Network Analysis of Emails Sent Within Enron

Frank Acquaye

April 12, 2020

1 Network Summary

1.1 Data Source

I wanted to explore the network structure of emails sent within Enron. I used preprocessed data from SNAP.

1.2 Summary Statistics of Enron Emails

Table 1 shows the summary statistics of Enron Emails Network.

Figure 1 shows degree distribution
Figure 2 shows a visualisation of the network using the Force Atlas 2 Algorithm.
Figure 3 shows a visualisation of the network using the Force Atlas 2 Algorithm with node and edge colouring.
Figure 4 shows a visualisation of the network using the Yifan Hu Algorithm.
Figure 5 shows a visualisation of the network communities.

Upon inspection one realises a huge cluster within the centre of all visualisations and less connections in the extremities. It's my perception that these clusters are mostly key stakeholders in Enron and perhaps top management within Enron. Since the data used was not labelled this might be a bit hard to confirm.

Statistic Name	Value
Nodes/Order	36692
Edges/Sizer	183831
Max Degree	1383
Min Degree	1
Mean Degree	10.020
Number of Connected Components	1065
Radius	7
Diameter	13
Clustering Co-efficient	0.497
Order of largest connected component	33696
Size of largest connected component	180811
Ratio of Size of entire Graph to Connected Component	0.983
Ratio of Order of entire Graph to Connected Component	0.918
Total Number of traingles	727044.0

