Analyzing Zachary Karate Dataset using NetworkX

D Shanmukha 22BDS019
G Leeladitya 22BDS024
Rajdeep Manik 22BDS048
Vansh Lal Tolani 22BDS061

Abstract

This project is part of the analysis of social net- works, specifically focusing on Zachary's Karate Club dataset, which is a renowned social network representation detailing interactions among 34 members of a university karate club. Collected by Wayne W. Zachary in 1977, this dataset includes 78 edges that illustrate the social ties among members. The objective of this study is to investigate the social structure of the club and how internal conflicts led to the emergence of two distinct factions. We applied various network analysis techniques to extract insights into community structures, centrality measures, and the influence of individual members within the network. Our findings highlighted key influential nodes, particularly the club leaders, and illustrated community divisions through clustering algorithms. This dataset serves as an exemplary case for under- standing the dynamics of social networks and the methods used for community detection in network science.

Index Terms

Centrality Measures,Path and Reachability analysis,Clustering and community Detection,Network models Epidemic models

I. Introduction

THe study of social networks, particularly in the context of small groups, provides a unique perspective on how influence, power, and social cohesion operate at a micro-level. In the case of the Karate Club, the dataset consists of 34 nodes (representing club members) and 78 edges (representing friendships or interactions). The club ultimately split into two factions due to internal conflict, and this split is clearly reflected in the network's structure. Understanding the underlying dynamics that led to this split is critical, as it provides insights into how similar divisions might occur in other settings, such as organizations, communities, or even political groups. By analyzing this dataset, we gain a model for interpreting and anticipating divisions within small networks.

Project Objectives and Approach :

The project aims to apply several key principles of social network analysis to the Karate Club dataset. Here, the primary objectives include:

Network Construction : The first step in this project is to construct the network graph for the Karate Club data. Each node corresponds to a member, while edges represent social interactions. Using network visualization tools, we will map the relationships visually, creating an intuitive representation of the club's social structure. This graph provides the foundation for further analysis, allowing us to easily observe sub-groups and connectivity patterns.

Centrality Analysis : Centrality measures are essential in identifying influential nodes within the network. The project includes calculating various centrality metrics:

Degree Centrality : Counts the number of direct connections each member has. In social terms, this measure often reflects popularity or reach within the network. Betweenness Centrality : Indicates how often a node acts as a bridge along the shortest path between two other nodes. Nodes with high betweenness centrality are critical for connecting different parts of the network and may play an essential role in group cohesion. Closeness Centrality: Represents how quickly a member can access others in the network, essentially capturing the idea of "social reach." Eigenvector Centrality: Measures influence by assigning scores based on a node's connections to other highly connected nodes, capturing both direct and indirect influence within the network. Community Detection: One of the main challenges in social network analysis is identifying community structures. Communities are groups of nodes more densely connected to each other than to the rest of the network. For this dataset, community detection methods such as the Girvan-Newman algorithm (based on betweenness centrality) can reveal clusters within the club that might have led to the eventual division. This part of the analysis is particularly relevant to understanding social cliques and factionalism within small groups.

Visualization of Network Structure: Visualization is key to interpreting network data. For the Karate Club dataset, network visualizations will help highlight clusters, influential members, and the strength of connections within the group. Using color-coding, node sizing, and other visualization techniques, the project will create an intuitive depiction of the network, making it easier to identify central figures, peripheral members, and distinct factions within the club.

Analyzing Network Cohesion and Structural Balance: Beyond centrality and community detection, the project also explores measures of network cohesion and structural balance. Cohesion metrics such as clustering coefficients and network density provide insights into the overall connectedness of the group. Structural balance theory, on the other hand, examines the tension or stability in triadic relationships (groups of three nodes) within the network, offering a deeper understanding of the potential for conflict or alignment within social groups.

Theoretical Foundations and Methodology:

The methodology for this project is rooted in the principles of graph theory, which provides the mathematical foundation for network analysis. Graph theory allows us to represent complex social structures as mathematical objects, enabling the use of algorithms and metrics to quantify relationships and structure. For instance, calculating shortest paths between nodes is essential for determining betweenness centrality, while adjacency matrices help in detecting clusters and communities within the graph.

To ensure a comprehensive analysis, this project incorporates Python libraries such as NetworkX for graph construction, manipulation, and metric calculation. NetworkX offers a robust framework for handling graph data, and its built-in functions allow efficient computation of centrality measures, community detection, and path analysis. Visualization libraries like Matplotlib and Plotly will also be used to create detailed, interactive visualizations that capture the nuances of the Karate Club's network structure.

Significance of the Study

Analyzing the Karate Club dataset is not just about understanding this specific group but about demonstrating the potential of social network analysis in addressing real-world issues. Networks are all around us, from social media connections to organizational hierarchies, and the principles derived from studying small networks can often be applied to larger, more complex systems. For example, in organizational settings, SNA can help identify key influencers, potential conflicts, and optimal communication pathways. In public health, understanding social networks can inform strategies for disease control and health interventions by identifying central figures who could spread information or behaviors quickly.

Furthermore, community detection and network visualization have applications in marketing, where identifying social groups and key influencers can aid in targeted advertising. The project, therefore, not only contributes to understanding the dynamics of a single karate club but also serves as a template for analyzing various types of social networks, highlighting the versatility and power of social network analysis as a tool for understanding human relationships.

In conclusion, the Karate Club dataset provides a unique and instructive case for applying SNA techniques, exploring network structure, and visualizing social influence and division. By following this structured approach, the project offers valuable insights into the mechanics of social connections and the factors that can lead to group cohesion or fragmentation.

Graph Representation :

