8/12/2024

Project Report

Introduction to Social Network Project

Submitted by: Mohamed Ahmed Mohamed Hossen,
I.D:22010211

Abstract

This project explores the structural properties and dynamics of the Karate Club graph, a well-known social network dataset. Using NetworkX and Matplotlib, we conducted a comprehensive analysis of the graph, focusing on various aspects such as node degree, average degree, special graph structures, connectivity, matrix representation, and graph representations (edge list and adjacency list). We also quantified the social structure through clustering coefficients and centrality measures, including degree centrality, closeness centrality, betweenness centrality.
Our analysis revealed key insights into the network's connectivity and the importance of specific nodes within the social structure. We enhanced our understanding by visualizing the graph with edge thickness based on weight and creating heatmaps to illustrate the distribution of node attributes. These visualizations provided a clear and intuitive representation of the network's characteristics.
The findings from this project demonstrate the utility of graph theory and network analysis in understanding complex social networks. The methodologies and tools used in this study can be applied to other datasets, offering valuable perspectives on the dynamics of various types of networks.

About the Dataset

The dataset used in this project is the Karate Club graph, a well-known social network dataset that represents the social interactions within a karate club. This dataset was first introduced by Wayne Zachary in his 1977 study, which analyzed the social dynamics and conflict within the club.

Dataset Description :

- Nodes: The graph consists of 34 nodes, each representing a member of the karate club.
- Edges: There are 78 edges in the graph, representing the social interactions or friendships between the members.
- Attributes: Each node and edge can have additional attributes, such as weights representing the strength of interactions.

Problems to be Addressed

This project aims to address several key problems related to the analysis and understanding of social networks using the Karate Club graph. The specific problems to be addressed include:

1. Identifying Key Members in the Network
   - Problem: Determining which members of the karate club are the most influential or central to the network.

2. Understanding Network Connectivity
   - Problem: Assessing the overall connectivity of the network and identifying any potential weak points or vulnerabilities.

3. Detecting Communities within the Network
   - Problem: Identifying subgroups or communities within the karate club that may have stronger internal connections.

4. Quantifying Social Structure
   - Problem: Measuring the degree to which members of the karate club tend to cluster together and form tightly-knit groups.

5. Visualizing Network Properties
   - Problem: Creating clear and informative visualizations that highlight important aspects of the network, such as node centrality and edge weights.

6. Representing the Network in Different Formats
   - Problem: Providing different representations of the network, such as adjacency matrices, edge lists, and adjacency lists, to facilitate various types of analysis.

Approaches Taken

Approaches that have taken place in the project structure; in order to solve presented issues :

1. Identifying Key Members in the Network
   - Approach: Use centrality measures such as degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality to identify key members.

2. Understanding Network Connectivity
   - Approach: Analyze the connectivity of the graph, including the identification of connected components and the robustness of the network.

3. Detecting Communities within the Network
   - Approach: Apply community detection algorithms to uncover clusters or groups within the network.

4. Quantifying Social Structure
   - Approach: Calculate clustering coefficients and average clustering coefficients to quantify the social structure of the network.

5. Visualizing Network Properties
   - Approach: Use NetworkX and Matplotlib to generate visualizations that vary node sizes and colors based on centrality measures, and edge thickness based on weights.

6. Representing the Network in Different Formats
   - Approach: Generate and analyze different representations of the graph to understand its structure from multiple perspectives.

Code Description

TECHNOLOGIES USED
PYTHON networx matplotlip numpy seaborn Pickle

ACTIVITY
IMPORTING LIBRARIES
LOADING AND DRAWING
CENTRALITY MEASURES
VISUALIZING CENTRALITY MEASURES

ACTIVITY
COMMUNITY DETECTION
DEGREE DISTRIBUTION
PRINTING CENTRALITY MEASURES
HEATMAP OF DEGREE CENTRALITY

Importing Libraries
The necessary libraries for graph analysis and visualization are imported. These include NetworkX for handling graph data, Matplotlib for creating visualizations, and Seaborn for generating heatmaps.

