

Introduction:

The Enron email dataset is a collection of emails that were exchanged within the Enron organization. For this assignment, we are analyzing a small subset of the same dataset.

While loading this dataset for the first time in Gephi, the network initially had 1102 nodes and 1802 edges. I applied a “Force Atlas” algorithm before applying any filters to the network. To abstract a more manageable network, I chose to filter the network using the “Giant Component” filter, filtering out the nodes with degree less than 2. This gives us the network of connected nodes that is a part of the largest cluster of the given network.

Then, I ran the modularity statistics to detect the communities in the extracted network. The nodes are sized by betweenness centrality and are color coded by modularity class. The following picture shows the community structure of the sub-network.

From the adjacent graph, we can observe six different communities that are color coded by red, brown, pink, blue, purple and green.

Upon careful analysis, I found that the red community comprises of a few executives of Enron and most Enron employees. People with bigger node size (in red) are Enron executives who connect other Enron employees to major executives of Enron. We can also see that there is an intense exchange of emails within the red community than in other community. The pink community comprises of CEO Kenneth Lay, COO Sally Beck, Vice President Danny McCarthy and other top Enron executives who were critical to the Enron organization, but were less involved in email exchanges. The brown community is composed of lawyers and attorneys who were not Enron employees,

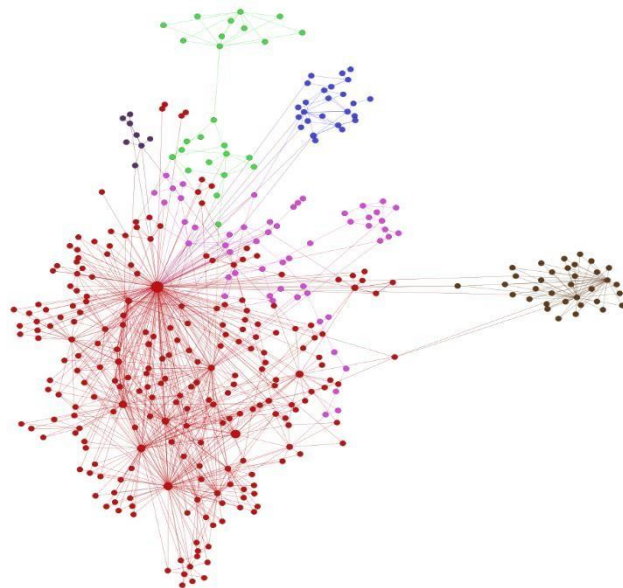


Fig 1: Enron Communities

but handled Enron financial legal issues. The blue community comprises of research teams and consultants that did research works for Enron. The purple community consists of a triads who served the role of financial officers and few executives of Enron who were Houston, Texas based. Finally the green community comprises of individuals who worked for a similar Enron like Oil and Energy based companies such as Calpine Power Company, Duke-Energy etc.

Relationship and Strength:

The keywords used in the email content, number of emails sent/received or exchanged between the nodes can be used to analyze the tie strength between the different nodes in the network.

Also, the email addresses can be used to differentiate the types of relationship shared between the nodes. For example, the tie strength between steven.kean@enron.com and mary.hain@enron.com will most likely be different than the tie strength between steven.kean@enron.com and

douglas@arterhadden.com. Two employees exchanging emails from the same organizational mailing address (*@enron.com) will have a higher degree of trust than two employees using different organizational email addresses. For example, one such emails encountered during the Enron email analysis frequently mentioned “Talking Points”. While I am guessing that the “Talking Points” must be some critical documents of Enron, it was shared only between the top level executives of Enron. On the same note, we can say that the Enron executives will never share these “Talking Point” documents to another employee with a mailing address of *@iepa.com or *@velaw.com.

Similarly, From, To, CC and BCC can be used to identify how close the relationship is between the nodes. A From-To relationship indicates a direct communication between two people while a From-CC-BCC relationship indicates a potentially weaker tie, where a person may simply be informed about a status and nothing else. For example, one email exchange between Jeff Dasovich and Richard Shapiro revealed that they had informal relationship rather than a formal professional relation. Most of the emails that were sent from Jeff Dasovich to other Enron employees were sent separately to Richard Shapiro (without any BCC or CC) that included his personal remark. One such email to Richard Shapiro said “Mike is fleshing out legal details”, while Mike is also an Enron employee. This personal remark tells us that Richard Shapiro and Jeff Dasovich must have shared a high level of trust as Jeff Dasovich is confiding his personal opinion to Richard Shapiro about another Enron employee.