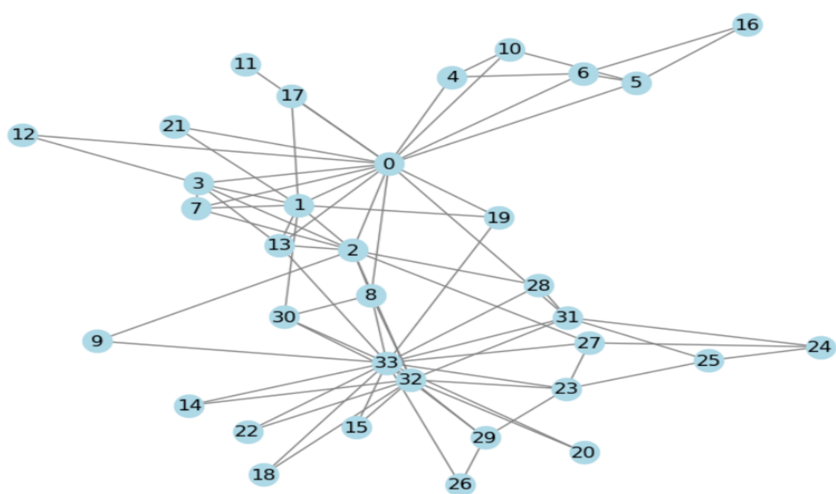


Figure : Karate Club Graph

II. Methodology

A. Centrality Measures

- **Degree Distribution**: The degree distribution is calculated to understand how connections are distributed among nodes in the karate club network. For each node i, the degree k_i represents the number of direct ties to other nodes. Given that social networks often exhibit a small number of highly connected nodes (hubs), we explore whether the karate network follows this pattern, which may reflect certain members acting as key influencers or central figures.
- **Cumulative Degree Distribution**: The cumulative degree distribution $P(K \geq k)$ helps us observe the probability that a randomly chosen node has a degree greater than or equal to k. This distribution is insightful in understanding whether the network follows a power-law distribution, typical in many real-world social networks. In the karate dataset, if a few members exhibit much higher connectivity, it might indicate their role as network leaders or mediators. analyzing the cumulative degree distribution in this dataset can tell you:
  1) Network Sparsity: The cumulative degree distribution can reveal whether most nodes have a low degree (few connections) or if there's a significant number of nodes with a high degree (many connections). For example, in a social network like the karate club, you might expect to see a few highly connected members (leaders or well-connected members) and many less-connected members.
  2) Hub Identification: By examining the tail of the distribution, you can identify nodes that serve as hubs. In the karate dataset, these are the individuals who have the most connections, likely representing central figures in the social structure.
  3) Community Structure and Polarization: If there is a clear break or change in the distribution, it might indicate factions or subgroups, which aligns with the known divisions in the karate dataset. This can be especially insightful when verifying community detection results.
- **Betweenness Centrality**: Betweenness centrality measures the extent to which a node lies on paths between other nodes. For a node v, its betweenness $C_B(v)$ is defined as:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

  where σ_{st} is the total number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of those paths passing through v. For the karate network, nodes with high betweenness centrality are likely the connectors or brokers between different clusters, perhaps influencing group dynamics. betweenness centrality reveals in this context:
  4) Key Mediators in the Network: Nodes with high betweenness centrality are crucial for maintaining communication across different parts of the network. In the karate club dataset, high-betweenness nodes would typically be the ones that connect subgroups or factions, helping bridge different parts of the network. These individuals are often essential for information flow and can influence both factions.
  5) Community Leaders: Since the karate club dataset historically reflects two factions (due to a conflict), nodes with high betweenness centrality often represent the leaders of these factions. For example, the dataset is known to feature prominent figures (such as "Mr. Hi" and the club's administrator) who connect various members and thus have high betweenness scores.
  6) Potential Power Brokers: : Nodes with high betweenness centrality can leverage their position to control the flow of information or influence interactions between groups. In a real-life scenario, these would be individuals who have strategic power because they control access to different subgroups. In the karate club, these individuals would likely be pivotal if either faction wanted to sway members to their side.
  7) Resilience and Network Vulnerability: If a high-betweenness node were removed, it could fragment the network, creating communication barriers between subgroups. This is particularly relevant for analyzing the club's dynamics: if a central person leaves (like one of the faction leaders), it could lead to further fragmentation.
- **Closeness Centrality**: Closeness centrality of a node v is the reciprocal of the average shortest path distance to all other nodes:

$$C_C(v) = \frac{1}{\sum_u d(v, u)}$$

  where $d(v, u)$ is the distance from v to u. In the karate network, nodes with high closeness centrality can access the majority of other members more quickly, indicating their potential to spread information or influence rapidly. what it represents for this dataset:
  8) Access to Information: Nodes with high closeness centrality can quickly access or spread information across the network. In the karate club context, this means that members with high closeness can efficiently communicate with or influence others. These individuals are well-positioned to stay informed and to keep others informed within the club's social structure.
  9) Influence Potential: High closeness centrality suggests that a member can influence many others with minimal "social distance," making them potentially persuasive or influential in swaying opinions within the network. This would be particularly relevant for key figures within each faction of the karate club.

10) Network Integration: A node with high closeness is generally more integrated within the network, as it has short paths to other members. In the karate club, members with higher closeness centrality could have stronger positions of influence across both factions due to their reachability.

11) Community Boundaries: Nodes with lower closeness centrality are often more peripheral, meaning they are more isolated or primarily connected within their subgroup. This is relevant in the karate club, where lower closeness might indicate members who are more tied to one faction and less connected to the other.

- **Eigenvector Centrality**: Eigenvector centrality assigns relative scores to nodes based on the centrality of their neighbors. Mathematically, it is derived from the eigenvector of the adjacency matrix A:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

where λ is the largest eigenvalue. Nodes connected to high-scoring nodes also score highly, which could help identify influential individuals linked to other important figures within the karate network. In the context of the karate club dataset, eigenvector centrality can provide insights into the following:

12) Influential Members: Nodes (members) with high eigenvector centrality are not only well-connected but are also connected to other well-connected members. This means they play a significant role in the network's overall structure and can influence other members more effectively.

13) Leadership and Group Dynamics: In the karate club context, higher eigenvector centrality might indicate members who are key leaders or influential figures within the group, potentially affecting decisions, behaviors, or group dynamics.

14) Community Structure: Eigenvector centrality can help identify clusters of members who are closely connected. By analyzing the centrality scores, you can discern which members are central to specific sub-groups within the karate club.

15) Resilience of the Network: Members with high eigenvector centrality are crucial for the network's integrity. If these members were removed, the network might become fragmented or less cohesive.

- **Katz Centrality**: Katz centrality is an extension of eigenvector centrality, adding a damping factor α to account for both direct and indirect connections:

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

where α and β are constants. In the karate dataset, this measure highlights nodes with broad influence, even if they aren't directly connected to all members.

- **PageRank**: Using a variation of eigenvector centrality, PageRank provides a score based on a random walk model. Nodes with higher PageRank scores are more likely to be visited during random traversals, making this metric relevant for assessing which members might be prominent through repeated interactions. In the Context of the Karate Club Dataset, PageRank Represents:

16) Relative Importance of Members: 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 (has many direct friends) but also connected to other influential members. This dual consideration makes PageRank a measure of both popularity and influence.

17) Influence Through Indirect Connections: PageRank takes into account not just the direct connections of a member but also the connections of their friends. This means that if a member is connected to other well-connected members, their own importance in the network increases. This indirect influence is crucial in understanding social dynamics.

18) Potential for Information Dissemination: Members with high PageRank scores are likely to be effective disseminators of information within the network. Their connections to other influential members mean that they can spread information quickly and efficiently, impacting group behavior and decisions.

19) Network Structure Insights: Analyzing PageRank can provide insights into the structure of the karate club network, revealing which members serve as hubs or bridges between different groups. Understanding these connections can help identify potential leaders or key communicators within the community.

20) Community Dynamics and Relationships: The PageRank scores can help illustrate the relationships and dynamics between different factions within the karate club. Members with high PageRank may play critical roles in connecting disparate groups, thereby influencing the overall cohesion and functionality of the club.

- **HITS Algorithm**: The HITS algorithm identifies "hubs" and "authorities" by iteratively refining scores based on neighboring nodes. In the context of the karate network, HITS could reveal key members who serve as authorities on information, as well as those acting as hubs for information dissemination. In the Context of the Karate Club Dataset, HITS Represents:

21) Authority Scores: Authority nodes are those that are highly linked to by other nodes. 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.

22) Hub Scores: Hub nodes are those that have many outbound links to other nodes. In this dataset, a member with a high hub score actively connects to many others, potentially promoting interaction, communication, and information sharing. Hubs are often influential in guiding the dynamics of the network by connecting members.

23) Community Dynamics: By distinguishing between hubs and authorities, the HITS algorithm provides insights into the structure and dynamics of the karate club. For example, a member may serve primarily as a hub by connecting various sub-groups, while another might be recognized as an authority due to their knowledge or experience.
24) Dual Role of Members: Some members may play dual roles in the network, functioning as both hubs and authorities. Understanding these roles can help in identifying key influencers within the club and their contributions to group dynamics.
25) Impact on Social Interactions: The analysis of hubs and authorities can shed light on how social interactions are structured in the karate club. It can help identify who drives conversations, who is sought after for advice, and how knowledge flows within the group.