Loading and Drawing the Graph
The Karate Club graph is loaded, representing the social interactions within a karate club. Various layouts are used to visualize the graph, including the default spring layout, random layout, circular layout, and spectral layout. These layouts help in understanding the structure and relationships within the network from different perspectives.

Centrality Measures
Centrality measures such as degree centrality, closeness centrality, and betweenness centrality are calculated. These measures identify the most important or influential nodes in the network. Degree centrality measures the number of connections each node has, closeness centrality indicates how close a node is to all other nodes, and betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes.

Visualizing Centrality Measures
The graph is visualized with nodes colored and sized based on their centrality measures. This visualization highlights the most central nodes, making it easier to identify key members within the network. A color bar is added to indicate the closeness centrality of nodes, providing a visual representation of their importance.

Community Detection
Communities within the graph are detected using the greedy modularity algorithm. This method identifies clusters or groups of nodes that are more densely connected to each other than to the rest of the network. The graph is then visualized with nodes colored based on their community membership, revealing the underlying community structure.

Degree Distribution
The degree distribution of the nodes is analyzed and plotted. This distribution shows the frequency of each degree value in the graph, providing insights into the connectivity of the network. Nodes with higher degrees are more connected and potentially more influential within the network.

Printing Centrality Measures
The centrality measures for the first five nodes are printed to provide a quick overview of the most central nodes in the graph. This summary helps in identifying key members and understanding their roles within the network.

Heatmap of Degree Centrality
A heatmap is created to visualize the degree centrality of all nodes in the graph. This heatmap provides a clear and intuitive representation of the distribution of centrality values, highlighting the most connected nodes.
This comprehensive analysis and visualization of the Karate Club graph reveal important aspects of the network's structure and dynamics. By examining centrality measures, community structure, and degree distribution, we gain valuable insights into the social interactions within the karate club. These techniques can be applied to other datasets to uncover patterns and relationships in various types of networks.

Project Screenshots

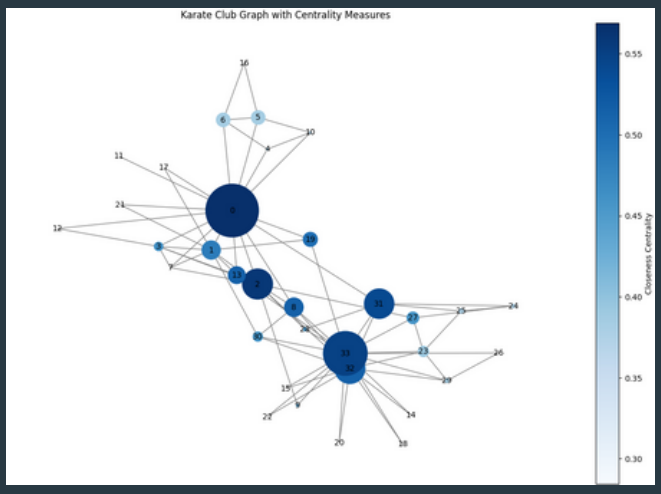

Image 1: The Karate Club Graph with Centrality Measures visualizes the relationships and importance of different nodes in a karate club social network. The size of the nodes represents their centrality, showing which members are the most central and influential within the group.