In the red community, we see that the nodes are involved in email exchanges extensively. At a glance, we see that the brown and blue community share a stronger relationship as they are involved in email exchanges that flow within their communities.

In the green community, we see that sandi.j.thompson@calpx.com and cabaker@duke-energy.com share high degree of trust between each other. Both of these node act as a bridge to connect two clusters. Mr. Sandi_j connects the Calpx employees to the other power company employees via Cabaker.



Fig 2. Bridge in the Enron Network

Also, the number of emails exchanged can tell us the strength of relationship between two nodes. A quick search on Enron email corpus revealed that there were at least 67 emails directly exchanged between Morris Brassfield and Rod Hayslett, and 34 emails (at least) were sent to Morris Brassfield from Rod Hayslett (without including any other individual in the thread). Similarly, 95 emails (at least) were sent by James.Sanders to Rod Hayslett while 166 emails were sent by Rod Hayslett to James Sanders. Few

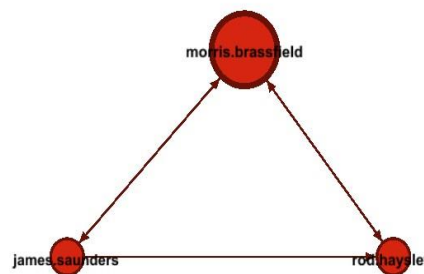


Fig 3. Triad

of the emails showed the loop interchange of emails

between these three nodes (they formed a triad). The loop exchange of emails indicate high degree of trust between the nodes. Similarly, Jeff Dasovich, Government Relations Executive and Richard Shapiro, Vice President of Regulatory Affairs alone share 582 sender/recipient email combinations in which they discuss the litigation issues (source: <https://roddersnotes.wordpress.com/2013/11/24/analyzing-keywords-in-enrons-email/>). The volume of mails shared exclusively between the two individuals reflect strong tie strength. Their tie strength can be addressed to the fact that both Richard Shapiro and Jeff Dasovich occupied similar position in the Enron (i.e Government affairs department). As such, they must have had similar interests in the outcome that relate to their position in the Enron structure. Hence, they must have had least distance between them both structurally and socially. Besides, there were a lot of personal remarks used in the email exchanges between them such as “he is trying to ensure a World War III” that hinted stronger tie between Shapiro and Dasovich.

Ego Networks:

An Ego-Network comprises of a focal node which is usually at the center and its corresponding nodes which are directly connected to the central node (also known as Alter). The figure below shows a level 2.0 Ego Network of Steven Kean.

The most important people in the network can be identified using the centrality measures. The nodes in our graph are characterized by betweenness centrality which also measures the node's or the ego's role as a connector/bridge between other nodes. It tells us how critical the node is to the network in its functioning as a bridging point between other nodes. Here, the size of each node correspond to its critical role to serve as a connector in the network. Table 1 shows top 5 individuals from the above network who scored the highest with respect to betweenness centrality.