- **Edge Betweenness Centrality**: Edge betweenness identifies edges that frequently appear on shortest paths between nodes, providing insight into connections that act as "bridges" between different network clusters. By calculating edge betweenness, we can see which relationships in the karate dataset are most critical for maintaining connectivity. In the context of the karate club dataset, edge betweenness centrality can provide insights into the following:
26) Bridge Edges: Edges with high betweenness centrality act as bridges or connectors between different parts of the network. In the karate club, such edges might represent crucial friendships that link different groups of members.
27) Community Structure: The karate club is known for its community structure, where members can be divided into two factions. Edges that have high betweenness centrality may help maintain communication between these factions or be pivotal during the conflict between the two groups.
28) Influence on Information Flow: High betweenness edges are critical for the flow of information across the network. They are often the shortest paths for communication between members. In the context of the karate club, this can indicate which friendships facilitate better communication between cliques or factions.
29) Vulnerability and Resilience: If an edge with high betweenness centrality is removed, it might significantly disrupt communication in the network, making the network more fragmented. This can be important for understanding resilience in social networks.

B. Path and Reachability Analysis

- **Shortest Path Length**: In the karate dataset, the shortest path length between pairs of nodes represents the minimum interactions required to connect members. Calculating all pairs' shortest paths allows us to analyze the efficiency of information flow across the network and provides a basis for assessing potential barriers in group communication.
- **Average Path Length**: The average path length in a network indicates the typical distance between nodes. For the karate club, this measure helps us understand how quickly information or influence could spread from one member to another.
- **BFS and DFS (Breadth-First Search and Depth-First Search)**: BFS and DFS are used to explore the connectivity structure. In the karate network, these algorithms help identify isolated clusters, allowing us to examine the connectedness of sub-groups and uncover any significant separations within the club.

C. Clustering and Community Detection

- **Local Clustering Coefficient**: The local clustering coefficient, given by:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

where e_i is the number of edges among the neighbors of node i and k_i is the degree of node i, shows the likelihood of a node's neighbors forming a complete clique. In the karate club, a high local clustering coefficient could indicate tight-knit groups within the club. In the context of a social network like the karate club, here's what the local clustering coefficient reveals:
1) Tightness of Friend Groups: A high local clustering coefficient for a member indicates that their friends are also likely to be friends with each other, suggesting a closely-knit friend group or clique around that individual.
2) Community Structure: Members with higher clustering coefficients are likely part of densely interconnected communities within the club. Such members might be involved in smaller, closely-knit subgroups rather than being spread across multiple groups.
3) Bridge Roles: Members with low clustering coefficients often connect otherwise unconnected parts of the network. If a member has a low clustering coefficient, their friends might not know each other, potentially positioning them as a bridge between different social circles within the club.
4) Indicator of Influence Spread: In networks with high clustering, influence or information spread is often rapid within tightly-knit groups, as each person is connected to many others in the same group. Nodes with a high clustering coefficient may influence their friends' friends more effectively than nodes with low clustering.
Example Calculation For example, if a node has three neighbors and two of them are also connected to each other, the local clustering coefficient for that node would be 0.67 (two out of three possible connections among neighbors).

- **Global Clustering Coefficient**: The global clustering coefficient provides an overall measure of clustering within the network. In the karate dataset, this metric helps to evaluate how well the members of the club are connected as a whole. Here's what the global clustering coefficient represents for the karate club dataset:
  5) Overall Cohesion: A high global clustering coefficient indicates that the network as a whole has a significant number of triangles, suggesting that friends of friends are likely to also be friends. This can signify a strong community structure where members are tightly interconnected.
  6) Community Formation: If the coefficient is high, it may suggest that the karate club consists of well-defined subgroups or cliques. Conversely, a low global clustering coefficient may indicate a more loose-knit network where members do not cluster as much.
  7) Social Dynamics: Understanding the global clustering coefficient can help analyze the social dynamics within the club. High clustering might correlate with certain behaviors, such as collaboration or support within groups, while low clustering might suggest a more diverse set of connections and interactions across different groups.
  8) Influence and Information Spread: A higher global clustering coefficient can facilitate quicker information spread within the network, as members are more likely to be interconnected. This could be crucial in scenarios like event organization or decision-making within the club.
  Example Calculation The global clustering coefficient is calculated as follows:
  $$c = \frac{3(Number of triangles)}{Number of connected triplets}$$
  Where:
  Number of triangles: The count of sets of three nodes that are all mutually connected. Number of connected triplets: The count of sets of three nodes where at least two nodes are connected.
- **Connected Components**: We define connected components as subnetworks within which nodes are reachable from each other. In the karate dataset, connected component analysis reveals if any members or groups are isolated from the main network, indicating sub-communities or factions within the club.
- **Girvan-Newman Algorithm**: This algorithm iteratively removes edges with the highest betweenness to split the network into distinct communities. Applied to the karate network, Girvan-Newman helps to identify naturally occurring divisions among members. Applying the Girvan-Newman algorithm on the karate club dataset will help us identify communities or clusters by iteratively removing edges with the highest betweenness centrality. Here's what we would get:
  9) Detection of Natural Communities: The karate club dataset historically represents a club that split into two factions. The Girvan-Newman algorithm should ideally divide the network into two primary communities that represent these two groups. By progressively removing high-betweenness edges, the algorithm reveals densely connected subgroups within the network, where members within each subgroup have more connections to each other than to members outside their group.
  10) Hierarchy of Splits: The Girvan-Newman algorithm can generate a hierarchy of splits, breaking down the network into smaller and smaller communities. After identifying two major communities, further iterations may split each community into smaller subgroups. This hierarchical community structure shows various levels of cohesion and division within the network, helping us visualize subgroups and how tightly members are bound.
  11) Community Structure and Roles of Key Nodes: Some nodes act as bridges or ”brokers” between communities. In the karate club, these could represent members who had connections in both factions and potentially played a mediating role. By identifying nodes with high betweenness centrality (the ones removed first), we can recognize the influential members who connected different parts of the club.
- **Louvain Method**: The Louvain method optimizes modularity to detect communities, providing a different approach from Girvan-Newman. For the karate dataset, Louvain helps discover cohesive subgroups, which may align with real-life friend groups or alliances. In the Context of the Karate Club Dataset, the Louvain Method Represents: Community Detection: 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.
  12) Modularity Optimization: The algorithm optimizes a modularity score, which quantifies the density of links inside communities as compared to links between communities. A higher modularity score indicates a more pronounced community structure.
  13) Understanding Social Dynamics: The identified communities can help understand social dynamics and relationships within the karate club. For example, one community may represent a subgroup of members with similar interests or affiliations.
  14) Identification of Key Groups: By applying the Louvain method, you can identify key groups that may play important roles in the club's activities, decision-making processes, or conflict resolution.
  15) Potential for Enhanced Collaboration: Insights from community detection can inform strategies for enhancing collaboration and interaction within and across communities, potentially leading to improved group cohesion.

16) Visualization of Community Structure: The communities identified by the Louvain method can be visualized in a graph, allowing for an intuitive understanding of how members are clustered together based on their interactions.

