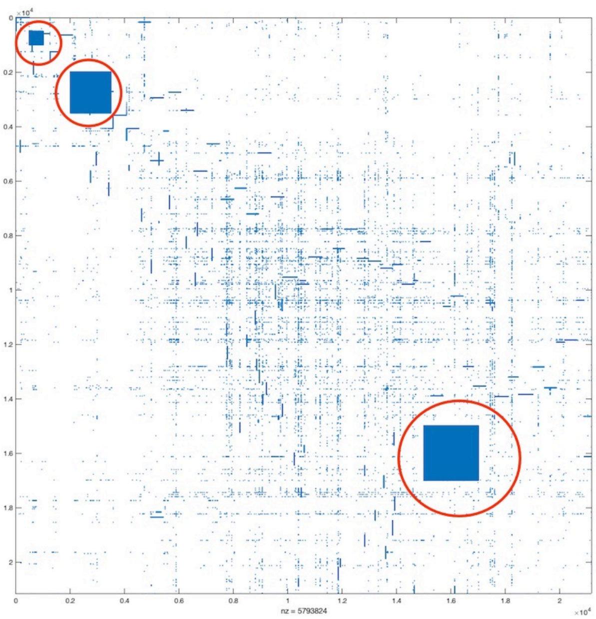
CS 235 Assignment

Yujia Zhai
SID: 862120464

Question 1:

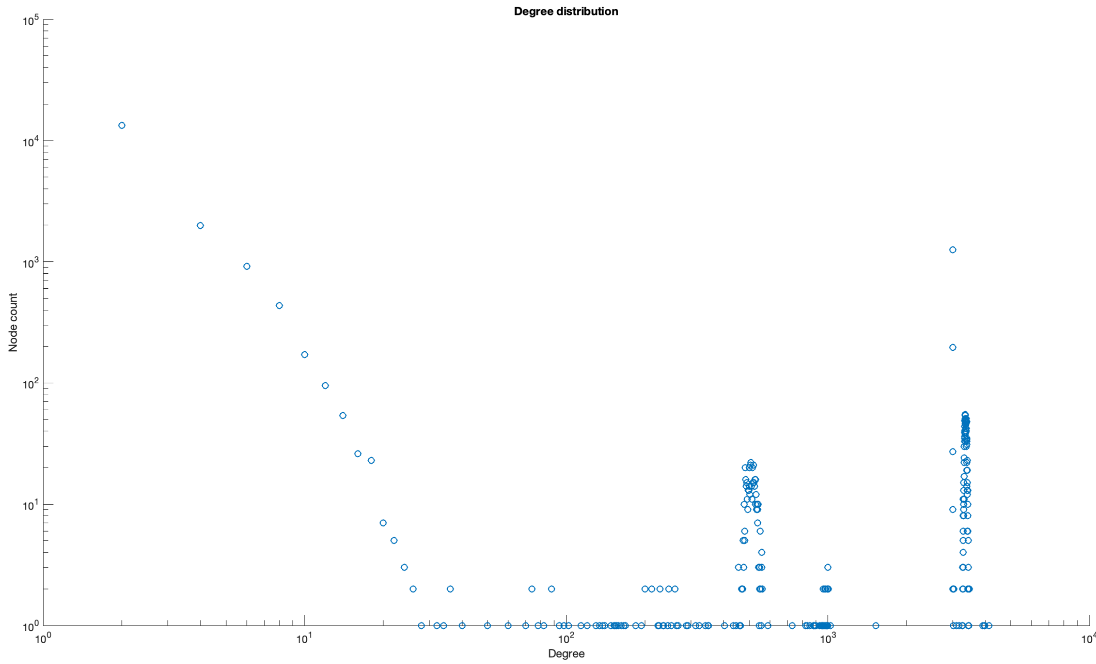
1. See the code deliverable.
2. Spy plot, with the abnormal blocks circled.