Table 1: Summary Statistics for Enron Emails Network

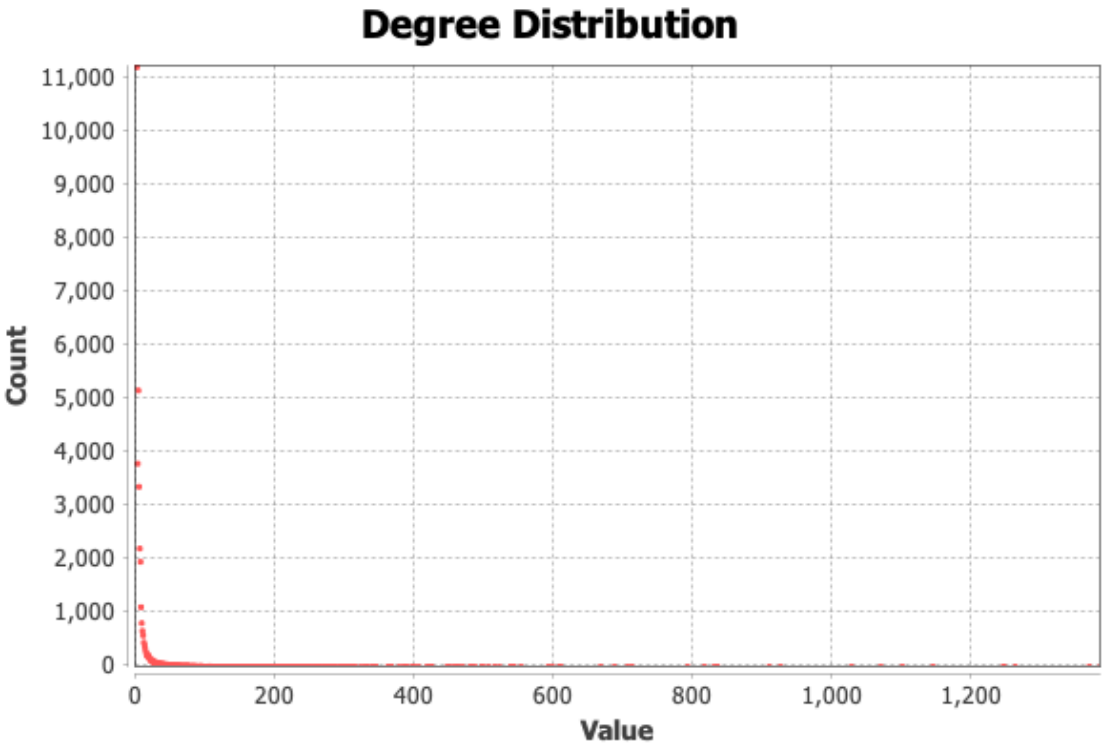

Figure 1: Degree Distribbution

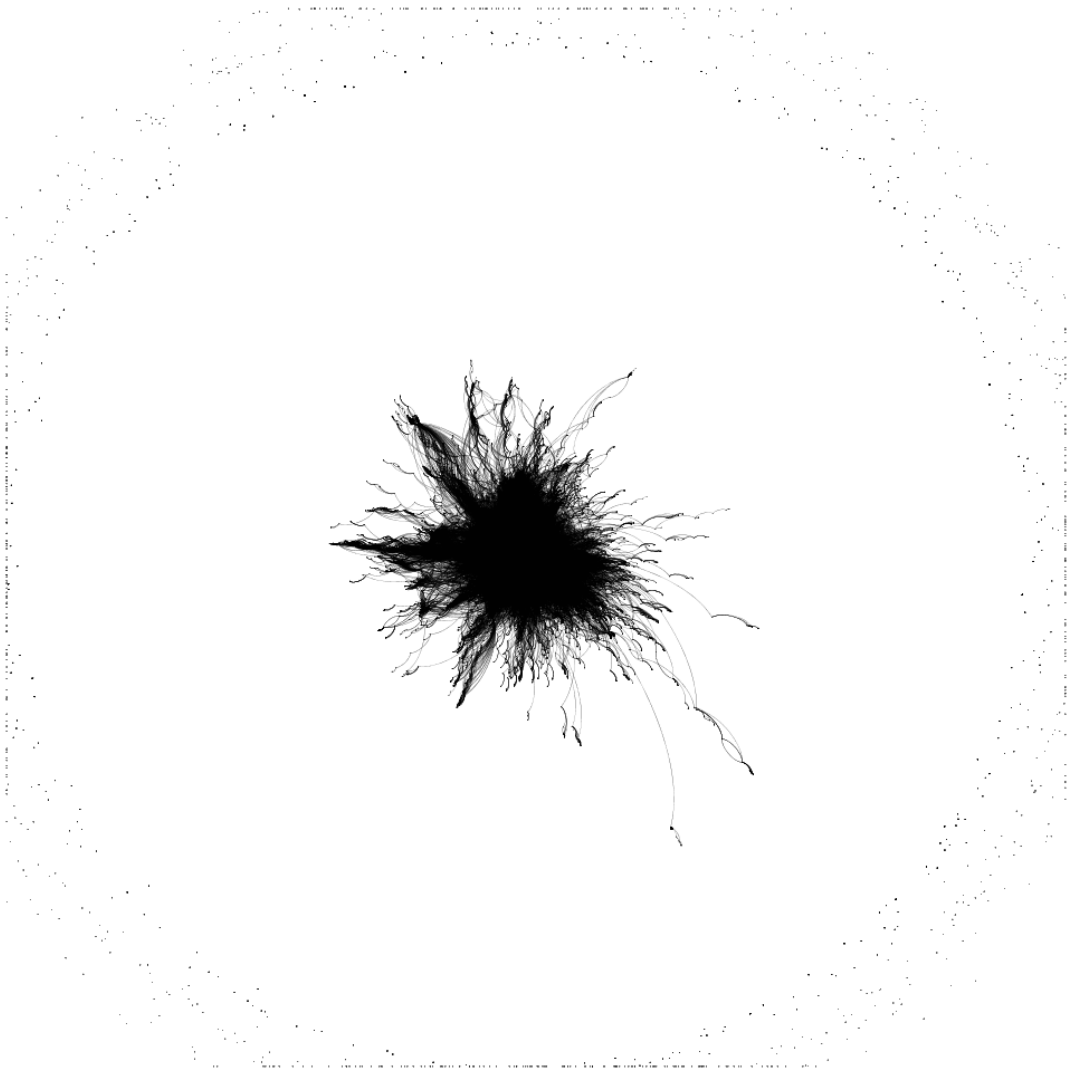

Figure 2: Network Structure of Enron Emails using Force Atlas 2

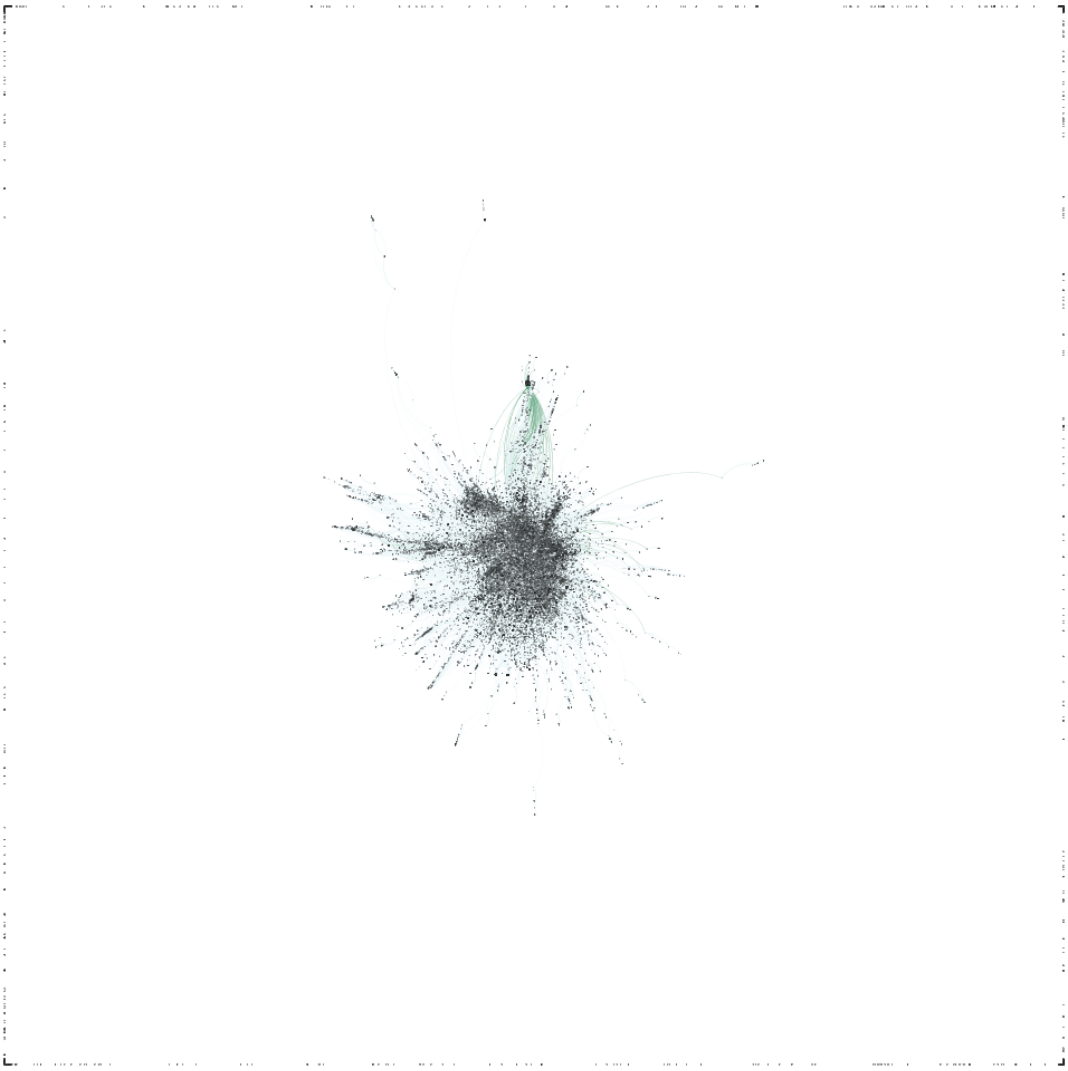