D. K-Cliques, K-Plex, K-Core, and Other Community Metrics

- **K-Clique**: A k-clique is a subset where each node connects to at least k others within the group, indicating fully connected communities. In the karate dataset, these groups reflect tightly connected members.
- **K-Clan, K-Club, K-Plex, K-Core**: Each of these k-based metrics provides different insights on the density and resilience of subgroups. For example, k-club analysis reveals clusters reachable within k steps, relevant for understanding members' indirect connections.

E. Network Models, Epidemic Models, and Other Properties

- **Watts-Strogatz and Barabási-Albert Models**: Comparing the karate network with small-world (Watts-Strogatz) and scale-free (Barabási-Albert) models can show if it resembles these commonly observed social network structures. Key Features of the Watts-Strogatz Model

  1) Small-World Property: The model generates networks that exhibit both high clustering (like regular lattices) and short average path lengths (like random graphs). This means that while individuals tend to have many local connections, they can also reach each other through a few intermediaries, reflecting the "small-world" phenomenon observed in real-life social networks.

  2) Parameterization: The model is defined by three parameters: N: The number of nodes in the graph (which corresponds to the number of members in the karate club). K: Each node is connected to K nearest neighbors (which reflects the degree of regularity). P: The probability of rewiring edges (which introduces randomness into the network).

  3) Analysis of Connectivity: By comparing the Watts-Strogatz model generated from the karate dataset to the actual karate club network, you can assess how well the model captures the existing connections and structures in the network. This can provide insights into community structure, the influence of individual members, and how information might spread through the network.

  4) Community Structure: Given that the karate club dataset represents a community of individuals with known social ties, applying the Watts-Strogatz model can help analyze the formation and robustness of communities within the network, as well as the potential for different group dynamics.

  Applications of the Watts-Strogatz Model to the Karate Club Dataset

  5) Understanding Network Dynamics: The model can help simulate how friendships may form or dissolve over time in social networks.

  6) 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, you could apply community detection algorithms to identify clusters of individuals with similar traits or interests.

  Applications of the Watts-Strogatz Model to the Karate Club Dataset

  7) Understanding Network Dynamics: The model can help simulate how friendships may form or dissolve over time in social networks.

  8) 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.

  9) Community Detection: Using the generated networks, you could apply community detection algorithms to identify clusters of individuals with similar traits or interests.

  Key Features of the Barabási-Albert Model

  10) Scale-Free Networks: The BA model generates networks that exhibit a power-law degree distribution, meaning that a few nodes (hubs) have a very high degree (many connections), while most nodes have only a few connections. This reflects the characteristic of many real-world networks, including social networks, where certain individuals are significantly more connected than others.

  11) Preferential Attachment: The model operates on the principle of preferential attachment, where new nodes are more likely to connect to existing nodes that already have a high degree. This mimics real-life scenarios where individuals are more likely to befriend popular or well-connected individuals.

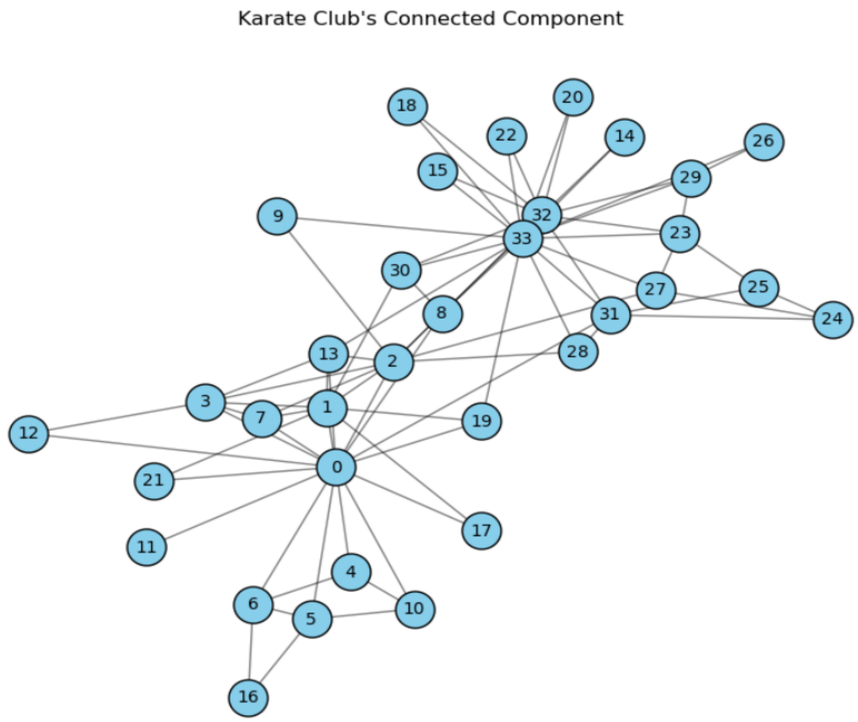
  12) Dynamic Growth: Unlike static models, the BA model allows for the growth of the network over time. It starts with a small number of interconnected nodes and adds new nodes one at a time, creating edges according to the preferential attachment mechanism. Applications of the Barabási-Albert Model to the Karate Club Dataset

  13) 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.

14) Community Structure: By comparing the BA-generated network with the actual karate club graph, researchers can assess the similarity in community structure and identify how closely the karate club aligns with a scale-free network.

15) 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.

- **SIS and SEIS Models**: SIS and SEIS are dynamic models of contagion or rumor spread. Applying them to the karate network simulates how information or influence might circulate and return over time, reflecting real-life dynamics in social groups.
- **Entropy and SimRank**: Entropy quantifies the disorder in node connections. SimRank provides a similarity measure, helping identify members with overlapping roles or interests.
- **Dunbar Number and Neighborhood Overlap**: Using Dunbar's concept, we assess whether the network size reflects realistic limits on stable social connections. Neighborhood overlap shows the strength of node relationships.



III. Result and Findings

Degree Distribution :

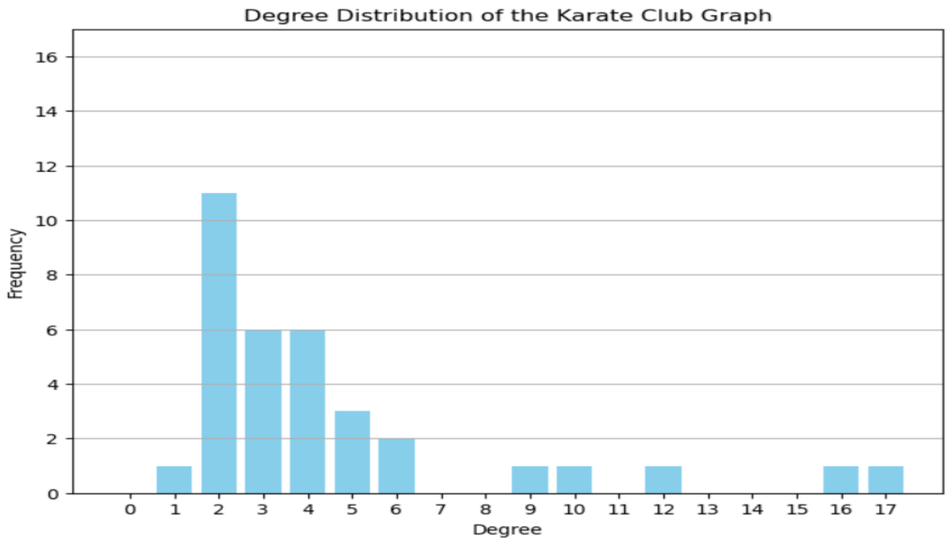


Figure 1 : Degree Distribution

Result :