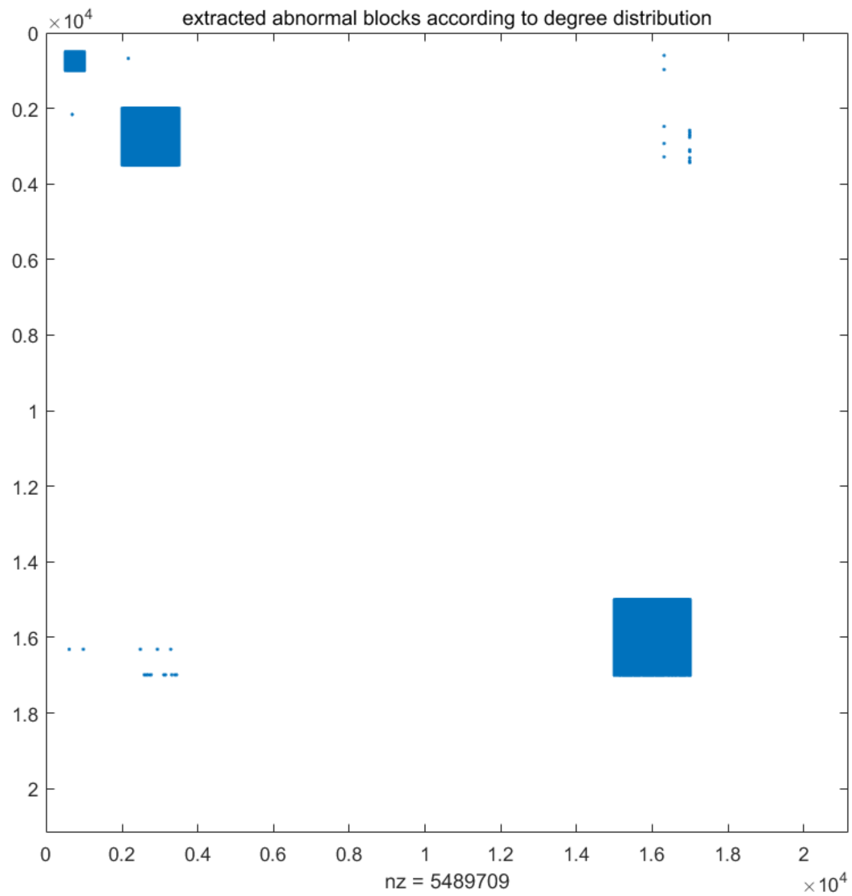
3. Three abnormal blocks are already circled.

Question 2:

1. See the code deliverable.
2. Degree distribution in log-log scale.



3. Compared with the clean data, we saw abnormal degrees when degree > 100 and degree count > 5. Say, in the original clean distribution when degree > 100, there's no a circumstance that degree count > 5 for that degree.



4. This (when degree > 100, there's no a circumstance that degree count > 5 for that degree) became the judge criteria for us to local abnormal nodes and to give a spy-plot for abnormal nodes. See the figure below; three abnormal blocks are extracted based on the degree distribution figure.

   So we have:
   a. Three abnormal blocks.
   b. basically there are three abnormal node intervals: [500 : 1000], [2000 : 3500], [15000 : 17000].

Question 3: SVD

1. Literature Survey

a. **Main problem**: When trying to detect/extract communities in large graphs, traditional methods including Spectral Clustering, Co-clustering (on bipartite graphs) are not able to work well on large graphs detections or have some obvious limitations. For example, spectral clustering failed to give large internal coherence with long chains. Meanwhile, graph partitioning methods lack internal coherence as well, which is NOT feasible “when applied to social graphs”.

b. **Main innovation**: This paper reports an interesting phenomenon: when plotting singular vectors of Mobile Call graph, those singular vectors often have clear sperate lines that align with axes. This phenomenon is called EigenSpoke. They also developed an community identification approach SpokEn to exploit EigenSpokes.

