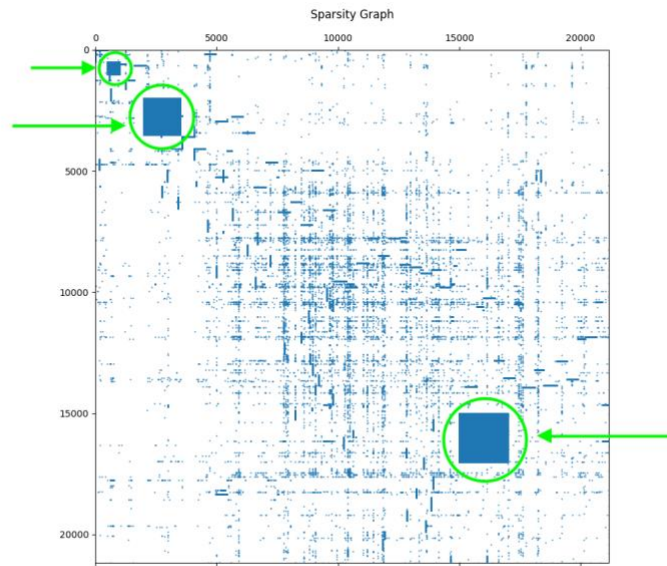


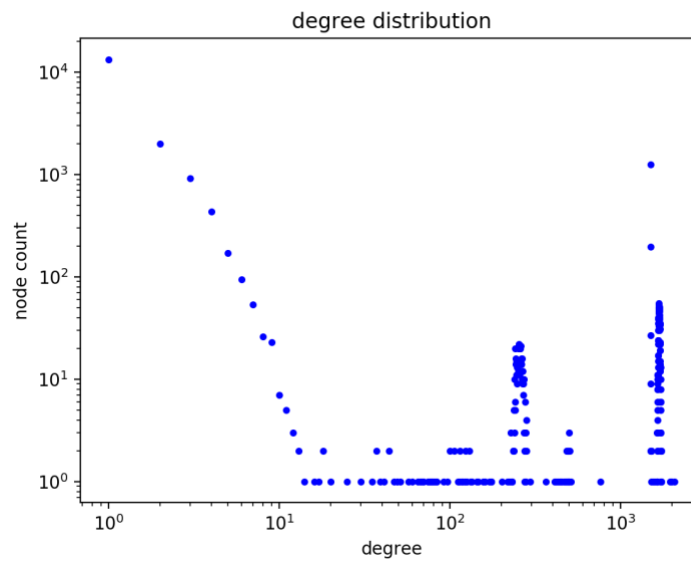
Paul Rodriguez
SID: 861152748

CS235 Assignment Report

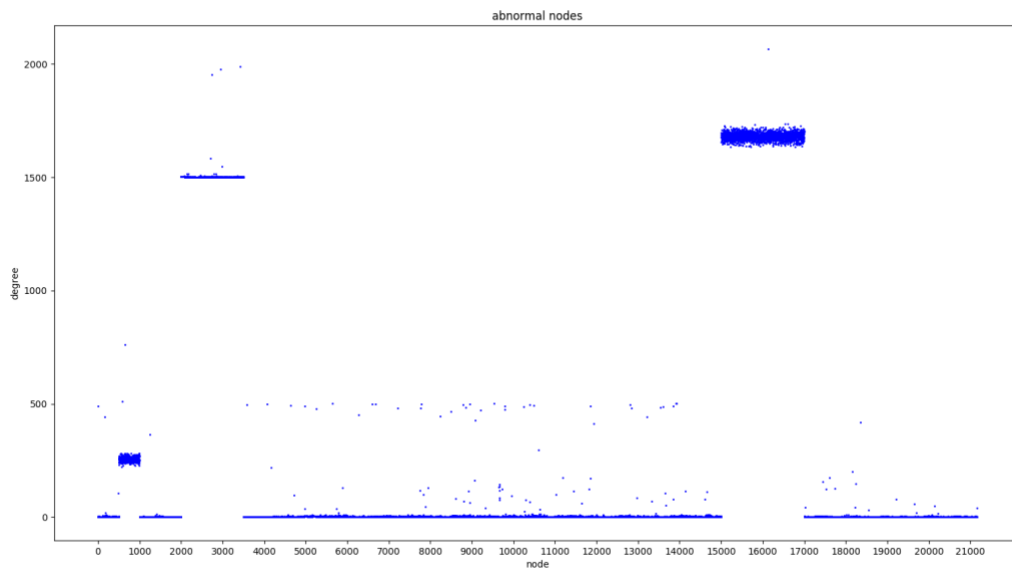
Question 1



Question 2



3. There are clear differences between this graph and the clean graph. This graph has an abundance of nodes with high degrees.



