GRAPH - Graph Mining - CS - PARIS - SACLAY (2023-2024)

Exploring Social Networks with Graph Theory

Lucas J. Velôso and Paul Canal

Professor: Nacéra Seghouani

March 19, 2024

Introduction

This project aims to analyze social network dynamics through graph theory, focusing on Facebook 'circles'. We explore various graph properties and centrality measures to uncover underlying patterns and community structures within the network.

Figure: The source code for our project can be found on GitHub clicking here.

Dataset Overview

The dataset from Stanford's SNAP features anonymized Facebook profiles, circles, and ego networks. With 4,039 nodes and 88,234 edges, it provides a rich playground for our analysis, emphasizing the network's average clustering coefficients and triangles.

facebook

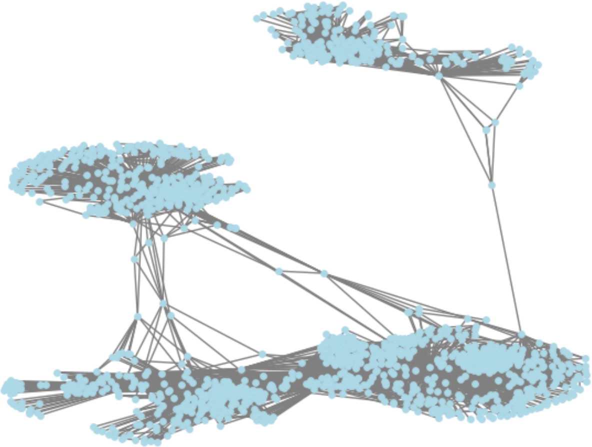
from
Meta

Methodology

- **Data Preparation:** Downloading and extracting the dataset.
- **Centrality Analysis:** We examine degree centrality, eigenvector centrality, PageRank centrality, and others to evaluate node importance.
- **Community Detection:** Implementing algorithms to detect communities within the network, using techniques like k-Core Decomposition, the Louvain Algorithm, and Girvan-Newman Algorithm.

Graph Visualization

Graphical representation of the network, highlighting its complex structure.



Centrality Analysis

Degree Centrality

Degree centrality measures the number of edges incident upon a node. It indicates the immediate risk of a node for catching whatever is flowing through the network.

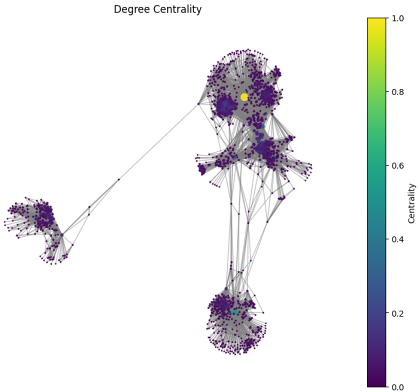


Figure: Nodes with high degree centrality can be seen as popular or influential figures within the network, having numerous direct connections.

Eigenvector Centrality

Eigenvector centrality measures a node's influence based on the number and quality of its connections. A node is considered highly influential if it is connected to many other influential nodes.

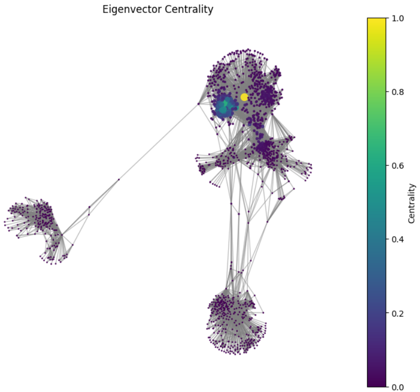


Figure: Using this graph, we can identify not just the most connected nodes, but also those connected to a network of well-connected peers.

Closeness Centrality

Closeness centrality measures how close a node is to all other nodes in the network. It is defined as the reciprocal of the sum of the shortest path distances from a node to all other nodes.

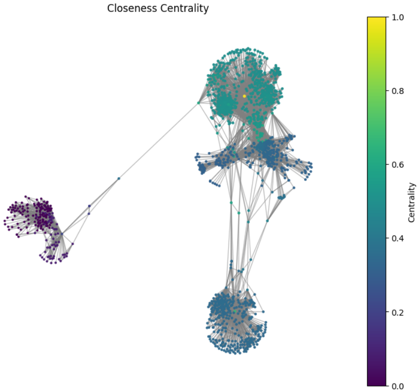


Figure: Nodes with high closeness centrality can quickly interact with all others directly or indirectly, signifying efficient information dissemination.

PageRank Centrality

PageRank centrality assesses the importance of a node within the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question.

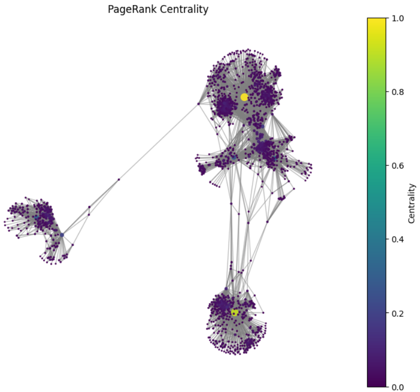


Figure: This graph helps identify not just well-connected nodes, but also those with qualitatively important connections, indicating a form of recursive

Community Detection

k-Core Decomposition

k-Core Decomposition identifies clusters based on the core level, where a k-core of a graph is a maximal connected subgraph in which all vertices have degree at least k. This method is useful for understanding the graph's structure and identifying influential communities.

Figure: Peeling off layers of less connected nodes to simplify graph analysis and identify influential communities. Here, we used k=5.

Louvain Algorithm

The Louvain Algorithm detects communities by maximizing modularity through a heuristic approach. It is iterative, aggregating nodes into communities at various scales, and is known for its hierarchical clustering capability.

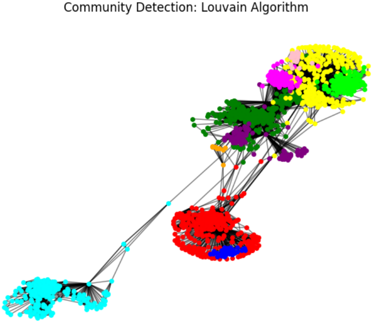


Figure: Hierarchical clustering approach to detect communities by maximizing modularity.

Girvan-Newman Algorithm

The Girvan-Newman algorithm identifies community structures by iteratively removing edges based on the edge betweenness centrality. This method progressively splits the graph into communities until the desired structure is obtained.

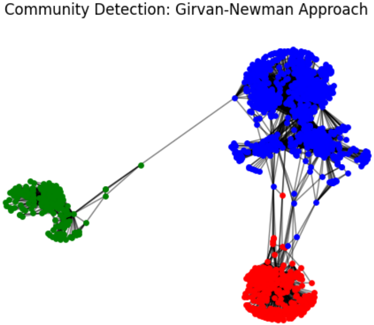


Figure: Progressive identification of community structures through iterative edge removal.

Conclusion

- Our analysis has delved into the intricate dynamics of social networks through graph theory, focusing on Facebook 'circles'.
- By examining various centrality measures and community detection algorithms, we've identified key patterns and structures within the network.

For more detailed information and access to our data analysis scripts, visit our GitHub repository: SVJLucas/GraphMining.

Questions?

Thank you for your attention!
Any questions?