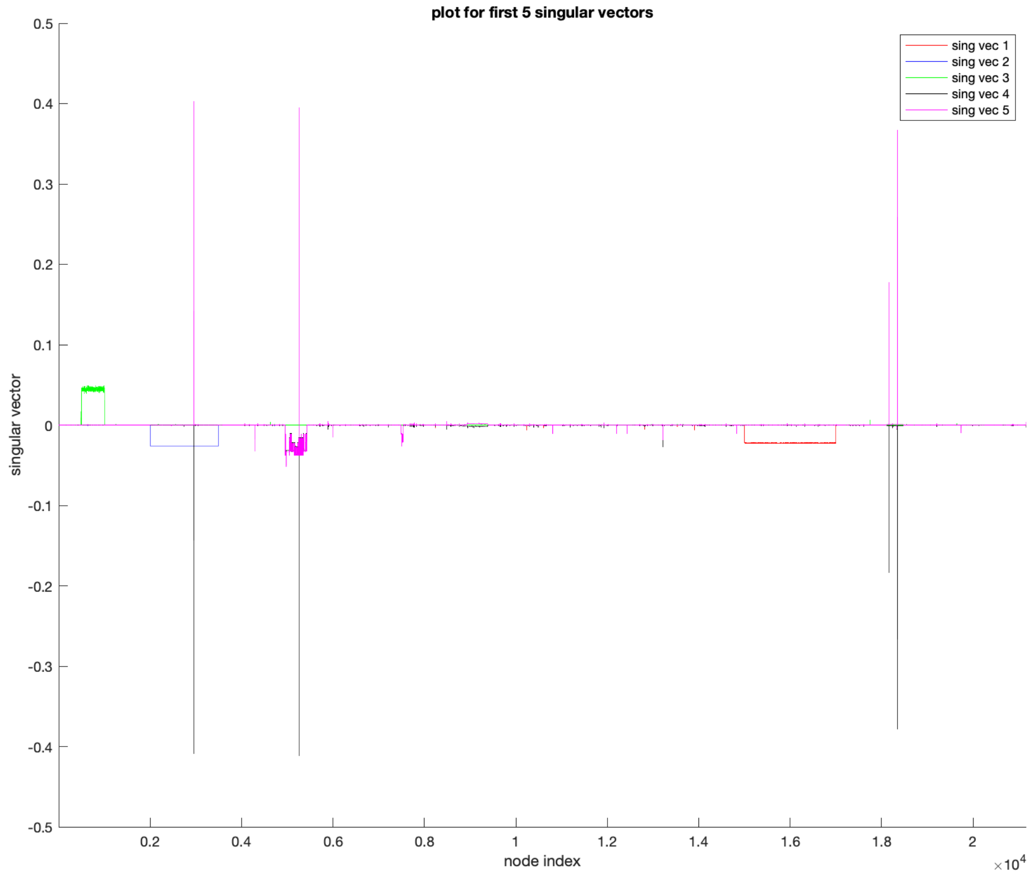
c. **Results**: EigenSpokes are strongly associated with the communities in sparse graphs, and their SpokeEn community-detection method based on EigenSpokes successes with real-world graphs.

