Griffin Austin

Trey Grossman

Caitlynn Koback

Jonathan Gregg

Our group decided to use the Facebook Social Circles dataset. We are interested in this dataset because it represents applicable, real-world data. No pre-processing was performed on our part, however the data was anonymized by the data aggregators. We would expect nodes with a high centrality value to represent individuals connected who are friends with many other people; that is, people who are central to their friend groups. The graph of degree distribution should have a power law relationship, as there will be many people with few friends and few people with many friends. We would expect this graph to have the small-world property, because although most people are not friends with most people on the site, most people can be connected with someone else through a small number of acquaintances. In fact, there is a notion that any two people can be connected by six or fewer intermediaries, known as the Six Degrees of Separation.

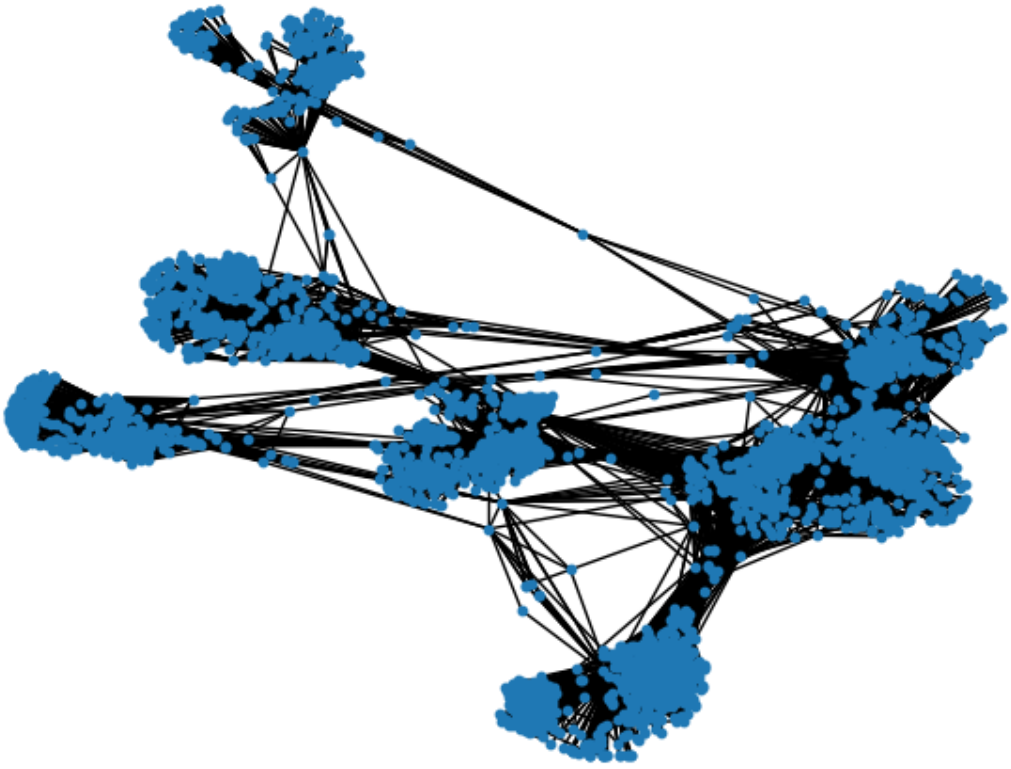


Figure 1: Graph visualization of data

The ten nodes with the highest degree are shown in table 2.

Node index	**Degree**
107	1045
1684	792
1912	755
3437	547
0	347
2543	294

2347	291
1888	254
1800	245
1663	235

Table 2: Top 10 nodes and their degree

The ten nodes with largest betweenness centrality are shown in table 3.

Node index	**Betweenness centrality**
107	0.4805
1684	0.3378
3437	0.2361
1912	0.2293
1085	0.1490
0	0.1463
698	0.1153
567	0.0963
58	0.0844
428	0.0643

Table 3: Top 10 nodes and their betweenness centrality

Find the 10 nodes with the highest clustering coefficient. Ties were solved by smallest node index.

Node index	**Clustering Coefficient**
32	1.0
33	1.0

35	1.0
42	1.0
44	1.0
46	1.0
47	1.0
52	1.0
63	1.0
70	1.0

Table 4. Top 10 nodes and their clustering coefficient

Find the top 10 nodes as ranked by eigenvector centrality.

