Enron Network Analysis

INTRODUCTION TO NETWORK ANALYSIS Final Project (PASS)

Dulari Narayanar

2021

Contents

[**Project Summary** 2](#_Toc84801029)

[Problem Description 2](#_Toc84801030)

[Network Data 2](#_Toc84801031)

[**Analysis Summary** 3](#_Toc84801032)

[Data Analyzed 3](#_Toc84801033)

[Network Details 3](#_Toc84801034)

[**a)** **Network Structure** 3](#_Toc84801035)

[**b)** **Network Metrics and Analysis** 5](#_Toc84801036)

# **Project Summary**

## Problem Description

Enron was a U.S. energy trading company that became embroiled in a major accounting and financial scandal in 2001.  As per Wikipedia at the end of 2001, it was revealed that Enron's reported financial condition was sustained by an institutionalized, systematic, and creatively planned accounting fraud, known since as the Enron scandal. Enron has become synonymous of willful corporate fraud and corruption. The scandal also brought into question the accounting practices and activities of many corporations in the United States and was a factor in the enactment of the Sarbanes–Oxley Act of 2002. The scandal also affected the greater business world by causing the dissolution of the Arthur Andersen accounting firm, which had been Enron's main auditor for years.

The Enron case data has been taken in this project for studying further its network.

## Network Data

The adjacency list from the [Data list](https://www.dropbox.com/s/u7kc0wyad0n7z9v/enron-all.csv?dl=1) as well as the [Enron email corpus](https://www.dropbox.com/s/inlwq7czntel8j0/enron_mail_20150507.tar.gz?dl=1)  were used to build a network in Gephi and to analyze it.

# **Analysis Summary**

## Data Analyzed

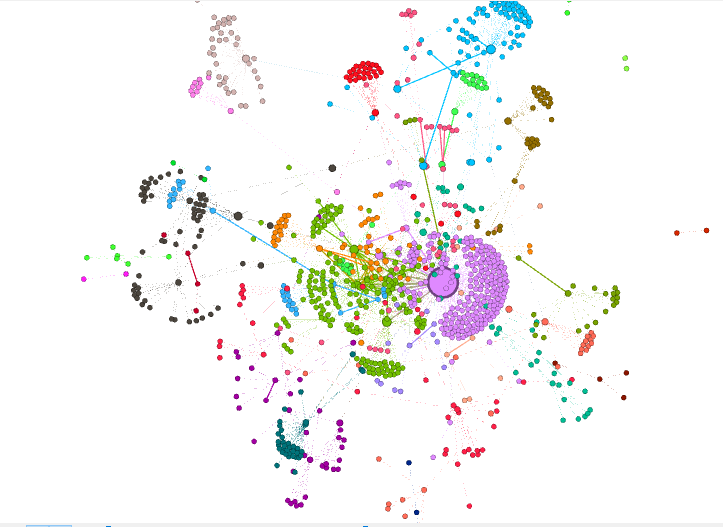
The data was imported as an adjacency list with undirected Graph type and with 1102 nodes and 1802 edges. Undirected was selected as the interest was more on the connection between two people in the company rather than who sent who the email. The Yifan Hu Graph layout was used.

The nodes were sized by betweenness centrality, and the color coding was by modularity class (by cluster)

## Network Details

### **Network Structure**

**Figure A: The Network Structure prior to Filtering**



The whole network consists of 3 to 4 main clusters in the center and has numerous fans towards the periphery of the network. Further analysis of the email exchanges suggest that these are broadcast emails sent by regulatory departments like Legal, Quality, regulatory and other Departments to address matters regarding certain policies.

The network looks like emails flow in one direction from the center outwards, with the well-connected nodes at the center.

**Figure B: The Network Structure based on a threshold of 9 and above emails**

Diagram

Description automatically generated

To Simplify the network and observe attributes and patterns the network was filtered to above 9 emails.