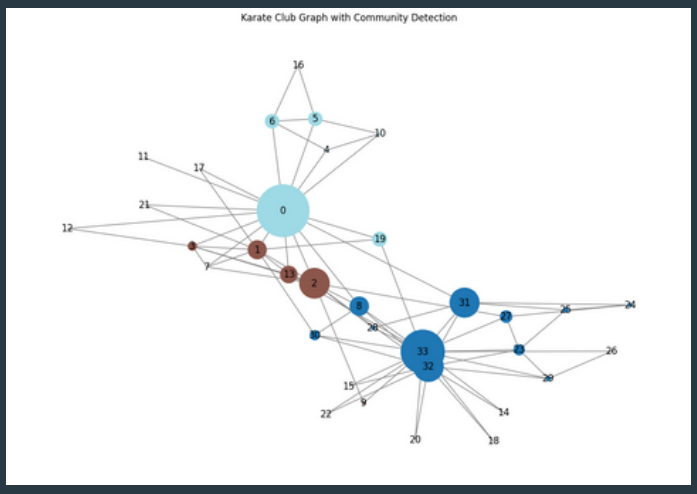

Image 1: The Karate Club Graph with Centrality Measures visualizes the relationships and importance of different nodes in a karate club social network. The size of the nodes represents their centrality, showing which members are the most central and influential within the group.

Project Screenshots

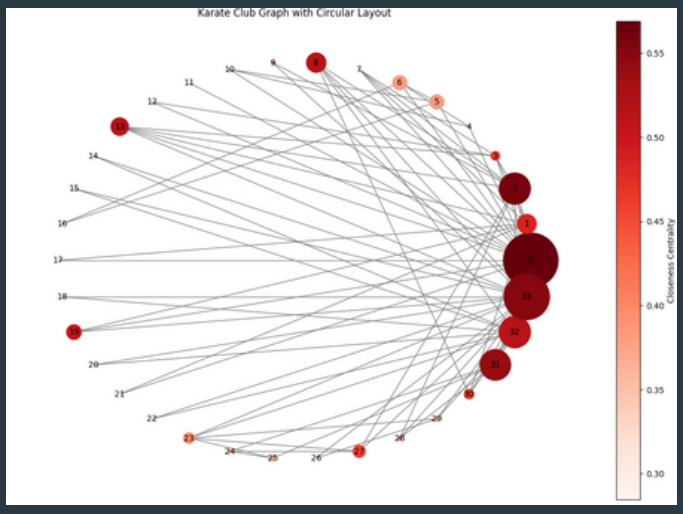

Image 3: The Karate Club Graph with Circular Layout presents a radial visualization of the social network, highlighting the interconnections between members. This layout emphasizes the overall structure and patterns of the network.

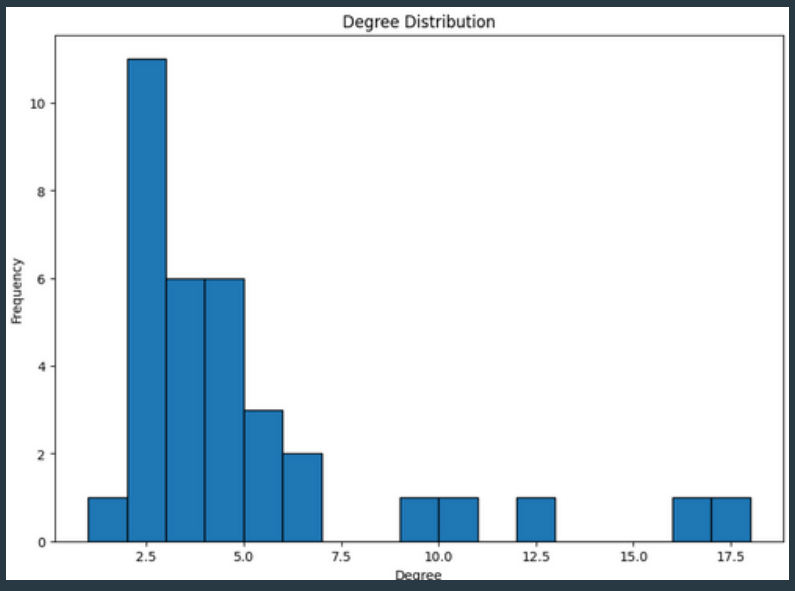

Image 4: The Degree Distribution graph shows the frequency of different node degrees (the number of connections each node has) within the karate club network. This distribution provides insights into the network's topology and the presence of hub nodes or high-degree members.