Figure 3: Network Structure of Enron Emails using Force Atlas 2 with colouring

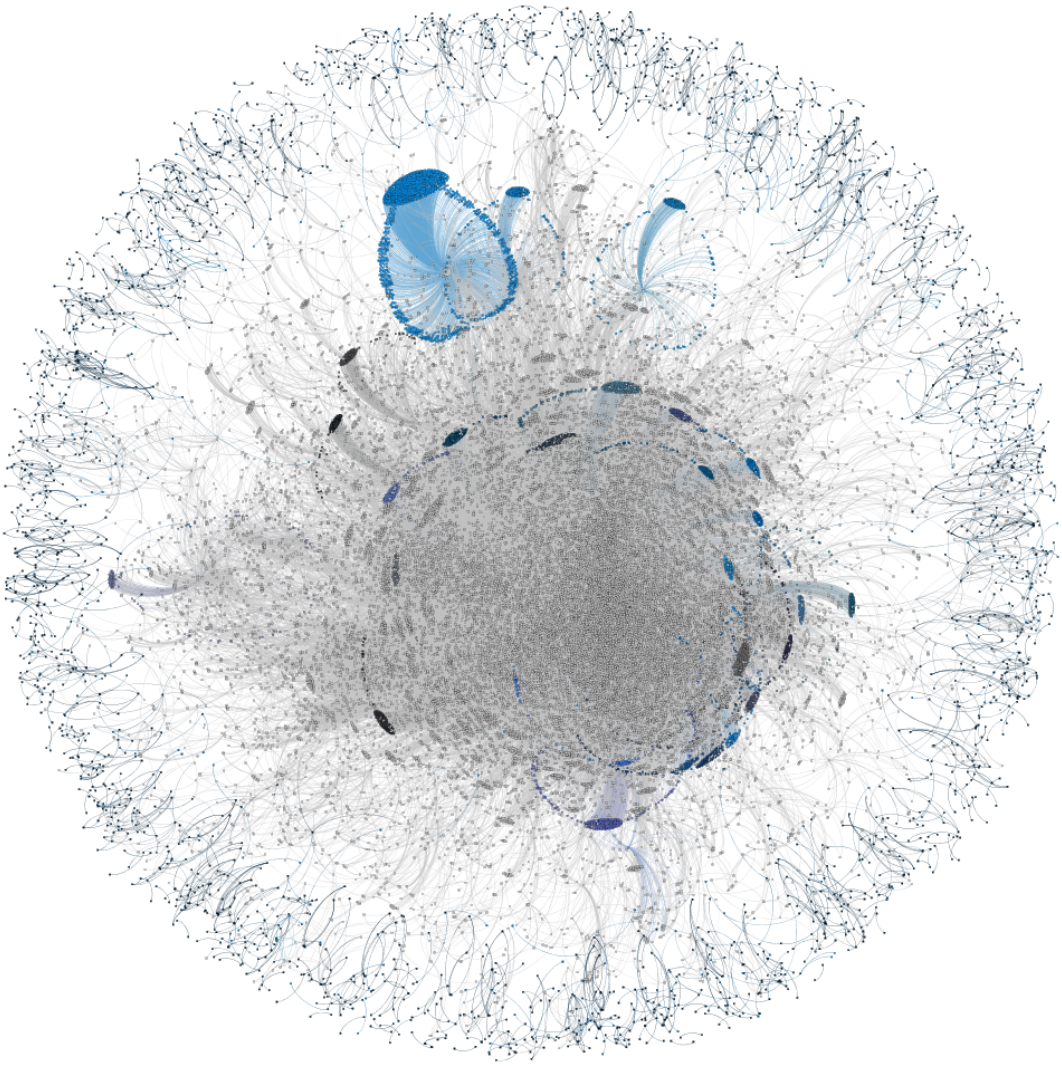

Figure 4: Network Structure of Enron Emails using Yifan Hu

2 Network Structure

Owing to the huge nature of the network most network, the assortative matrix was computed using a subgraph obtained from the first 250 nodes of the network according to the page-rank algorithm.

3 Community Detection

The image in *Figure 5* is incomprehensible hence I used the first 250 nodes of the network according to the page-rank algorithm.

