

Data 620 Week 2 Assignment 1

Enron Email Network Dataset Summary

The Enron Email Network dataset contain 36,692 nodes. The summary statistics are as follows:

Dataset statistics	
Nodes	36692
Edges	183831
Nodes in largest WCC	33696 (0.918)
Edges in largest WCC	180811 (0.984)
Nodes in largest SCC	33696 (0.918)
Edges in largest SCC	180811 (0.984)
Average clustering coefficient	0.4970
Number of triangles	727044
Fraction of closed triangles	0.03015
Diameter (longest shortest path)	11
90-percentile effective diameter	4.8

Project Resources:

GitHub Repository: <https://github.com/tagensingh/DATA620-W2-A1>

Jupyter Notebook: <https://github.com/tagensingh/DATA620-W2-A1/blob/main/W2-A2-ENRON.ipynb>

Gephi Project File: https://github.com/tagensingh/DATA620-W2-A1/blob/main/data620_gephi_w2_a2.gephi

Dataset Source:

- J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney. [Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters](#). Internet Mathematics 6(1) 29--123, 2009.
- B. Klimmt, Y. Yang. [Introducing the Enron corpus](#). CEAS conference, 2004.

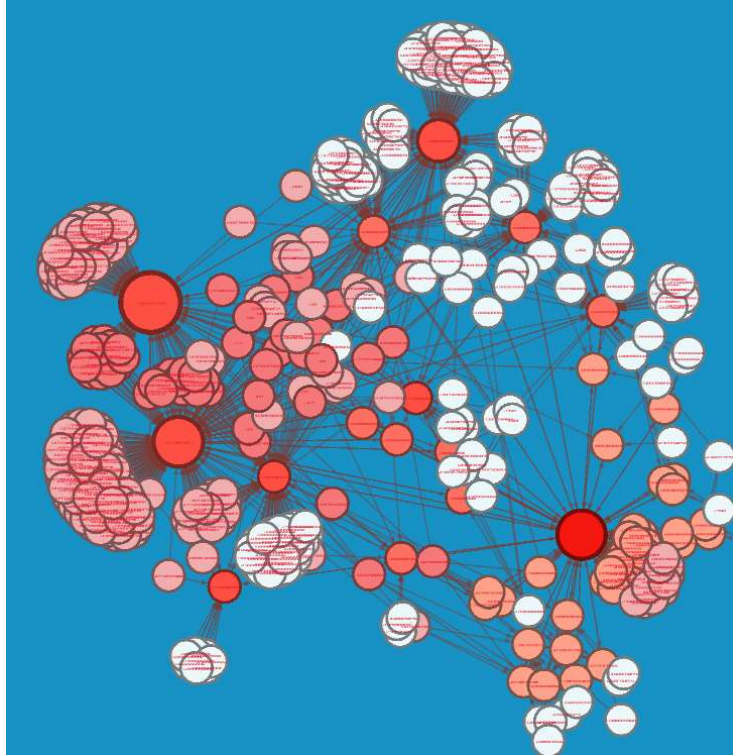
Collaborators:

Ramnivas Singh

Deepak Sharma

Tage Singh

Data subset – 500 records



Eigenvector Centrality Report

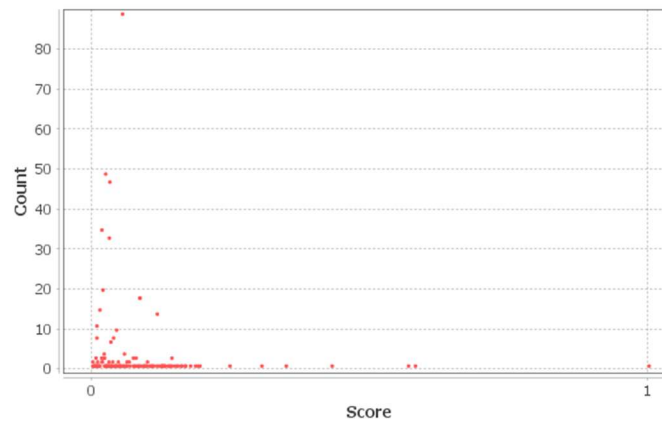
Parameters:

Network Interpretation: undirected

Number of iterations: 1000

Sum change: 3.1820143115450007E-4

Eigenvector Centrality Distribution

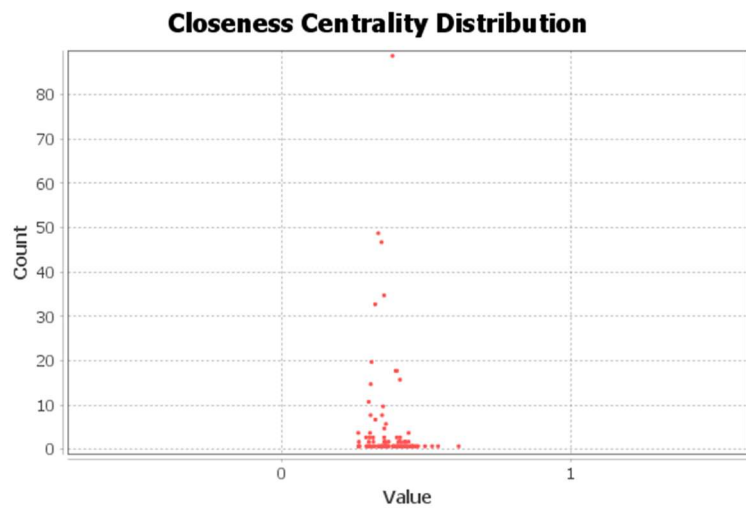
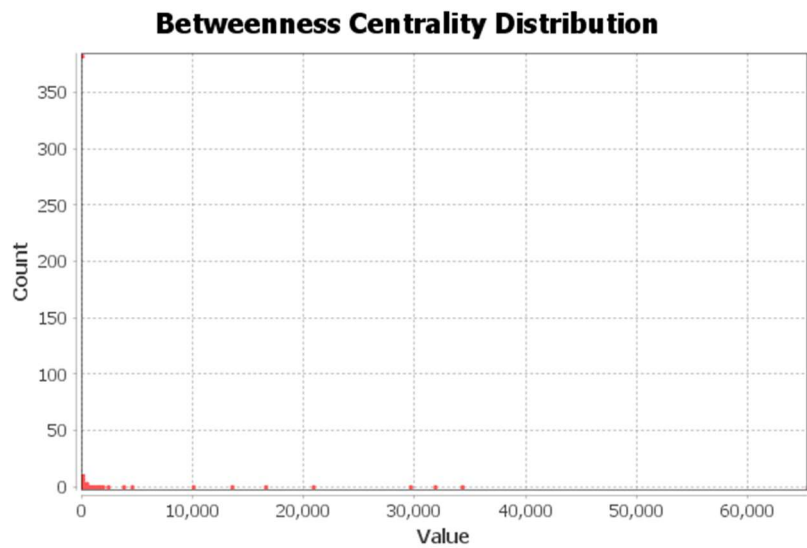


Graph Distance Report

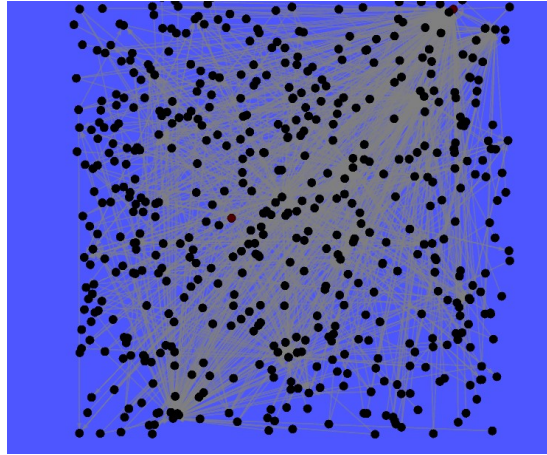
Parameters:

Network Interpretation: undirected

Results: Diameter: 4
Radius: 2
Average Path length: 2.871765568397447



Data subset – 1,000 records



Eigenvector Centrality Report

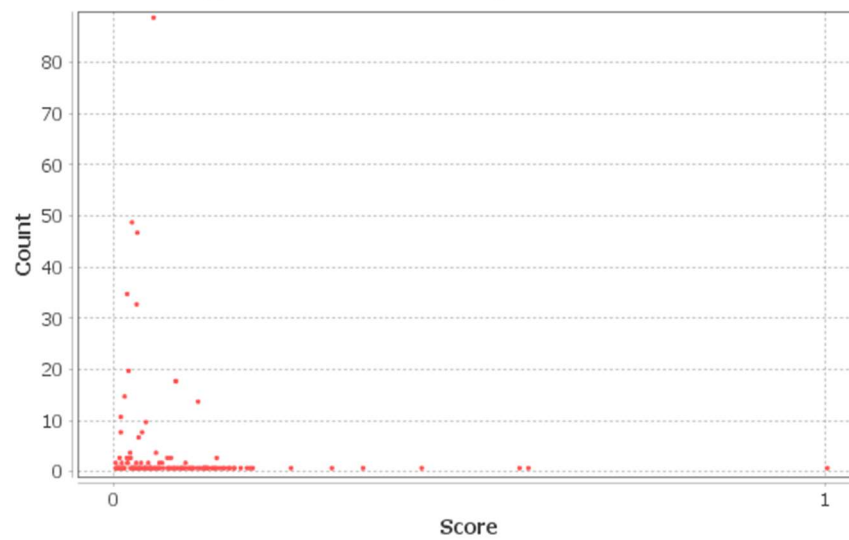
Parameters:

Network Interpretation: undirected

Number of iterations: 1000

Sum change: 3.1820143115450007E-4

Eigenvector Centrality Distribution



Graph Distance Report

Parameters:

Network Interpretation: undirected

Results:

Diameter: 4
Radius: 2
Average Path length: 2.871765568397447

