

Jaejoong Kim

Professor Kontothanassis

DS 210

14 December 2023

Final Project Report: Analysis of Social Circles in the Facebook Dataset

For my final project in DS210, I used the “Social Circles: Facebook” dataset, which encompasses a network of 4039 nodes and 88234 edges, representing mutual friendships on Facebook. The dataset offers a comprehensive look at user connections, featuring node features, circles, and ego networks. My interest in social networking and its implications in career development led me to select this dataset, particularly to explore the underlying patterns of connectivity and social reach within a large social network.

1. Project Objective

The primary goal was to analyze the Facebook dataset to uncover insights into the average distances between users, shortest paths for new connections, overall network connectivity, and unique characteristics of the most connected individuals. This exploration intended to reveal the underlying structure of social networks and their implications in real-world networking.

2. Methodology

a. Data Processing and Graph Construction

The dataset was processed using Rust programming, focusing on constructing an undirected graph to represent the Facebook network. Each node corresponded to a user, and edges represented mutual friendships. Key functions implemented in the codebase include:

- ``read_file``: Reads and imports the data, transforming it into a list of edges.
- ``Graph::create_undirected``: Constructs an undirected graph from the edge list.

b. Analysis Techniques

- Breadth-First Search (BFS): Utilized to traverse the graph, starting from a node and exploring all reachable nodes. This approach helped in calculating the average distance between pairs of vertices.
- Average Distance Computation: The ``compute_average_distance_bfs`` function aggregated distances and computed the average, providing insights into the typical separation between users in the network.
- Subgraph Analysis: To understand variations in connectivity based on network size, I created subgraphs of varying node counts (100, 200, 300, 500, 1000, 2000, 3000, 4000) and calculated average distances within these subgraphs.
- Component Analysis: Employing BFS, the number of connected components in the graph was determined, revealing the level of social cohesion in the dataset.

3. Results and Observations

```

Finished dev [unoptimized + debuginfo] target(s) in 0.02s
Running `target\debug\DS210_Final_Project.exe`
Vertex with highest connectivity has 1045 connections.
Vertex with lowest connectivity has 1 connections.
Graph's average connectivity: 43.69
Total connected components in the graph: 1
Average distance among nodes in a subgraph of 100 nodes: 1.7495005
Average distance among nodes in a subgraph of 200 nodes: 2.3522105
Average distance among nodes in a subgraph of 300 nodes: 3.5051908
Average distance among nodes in a subgraph of 500 nodes: 5.358043
Average distance among nodes in a subgraph of 1000 nodes: 5.389848
Average distance among nodes in a subgraph of 2000 nodes: 5.08051
Average distance among nodes in a subgraph of 3000 nodes: 4.1217813
Average distance among nodes in a subgraph of 4000 nodes: 3.7007093
Average distance among all pairs of nodes: 3.6925068

```

- Average Distance: The average distance between pairs of vertices in the entire dataset was found to be approximately 3.69, suggesting a relatively small world phenomenon within the Facebook network.

- Subgraph Analysis: A parabolic pattern was observed in the average distances within subgraphs. The distance initially increased with the number of nodes, peaking at 500 nodes, and then decreased as the node count approached 4000.
- Component Analysis: The entire graph constituted a single connected component, indicating high interconnectivity.
- Degree Analysis: The most connected individual had 1045 connections, while the least connected had only 1. This disparity highlighted the presence of influential nodes in the network.

4. Conclusion

The analysis of the Facebook dataset revealed significant insights into the structure of social networks. The average distance of around 4 people aligns closely with the concept of 'six degrees of separation'. The subgraph analysis showed that network size influences connectivity patterns, displaying a parabolic trend in average distances. The single-component nature of the graph underscores a high level of interconnectedness among users. These findings contribute to understanding social networks' dynamics and can inform strategies for networking and community building.

5. How to Run the Project

- Setup: Ensure Rust and Cargo are installed on the system.
- Compilation: Navigate to the project directory and run ``cargo build``.
- Execution: Execute the program using ``cargo run``.
- Output: The program outputs various statistics, including average distances, component counts, and degree information.