Node 33 has highest degree(17)

Node 11 has lowest degree(1)

Cumulative Degree Distribution :

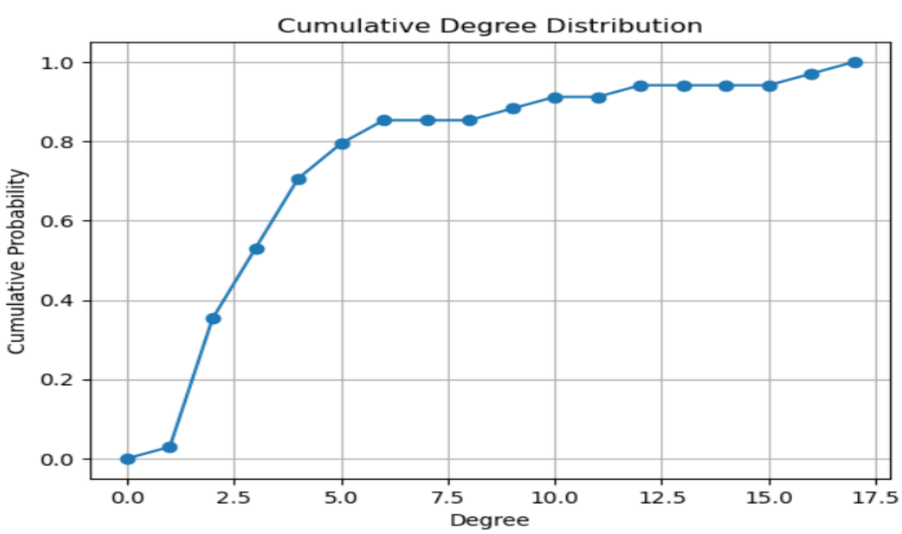


Figure 2 : Cumulative Degree Distribution

Hub score :

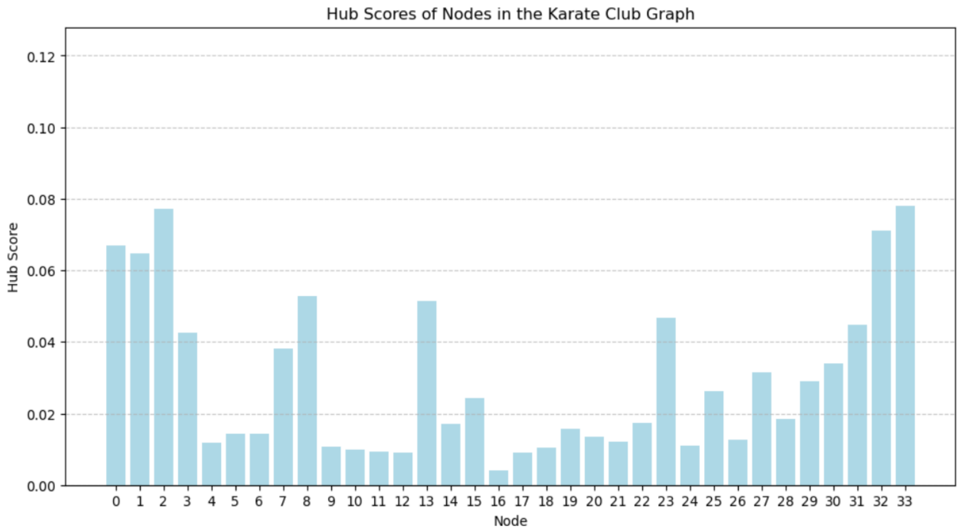


Figure 3 : Hub score

Result :

Node 33 has highest Hub score

Node 11 has lowest hub score

Authority score :

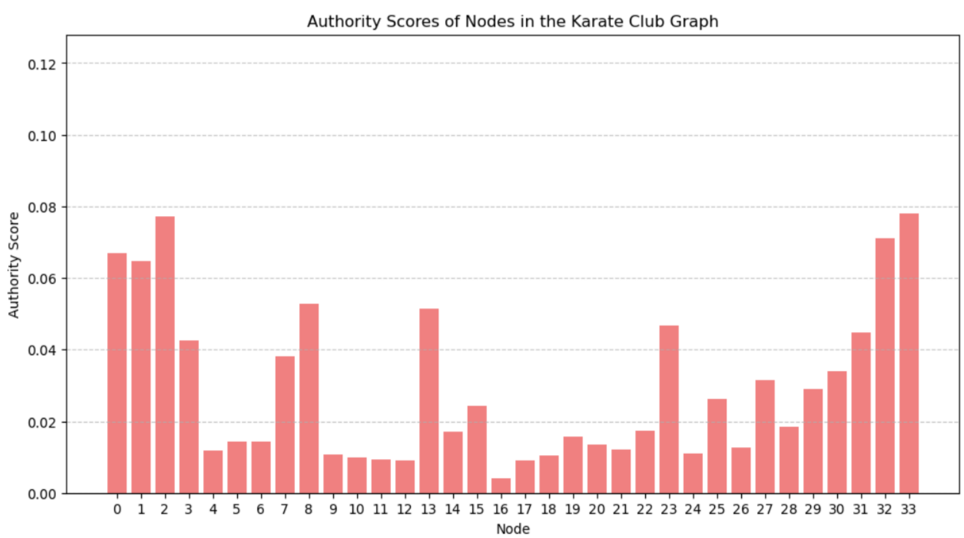


Figure 4 : Authority score

Result :

Node 33 has highest Authority score

Node 11 has lowest Authority score

Betweenness Centrality :

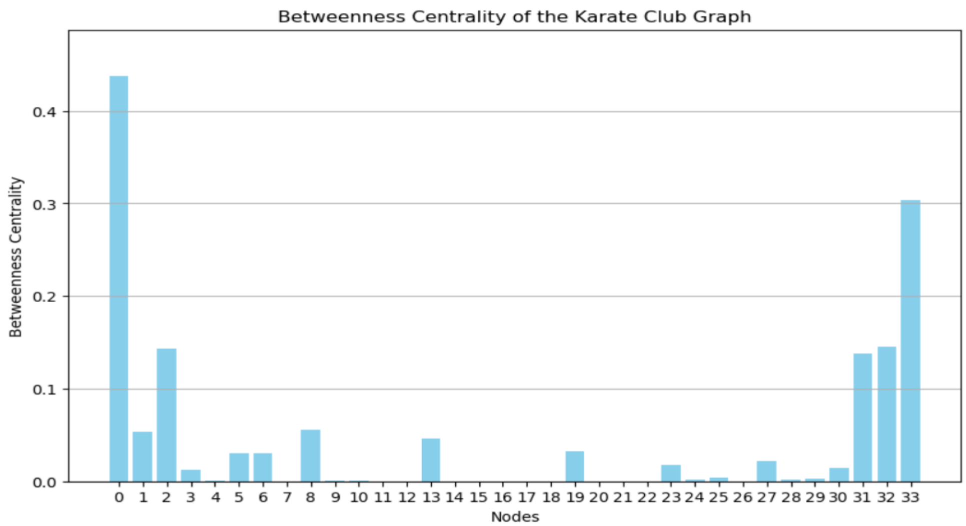


Figure 5 : Betweenness Centrality

Result :

Node 0 and 33 have the highest Betweenness Centrality

Closeness Centrality :