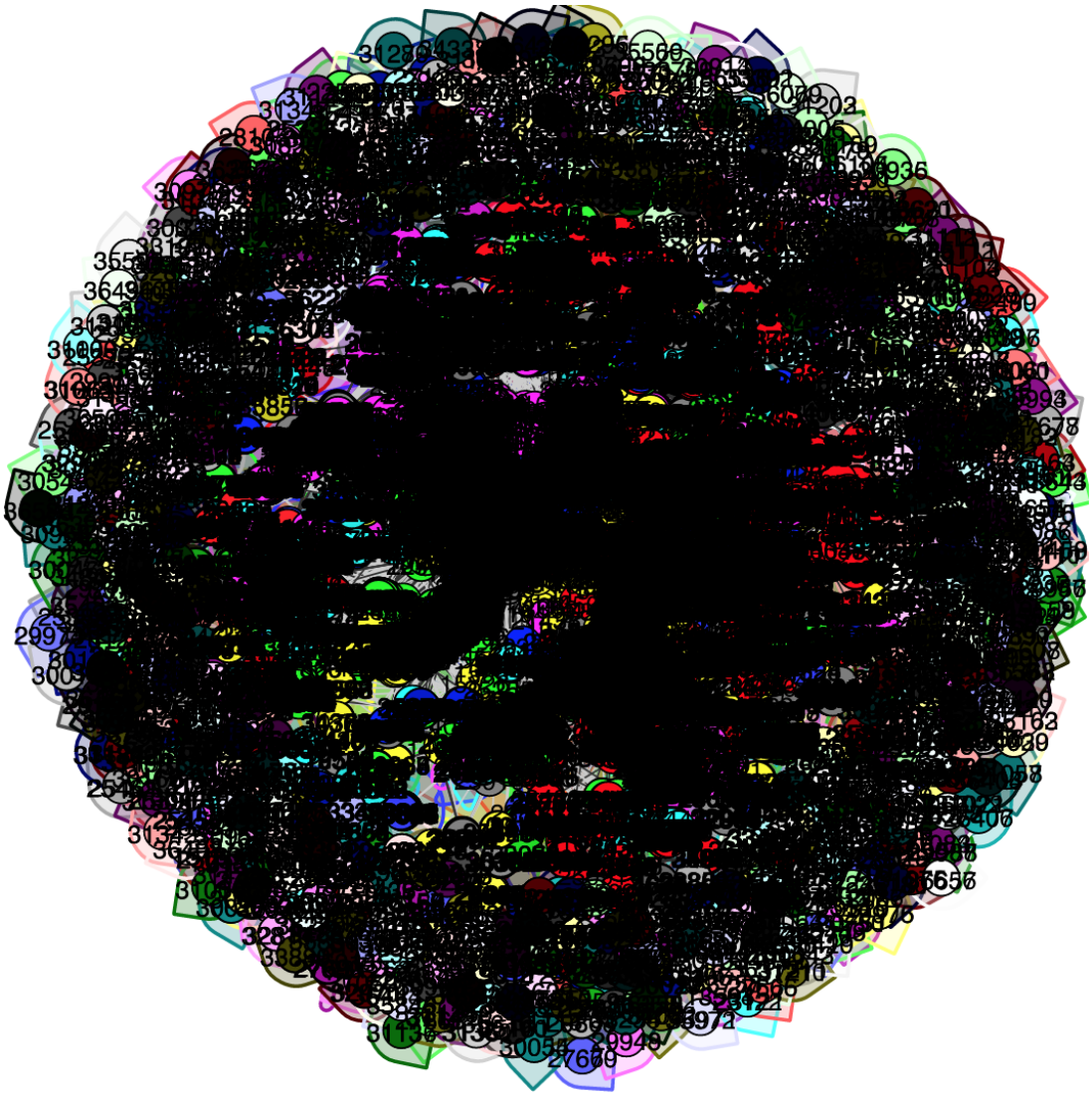

Figure 5: Communities in Enron Emails Network