2. Analysis.
   a. Find the number of singular value/vectors that reconstructed 90% of original graph.

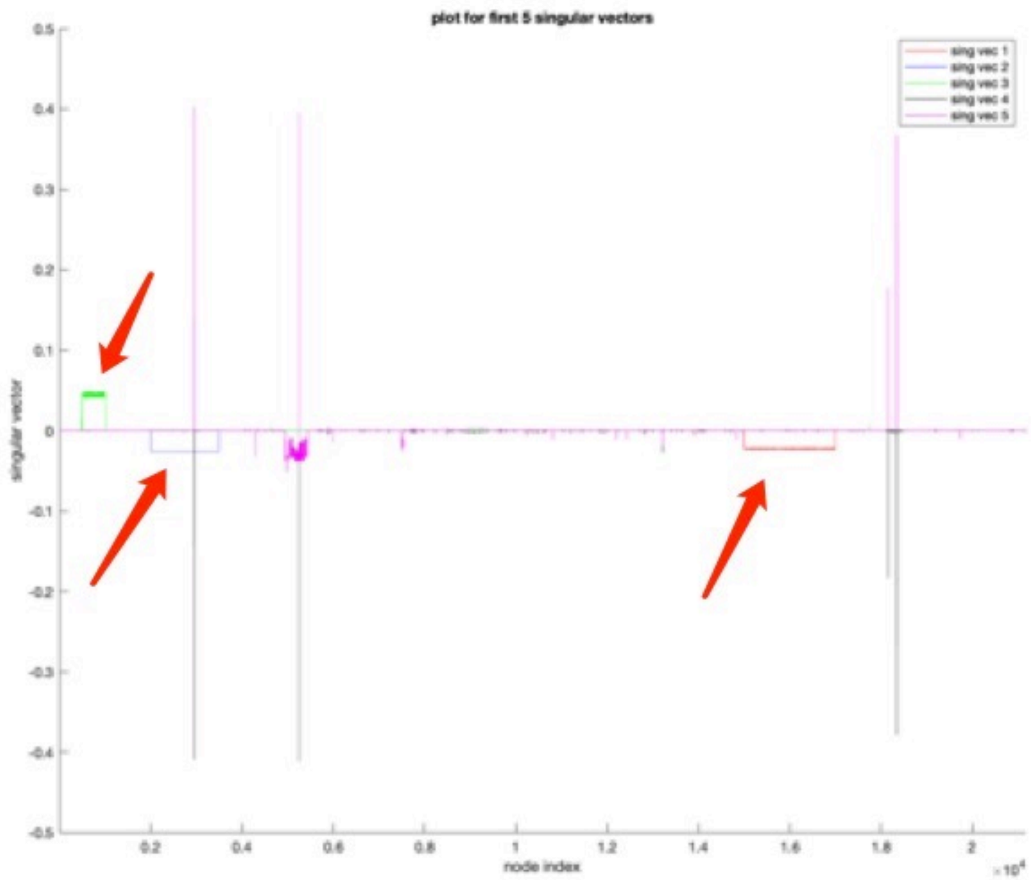
$$loss\ rate\ =\ \frac{||S_r - S_k||_2}{||S_k||_2}$$

   When loss rate <= 10%, I say it reconstructed 90%.
   According to the code in the deliverable, when r = 3, the loss rate is 0.016574.

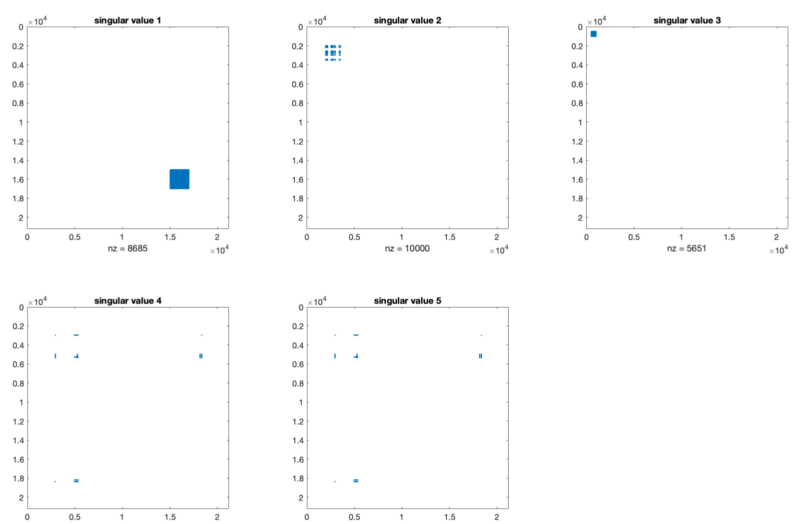
   b. Plot the left singular vectors (top 5 SVs).



c. There are three abnormal blocks. See from three arrows.


plot for first 5 singular vectors
sing vec 1
sing vec 2
sing vec 3
sing vec 4
sing vec 5
singular vector
node index

d.


singular value 1
nz = 8685
singular value 2
nz = 10000
singular value 3
nz = 5651
singular value 4
nz = 775
singular value 5
nz = 775