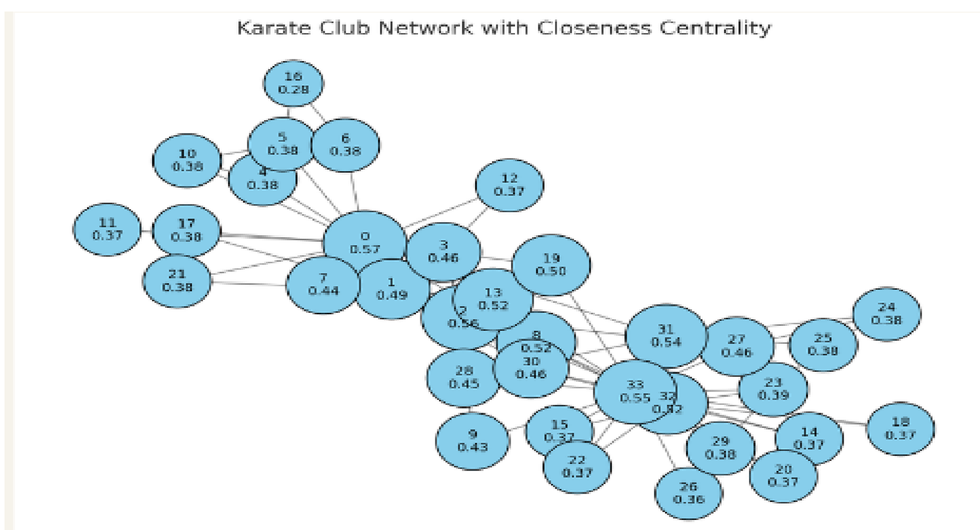


Figure 6 : Closeness Centrality

Result :

Node 0 has the highest closeness Centrality

Eigen vector Centrality :

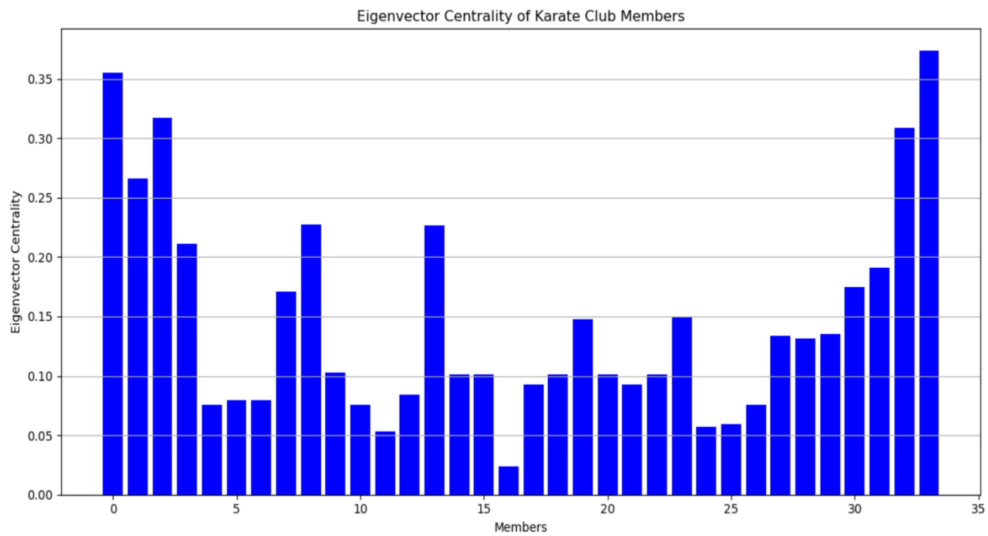


Figure 7 : Eigen vector Centrality

Result :

Node 0 and 1 have the highest Eigen vector Centrality

Shortest path length :

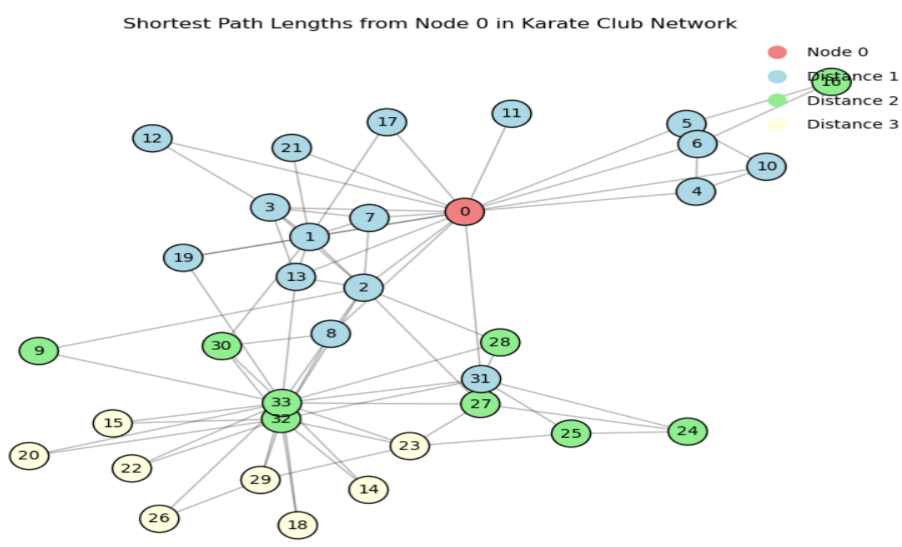


Figure 8 : Shortest path length

Result :

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 :

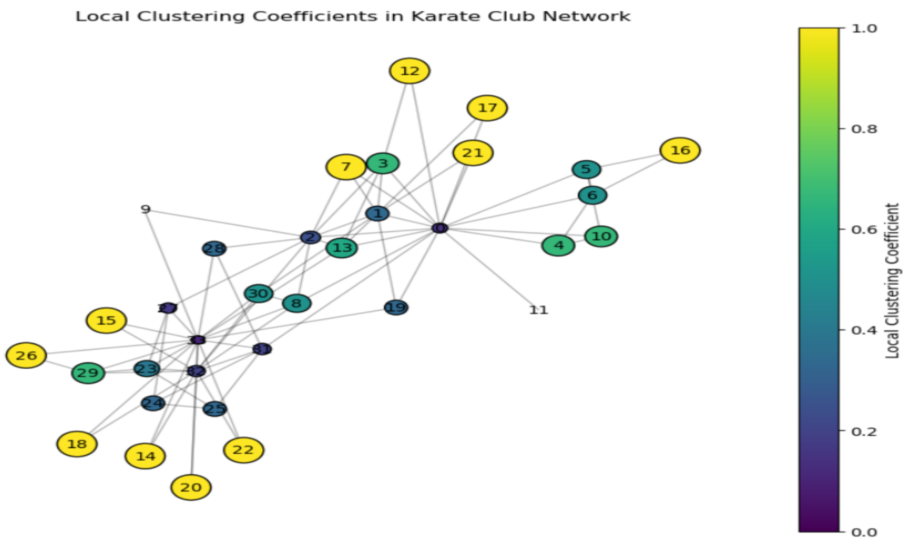


Figure 9 : Local clustering coefficient

Result :

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.