Node index	**Eigenvector centrality**
1912	0.09540696149067629
226	0.08698327767886552
2206	0.08605239270584342
2233	0.08517340912756598
2464	0.08427877475676092
2142	0.08419311897991796
2218	0.0841557356805503
2078	0.08413617041724979
2123	0.08367141238206226
1993	0.0835324284081597

Table 5. Top 10 nodes and their eigenvector centrality

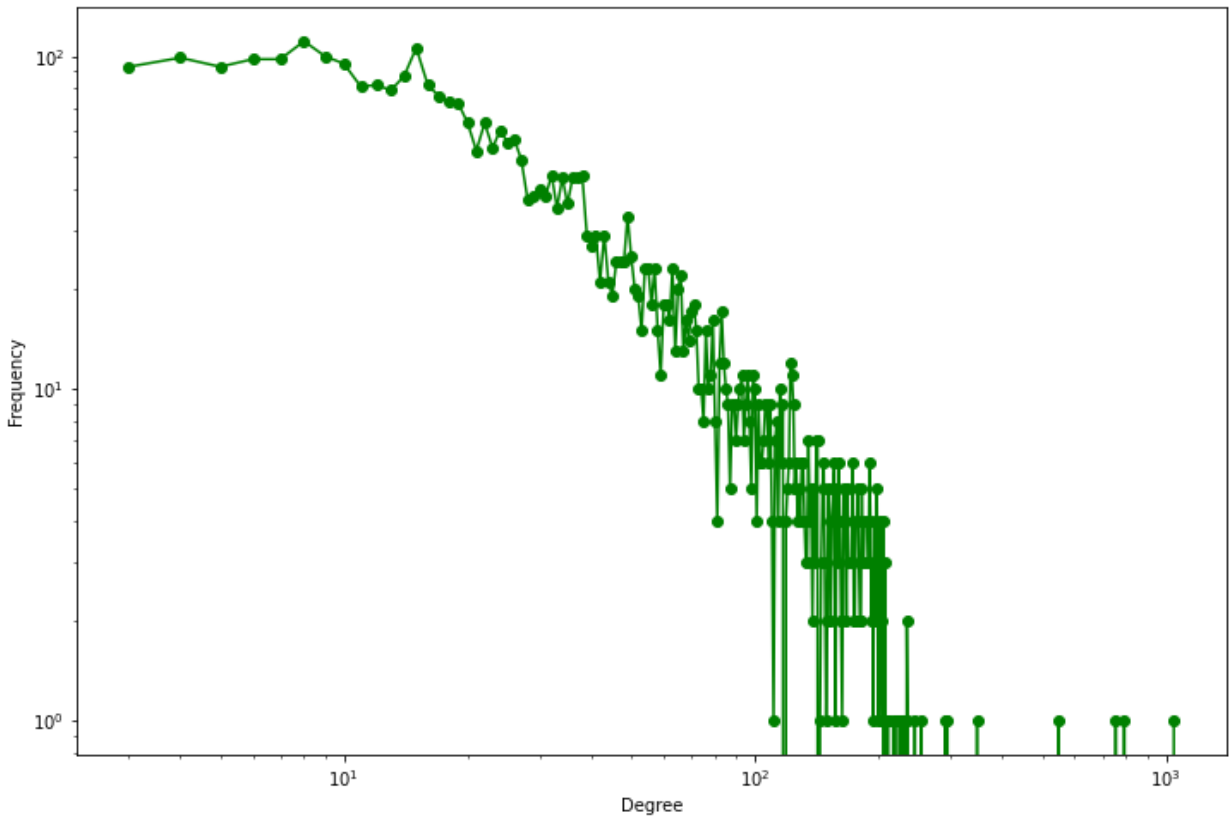
Find the top 10 nodes as ranked by Pagerank.

Node index	**Pagerank**
3437	0.0076145868447496
107	0.006936420955866117
1684	0.006367162138306824
0	0.006367162138306824
1912	0.003876971600884498
348	0.002348096972780577
686	0.002219359259800019
3980	0.0021703235790099928
414	0.0021703235790099928
698	0.0013171153138368812

Table 6. Top 10 nodes and their Pagerank ranking

The Pagerank, Betweenness Centrality, and Degree tables share at least 50% of nodes between them. For the most part, the shared nodes within the categories do not vary significantly, and many share ranks across the tables or only differ by one placement. It makes sense for these categories to share nodes because the degree of the nodes plays a significant role in the calculation of betweenness centrality, and Pagerank relies on both of those categories in turn.

The graph's average shortest path length is fairly low at 3.6925068496963913, and the average clustering coefficient is relatively high at 0.6055467186200876. This implies the graph exhibits small-world behavior.



Graph 7: Degree distribution

Based on the Degree Distribution graph, the graph does not display power-law behavior.