4. a. Based on the degree distribution, we have 3 abnormal blocks illustrated in the graph above. We plot the node with its respective degree and get a graph that compares all the nodes and their degrees.
- b. Nodes in the approximate range 750-1000 belong to block 1, nodes in the approximate range 2000-3400 belong to block 2, and nodes in the approximate range 15000-17000 belong to block 3.

Question 3

1. What is the main problem? The main problem for this paper arises from trying to understand social groups from a given network. These groups can follow certain trends and show up clustered in a dataset. This paper tries to find the different social groups that can exist in a network using the phone call network of mobile users. This problem also tries to understand how these groups are formed as well as trying to understand how new products and services can be offered to these distinct groups.

What is the main idea? The main idea of this paper is the understanding the phenomenon of 'spokes' that appear in graphs of singular vectors. The use of traditional methods such as spectral clustering and multi-level graph partitioning techniques did not yield good communities in the graphs so a different method of graphing had to be developed. This is where singular vectors come into play and show the different communities found in the networks.

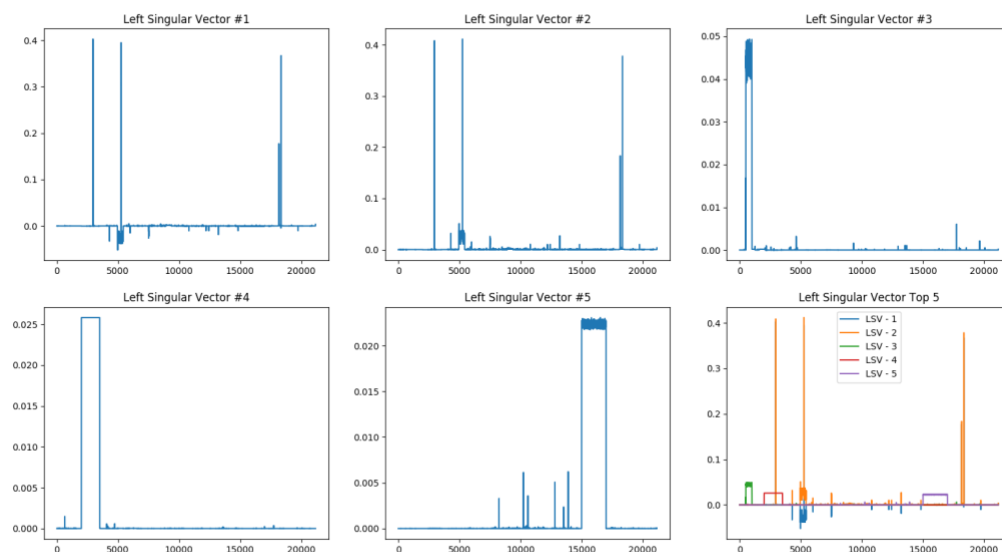
What are the results? Using singular value decomposition, they plot the U matrix against each other for every node and find lines that are straight and aligned to each axis. This the pattern they call as "EigenSpokes". They tested to see if this phenomenon appeared in other real-world graphs and surprisingly obtained similar results. Using these types of graphs, the authors

were able to find relevant communities in each of the datasets and were able to see how big these communities were.

2. a. The Rank for which 90% of the original data is reconstructed correctly is 80. The sparsity graph precision is 0.1 meaning that any values less than this value will not be plotted.



- b. The first 5 graphs (right-to-left, top-to-bottom) are the graphs corresponding to the top 5 left singular vectors graphed individually and the bottom right-hand graph is all of the left singular vectors graphed together.



- c. The abnormal blocks found in our graph are the big spikes in the last colorful graph.
- d. The first 5 graphs (right-to-left, top-to-bottom) are the graphs corresponding to the top 5 matrices graphed individually and the bottom right-hand graph is all of the matrices graphed together.