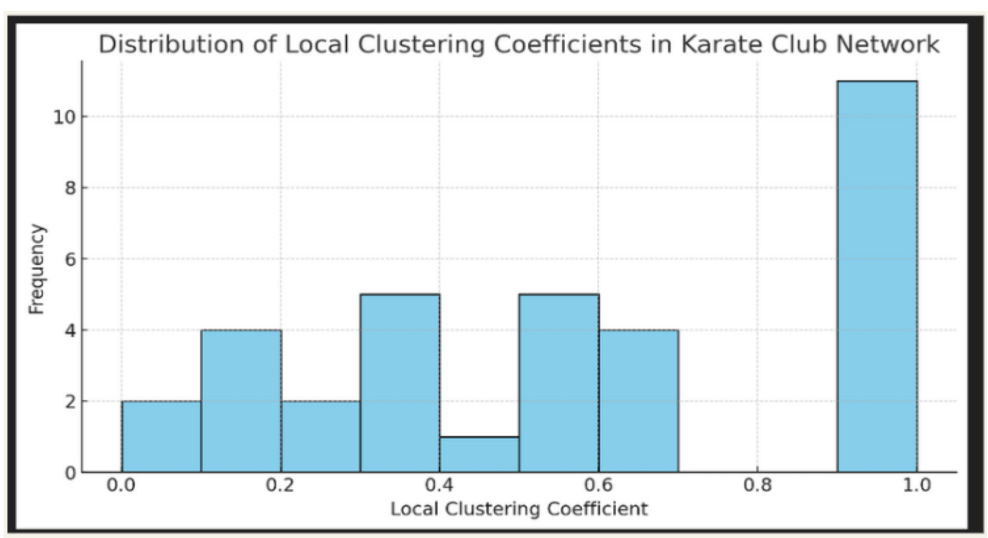


Figure 10 : Distribution of Local clustering coefficient

Edge betweenness centrality :

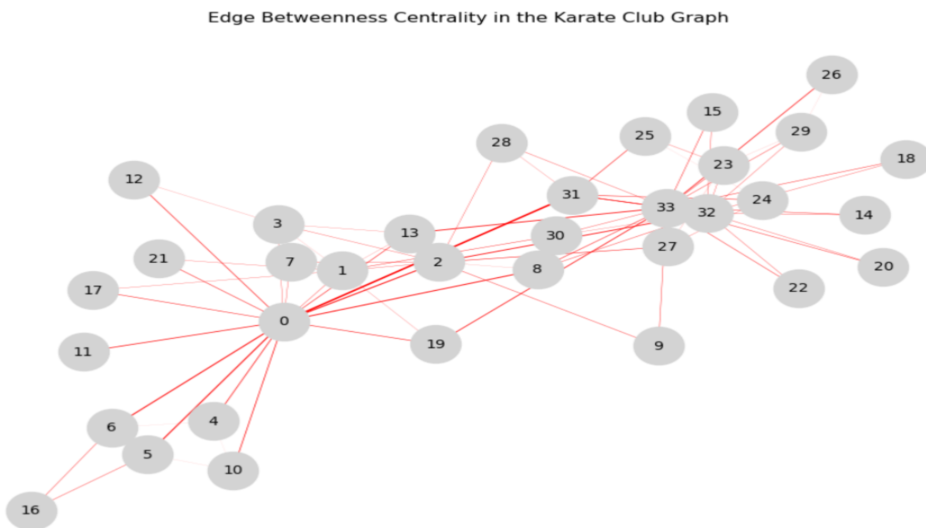


Figure 11 : Edge betweenness centrality

Result :

Thicker edges represent friendships that play a crucial role in connecting different members of the karate club.

Katz centrality :

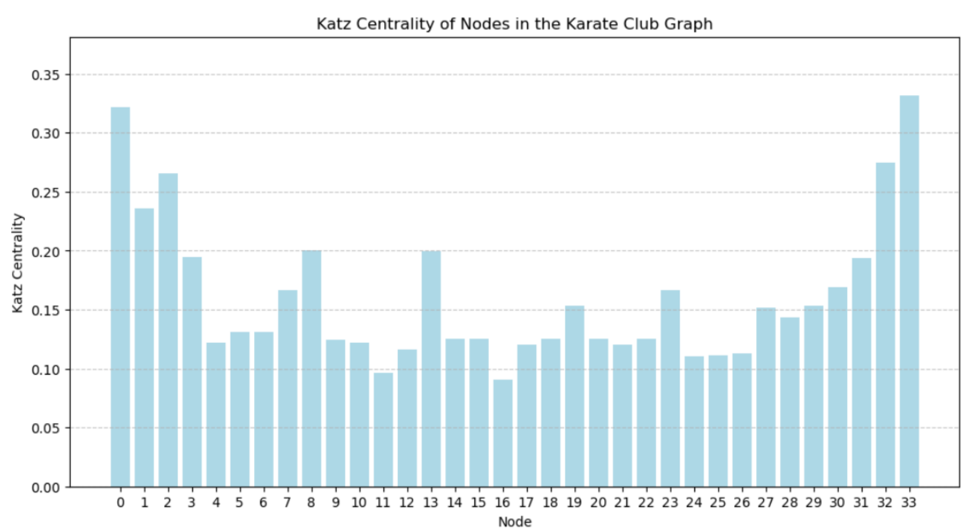


Figure 12 : Katz centrality

Result :

Node 33 has higher katz centrality(0.3314) followed by node 0(0.3213)

Node 16 has low katz centrality(0.0907)

Page Rank :

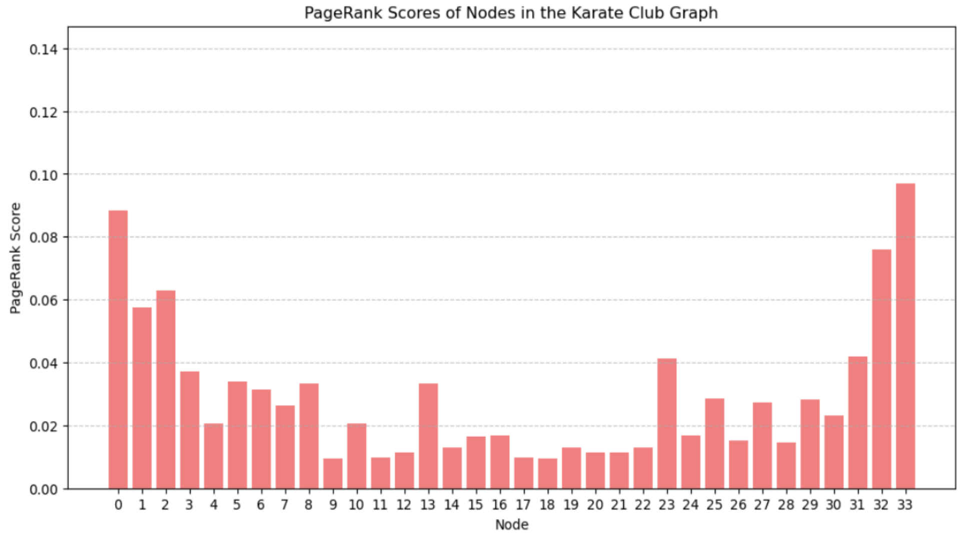


Figure 13 : Page Rank

Result :

Node 33 has higher page rank(0.0970)

Node 9 and 18 has low page rank(0.0095)

Core number :