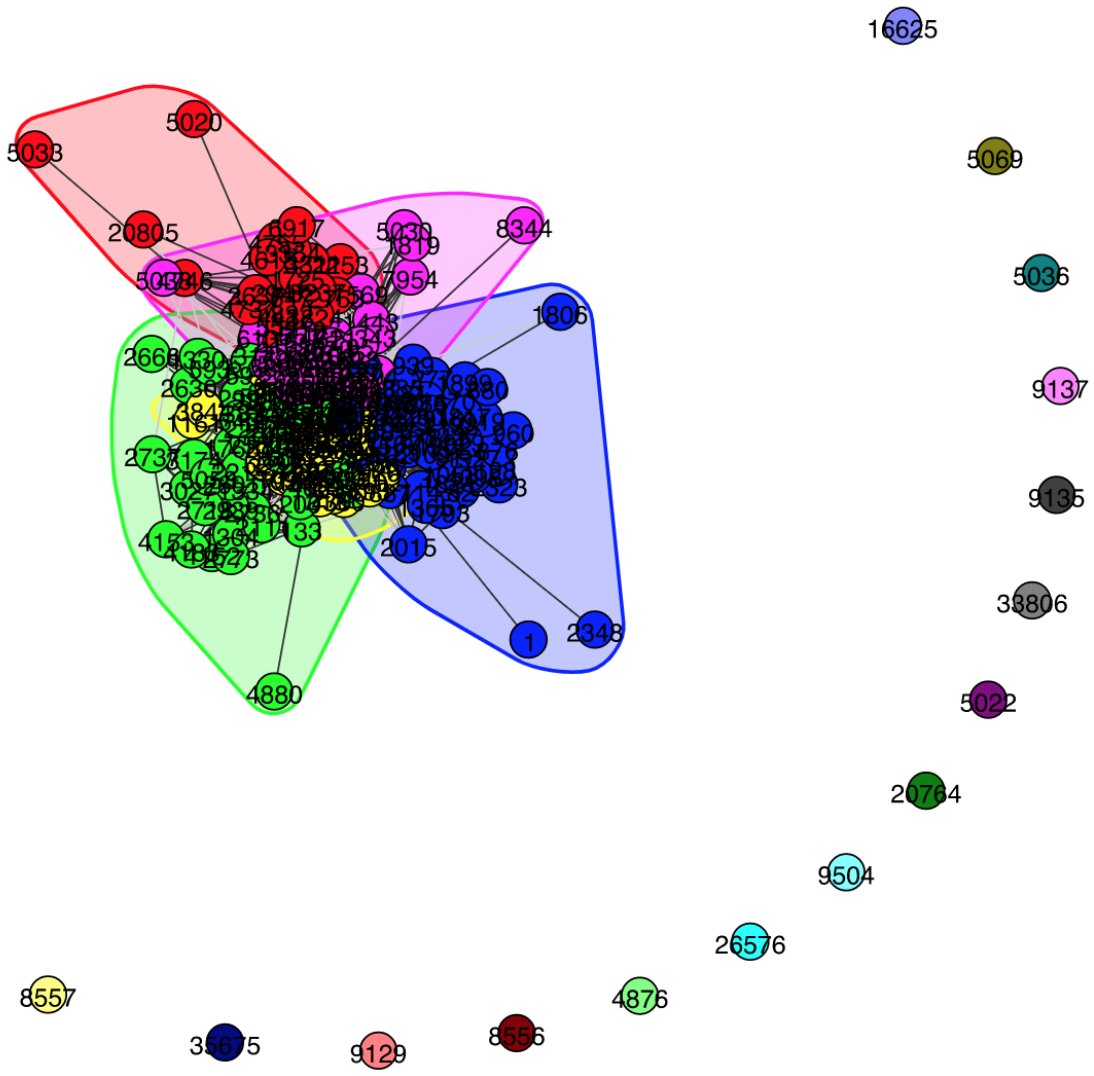

Figure 6: Communities in Top 250 Enron Emails Network According to Page Rank