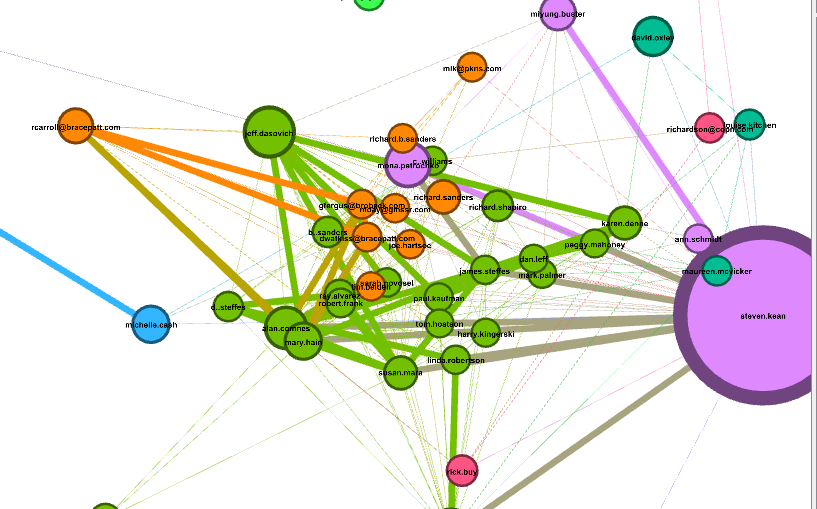
### **Network Metrics and Analysis**

Centrality

The centrality measure used was Betweenness centrality with Steve Kean showing the highest centrality. Steven Kean also was the center of one cluster (purple) and more details about the cluster is explained in the cluster section. This makes him the most important person in the network as he formed the main bridge between many other cluster centers as well.

Tie Strength

**Figure C: Tie Strength in the Network Structure based on a threshold of 9 and above emails**



The thickness of the lines shows the number of emails in which 2 people are connected. It was observed in the purple cluster, that Steven Kean and Miyung Buster have a strong tie strength. Analyzing the emails show that Miyung Buster although sent emails to many recipients who did not correspond with other Enron personnel, there were a significant number of emails between Steven and Miyung with regards to the California Power issue. The same goes for the tie strength between Mona Petrochko and Steven Kean, whose emails also correspond to the same California matter. Steven Kean also had strong tie strengths with some of the Enron government relations employees (green cluster) in which John Sheik, Susan Mara, Mary Hain, Alan Comnes, James Steffes and Richard Shapiro were always on the cc list on the same above matter.

Steven Kean also had very strong tie strengths between some of the nodes in the green cluster. A closer analysis of the emails of the green cluster nodes shows that all of them were related in some ways to government relations.

Thus, these were the strongest relationships in the Enron network during the time the email data were collected.

Clusters, Cliques and Communities

When the threshold was set to 9 emails and higher, 4 main communities (clusters) emerged centering around Steven Kean, Jeff Dasovich and R Carol and David Oxley. The purple cluster seem to main correspondents to Steven Kean on California matter while the green cluster represented Government relations staff and the orange cluster represented the top management of Enron.

Analyzing the email database online shows that Steven Kean has had 7452 email exchanges with many employees and the topic of discussion were on staff and human resources related, and thus was a top executive who did a lot of broadcast emails to employees or part of the email chain. Many of the emails though he is in the copy list rather than the sender or sent. A google search shows he was the Head of Enron's Chief of staff, which tallies why he had so many email exchanges.

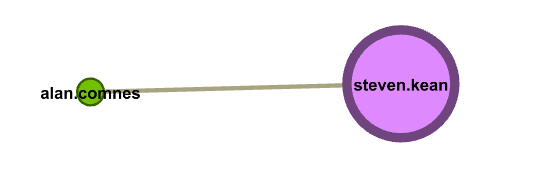
The email database also shows that Jeff Dasovich has had 18208 email exchanges with many employees and the topic of discussion were on regulations, and thus was a top executive who corresponded with the teams who dealt with policies. A google search shows he was a government relations executive for Enron.

In similar fashion both David Oxley and R Carroll (who is an external associate since he has a non-Enron email address) had multiple email exchanges. David Oxley email contents were Human resources related while R Carroll seemed to have been liaising with Enron’s litigation team mainly.

From the clusters it can be observed that most of the email content during the period this database was taken where on matters relating to staff, litigation, California scandal and government legislation and policy matters. This email database is thus mainly from the period when the trials were taking place.

No cliques could be observed in the network. There does not seem to be email connections between people in any one of the clusters, but the flow of email seems to be more going from higher management to mid management to employees.

Density

The highest density of emails was between Alan Comnes and Steven Kean and the email correspondence seemed to be on matters regarding regulatory affairs with over 35 emails. 

Egocentric Networks

It is interesting to notice that there were not any egocentric networks for the high centrality characters. The company may have been closing and email activity was winding, and thus mostly broadcast emails were being sent to employees or important email to only relevant staff.

It may also represent the culture of the company, whereby communications were one way as described earlier in a top to down manner.