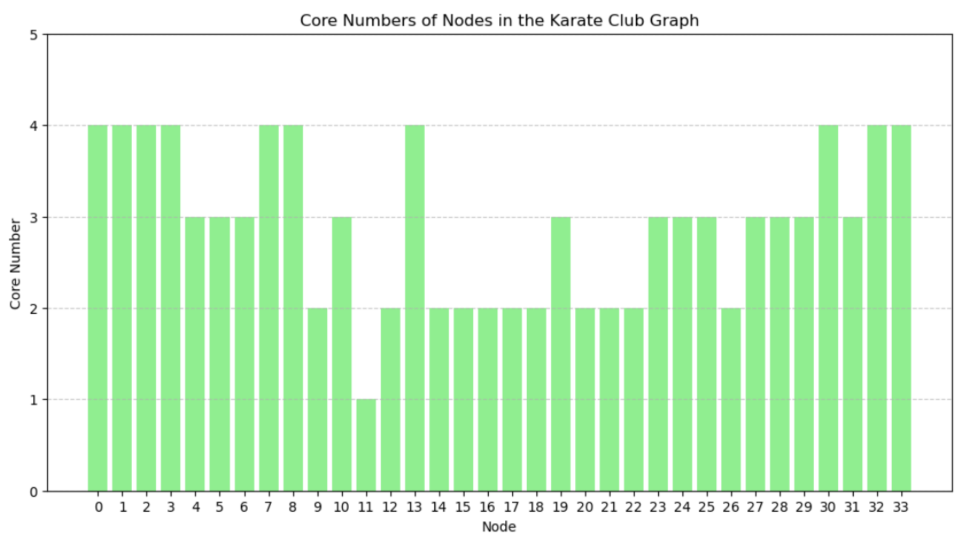


Figure 14 : Core number

Result :

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

Watts-Strogatz model :

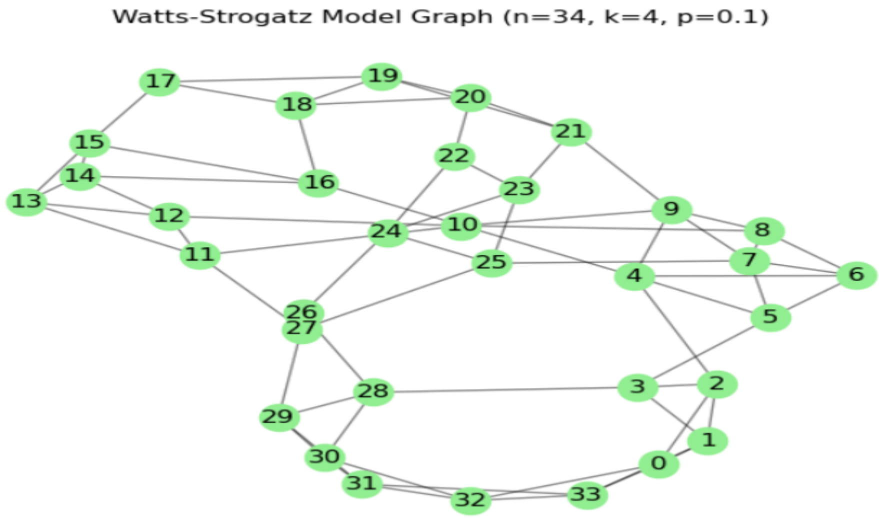


Figure 15 : Watts-Strogatz model

Barabási-Albert model :

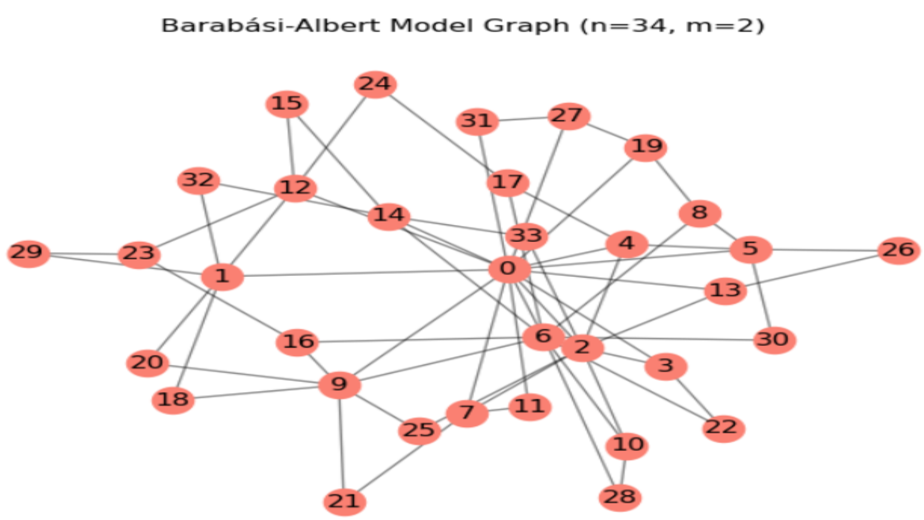


Figure 16 : Barabási-Albert model

Louvain method :

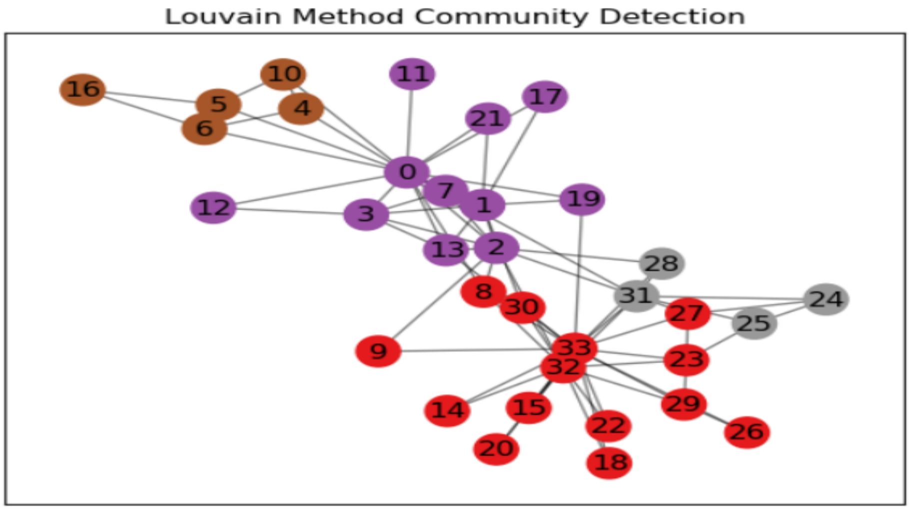


Figure 17 : Louvain method

IV. Conclusion

In summary, our analysis of the Zachary Karate Club network revealed critical insights into its structure, connectivity, and influential members. By using centrality measures, we identified high-degree nodes that play vital roles in sustaining cohesion and impacting the overall network, contributing to the club's division. Community detection algorithms like Girvan-Newman and Louvain highlighted distinct clusters within the network, mirroring the natural segmentation among members and illustrating robust social ties around key figures.

This analysis emphasizes the value of social network analysis in understanding similar systems. Insights into central figures and community boundaries can guide strategies to improve communication, reduce factionalism, and strengthen cohesion in various settings, from organizations to online communities. Moreover, the methods used here have broader applications, such as identifying influencers for targeted interventions in organizational, marketing, and public health contexts. The Karate Club network thus serves as a powerful example of how network science can reveal complex social dynamics and inform strategies for managing group interactions in diverse contexts.

References

[1] Zachary's Karate Club Dataset. : https://networkdata.ics.uci.edu/netdata/html/zacharyKarate.html

[2] Wikipedia contributors. (2023,November 5). HITS algorithm. *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/HITS_algorithm

[3] Rogers, I. (2002). *The Google PageRank Algorithm and How It Works*. IPR Computing. Retrieved from https://www.iprcom.com/papers/pagerank/

[4] Wikipedia contributors. (2023,November5). Social network analysis. *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/wiki/Social_network_analysis