Joyce Bozeko

joyceb@vt.edu

CS 5664 Homework 1

# Presentation of the dataset

Facebook\_combined.txt.gz is the dataset that contains a graph of the social network of Facebook users. It is a graph where each node represents a user, and each edge represents a friendship connection between two users. In this text file, the edges are represented as pairs of node IDs, where each pair represents a connection between two users.

# Basic analysis

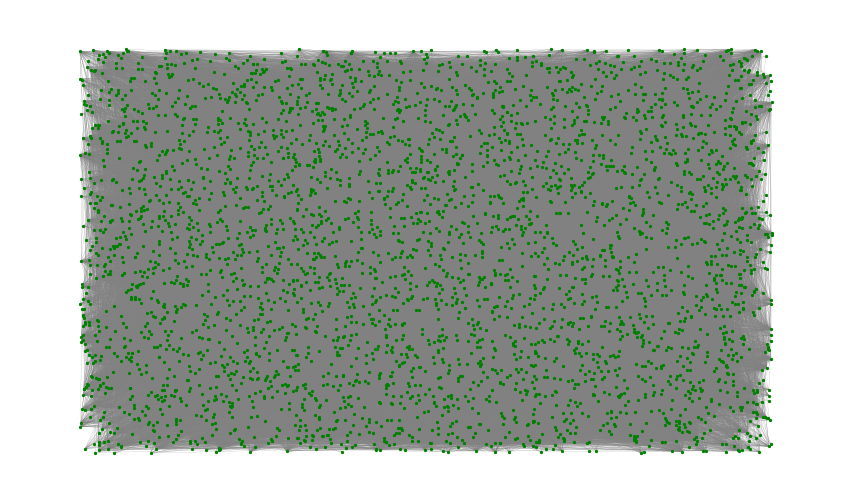
First of all, after loading the graph into a networkx object, we should look at the general characteristics of the graph.

**The number of nodes:** 4039

**The number of edges:** 88234

**The average degree:** 43.691

Bellow we have the plot of that graph:



# Centrality analysis

**-Degree centrality**: after analysis, the node with the maximum degree in the graph is the node 107. That node is linked to 25.8% of the total nodes of the graph

Top 5 nodes: (107, 0.258), (1684, 0.196), (1912, 0.186), (3437, 0.135), (0, 0.0859).

The degree centrality histogram shows that majority of Facebook users have degree centralities under 0.02.

**-Betweenness centrality**: after analysis, the node 107 has a betweenness centrality of 0.48, meaning it lies on almost half of the total shortest paths between other nodes. That shows that the node is most frequent in the shortest paths between 2 nodes

Top 5 nodes: (107, 0.48), (1684, 0.33), (3437, 0.23), (1912, 0.22), (1085, 0.14)

The betweenness centrality histogram shows that most nodes have betweenness centralities below 0.05

**-Closeness Centrality**: The graph analysis show us that the node 107 hasthe best closeness centrality, meaning it lies on almost half of the total shortest paths between other nodes. That shows that the node is most frequent in the shortest paths between 2 nodes

Top 5 nodes: (107, 0.45), (58, 0.39), (428, 0.39), (563, 0.39),(1684, 0.39),

The closeness centrality histogram shows that closeness centralities are distributed over various values.

# Clustering coefficient

**Average clustering coefficient**: 0.605

**The number of triangles**: 1612010

# Heavy-tail analysis

**-degree statistics:**

**Mean degree:** 43.691

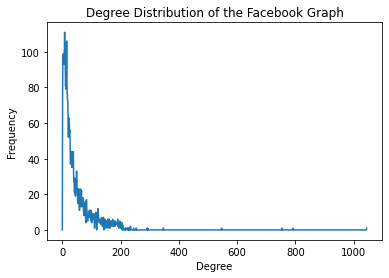
**Standard deviation of degree:** 52.414

**Minimum degree:** 1

**Maximum degree:** 1045

**Degree assortativity**: 0.0635

Bellow graph shows the degree histogram:



The power-law exponent is 5.99

The goodness-of-fit of the power-law distribution is 0.351

# Comparison with a Watts-Strogatz graph

The Watts-Strogatz model is a well-known model for generating random graphs that exhibit small-world properties. The model generates a graph with a small average shortest path length and a high degree of clustering, like many real-world networks. We generated a Watts-Strogatz graph with similar parameters and compare the average shortest path length, clustering coefficient, and degree distribution of the two graphs.

The average shortest path length of the Facebook graph is 3.6925068496963913

The average shortest path length of the Watts-Strogatz graph is 2.997013148547575

The average clustering coefficient of the Facebook graph is 0.60

The average clustering coefficient of the Watts-Strogatz graph is 0.53

We can see that the facebook graph and the Watts-Strogatz graph have almost the same average clustering coefficient but there is a big difference for the average shortest path.