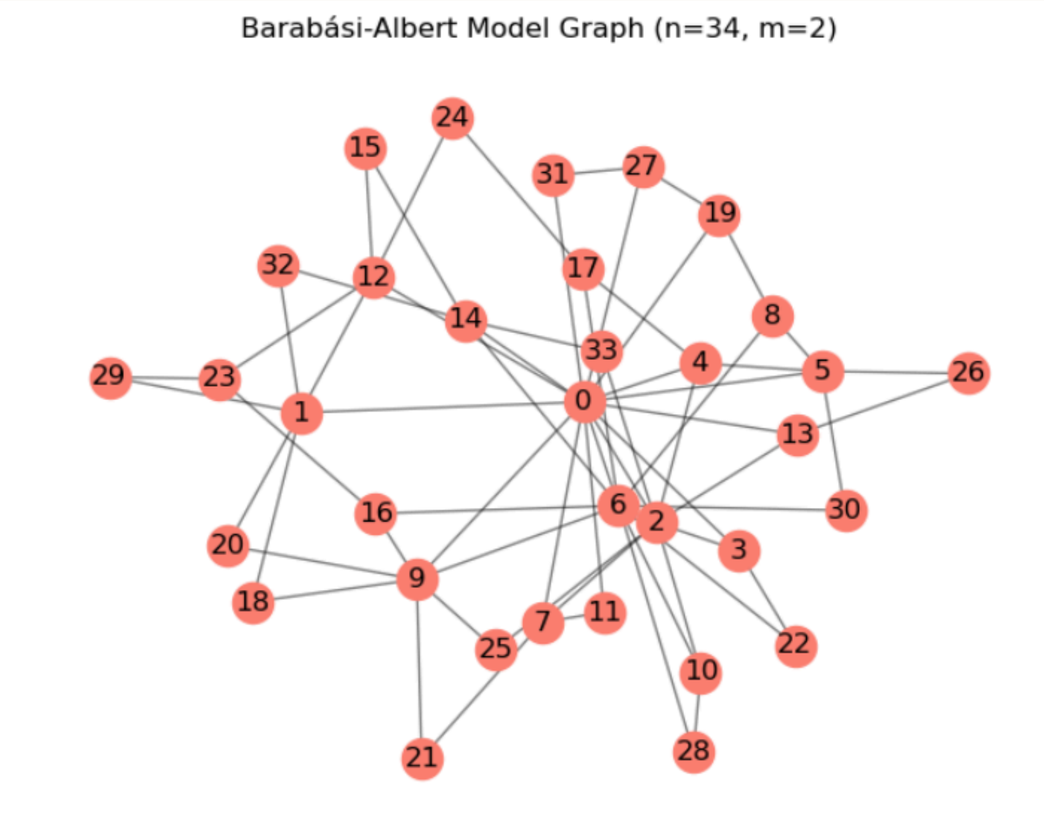
Watts-Strogatz Model

- Understanding Network Dynamics: The model can help simulate how friendships may form or dissolve over time in social networks.
- Predicting Behavior: By analyzing how individuals are connected in both the original and Watts-Strogatz-generated networks, one might be able to predict how changes in friendships could impact the overall network behavior.
- Community Detection: Using the generated networks, we could apply community detection algorithms to identify clusters of individuals with similar traits or interests.



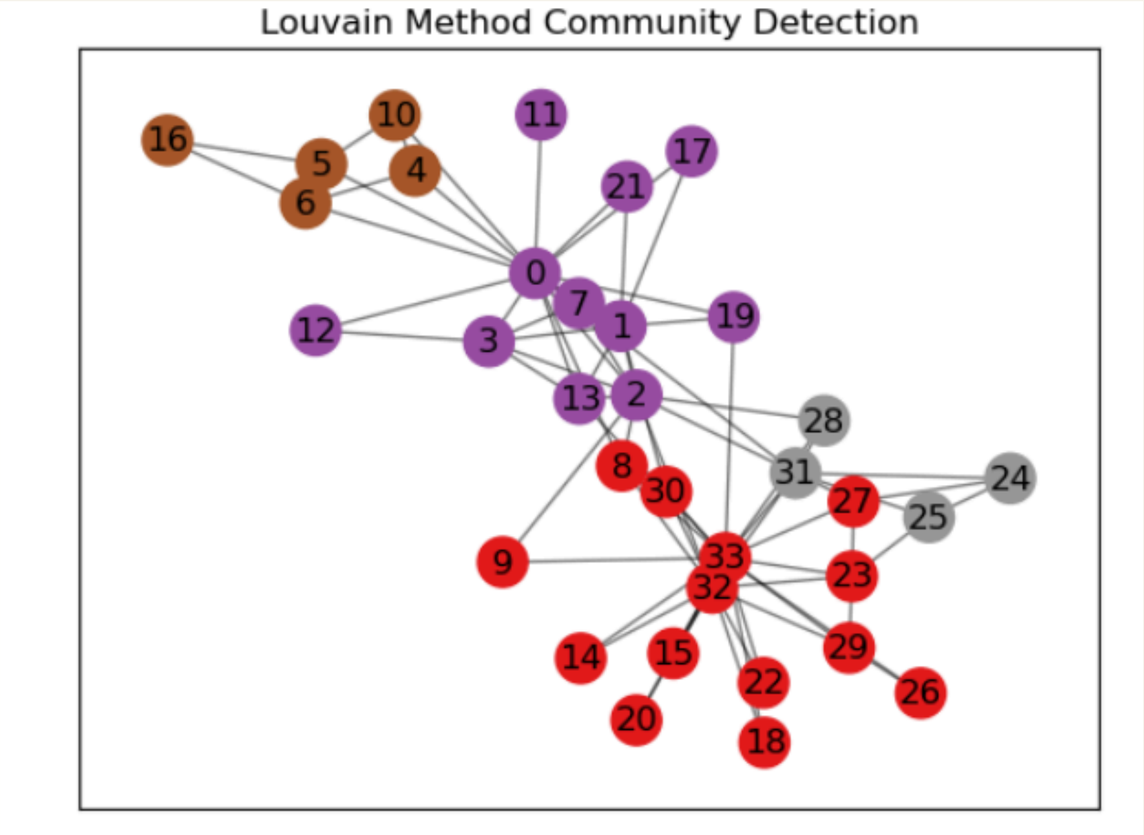
Barabási-Albert Model

- Understanding Network Dynamics: The model helps analyze how the karate club network might evolve over time. It can simulate how friendships develop in social networks as individuals join and connect with already well-connected members.
- Community Structure: By comparing the BA-generated network with the actual karate club graph, we can assess the similarity in community structure and identify how closely the karate club aligns with a scale-free network.
- Identification of Influential Members: The BA model can help identify key individuals (hubs) within the network. This can be useful for understanding the roles that certain members play in influencing social dynamics, information spread, or community cohesion.



Louvain Method

The Louvain method will reveal the underlying community structure within the karate club, identifying groups of members who interact more frequently with each other compared to those outside their community.



- Average Path Length of Karate Club Graph: 2.408
- Entropy of the Karate Club Graph: 1.9805
- Dunbar Number for the Karate Club Graph: 14

Thank You