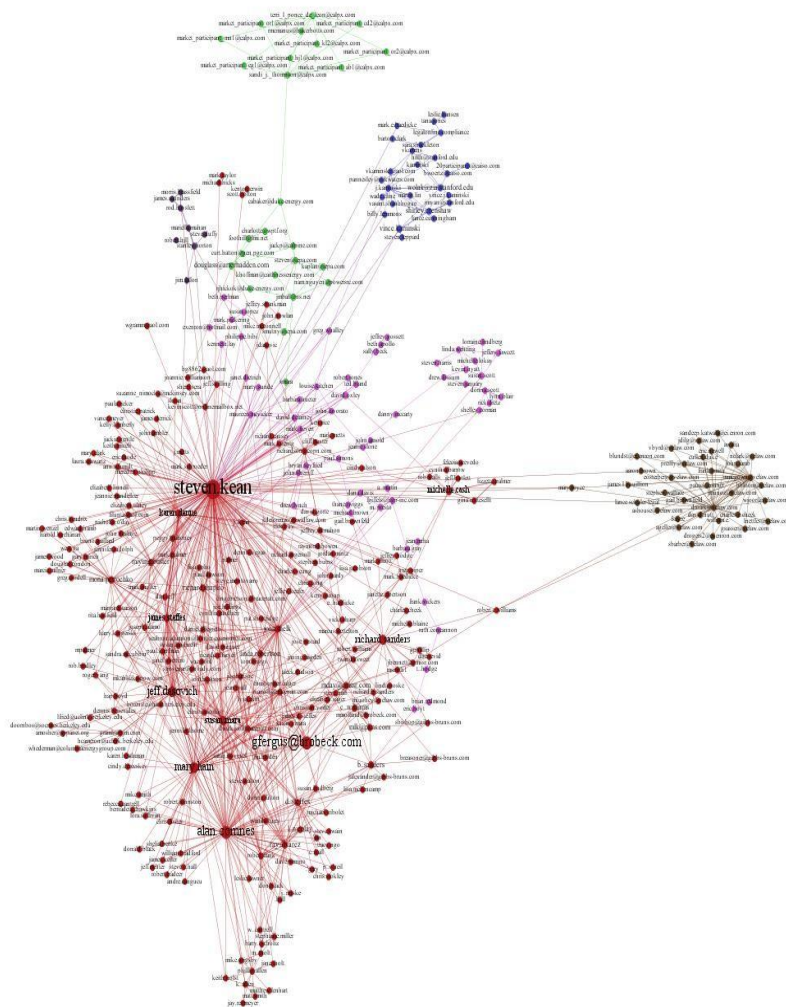


Fig 3: Network Visualization

As we can see, Steven.Kean is the node with the highest betweenness centrality, which means he is at the center of all important resources on the network. He has greater access to information and communicates his opinions to others most efficiently and most often. In the visualization (viz.) graph, we see that he has direct access to some important people of Enron like Kenneth Lay (CEO and Enron founder) and Jeff Skilling (CEO). He was also involved in recruiting employees at Enron. The database search results show that he connected most of the Enron department with each other as he forwarded most of the cross-department emails to other Enron employees. As per search results, he also sent some direct emails to Kenneth Lay regarding the Enron's work environment.

Label	Position
steven.kean	Vice President & Chief of Staff
gfergus@brobeck.com	Attorney
alan.comnes	Director Government and Regulatory Affairs
mary.hain	In House Lawyer
jeff.dasovich	Government Relation Executive

Table 1: Top 5 people per Betweenness centrality

The second node with the highest centrality is Attorney Gray Fergus. Although Gray Fergus (gfergus@brobeck.com) was not officially an Enron employee, but he was an influential member of the Enron network. Going through the Enron emails, we see that he was one of the primary authors of a third Enron memo that was designed to manipulate California's Energy market that led to the beginning of Enron scandal. From the viz. graph, we see that he has direct access to Enron executives like Richard Sanders (VP, Enron Wholesale Services) and Alan Comnes (Director Government and Regulatory Affairs). Going through the Enron email database, Gray Fergus has sent direct From-To emails to Richard Sanders without any CC/BCC hinting that they must have had certain level of trust with each other. Also, we can note that few pink and red nodes are closer. A closer look at the label of these nodes tell us that these nodes were the top faces of Enron scandal. Top Enron executives who were prosecuted in the scandal such as Kenneth Lay (pink), Susan Lopez (pink), Jeff Skilling (red) Jefferey Shank (red), Richard Causey (red), Mark. Koenig (red), Kenneth D. Rice (red) and Greg Whalley (blue) are close together. So we can say that although these nodes represent top executives but the ones in red were involved in more email exchanges.

Clique:

A clique in the given Enron network is the one in which each node is connected to every other node. The clique is useful in identifying the group of individuals who are fully interconnected to each other. In the context of Enron, a clique is a set of nodes that has either To or From emails, sent or received for other nodes within a given subset. In the given figure, we see that Steven.kean, james.steffes, jeff.dasovich, mary.hain, susan.mara, alan.comnes together

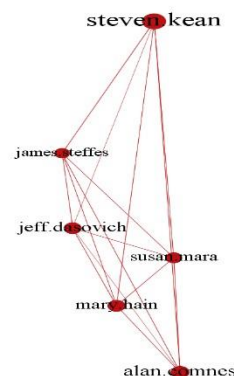


Fig 4. Clique in Enron Network

form a closed clique. They had frequent exchange of emails. This can be attributed to the fact that all of them were top executives of Enron and needed to exchange Enron-related information to one another.

Network Density: Density of a network can be defined as a total number of connections versus total number of possible connections. The Density value of a network ranges from 0 to 1, with 1 indicating a fully connected network. The network density for the given network is 0.008, which is quite low. The possible reason for this low value can be addressed to the presence of numerous unbalanced triads and few isolates.

References:

1. <http://www.igi-global.com/dictionary/ego-network/9203>
2. <https://rodgersnotes.wordpress.com/2013/11/24/analyzing-keywords-in-enrons-email/>
3. http://enrondata.org/assets/edo_enron-custodians-data.html
4. <http://james-irwin.github.io/>