Project Screenshots

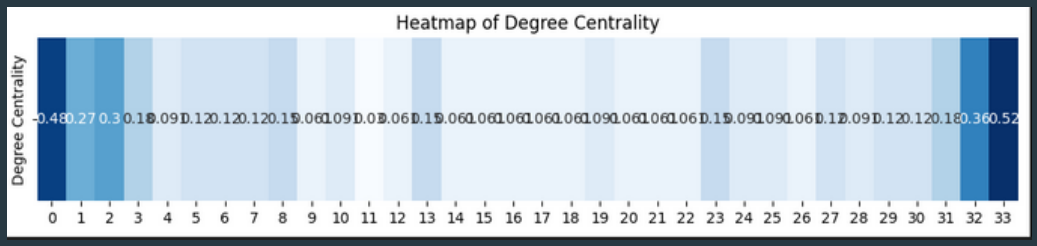

Image 5: The Heatmap of Degree Centrality illustrates the degree centrality values for each node in the network. Degree centrality simply counts the number of connections a node has. The heatmap uses a color scale ranging from dark blue to light blue to represent the relative degree centrality of each node, with higher values shown in darker shades.

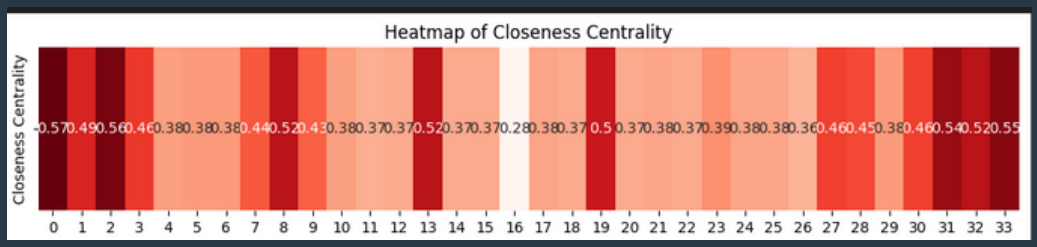

Image 6: The Heatmap of Closeness Centrality depicts the closeness centrality values for each node in the network. Closeness centrality measures how close a node is to all other nodes in the network. The heatmap uses a color scale ranging from dark red to light orange to represent the relative closeness centrality of each node, with higher values shown in darker shades.

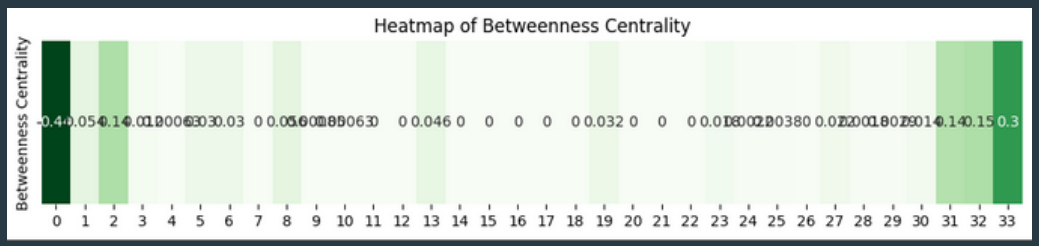

Image 7: The Heatmap of Betweenness Centrality visualizes the betweenness centrality values for each node in the network. Betweenness centrality measures how often a node lies on the shortest paths between other nodes. The heatmap uses a color scale to represent the relative betweenness centrality of each node, with higher values shown in darker shades of green.

Conclusion

This project has provided a comprehensive analysis of the Karate Club graph, uncovering key insights into the network's structure and dynamics. Through the use of various centrality measures, community detection algorithms, and visualizations, we have gained a deeper understanding of the social interactions within the karate club.

As I continue to refine and expand this project, further developments will be made to enhance the analysis and visualization techniques. These improvements will aim to provide even more detailed insights and a better understanding of the network's properties. The project will remain under active development until the final presentation day, ensuring that the latest methodologies and findings are incorporated.

Project Github link:
click here

Thank You