

ENRON EMAIL LIST

The main purpose of this part of the project is to find all the email recipients and senders from the Enron data set, and summarize and graphically represent the results. The first task was to extract the header of the emails, and then analyze the information. This included looking for Bcc (Blind Carbon Copy) and Cc (Carbon Copy) fields.

Preliminary Investigations. This was accomplished using regular expressions, and we found that if we do not include Bcc and Cc, we see that the person who emailed anyone the most was Jeff Dasovich. If we include Bcc and Ccs, the list remains the same. We may now want to

Rank	Sender	Number of Emails
1	jeff.dasovich@enron.com	131302
2	veronica.espinoza@enron.com	110601
3	rhonda.denton@enron.com	90730
4	cheryl.johnson@enron.com	74632
5	jae.black@enron.com	59196

TABLE 1. The top 5 people who sent email from the Enron email list.

investigate who Jeff was emailing, and we at least two ways of doing that. The first is to build another frequency table, as follows: In which we can see the number of times Jeff emailed his

Rank	Sender	Recipient	Frequency
1	jeff.dasovich@enron.com	richard.shapiro@enron.com	2909
2	jeff.dasovich@enron.com	paul.kaufman@enron.com	2768
3	jeff.dasovich@enron.com	susan.mara@enron.com	2746
4	jeff.dasovich@enron.com	james.steffes@enron.com	2725
5	jeff.dasovich@enron.com	karen.denne@enron.com	2489

TABLE 2. The top five people that Jeff Dasovich emailed.

top 5 recipients. We may also look at a network graph;

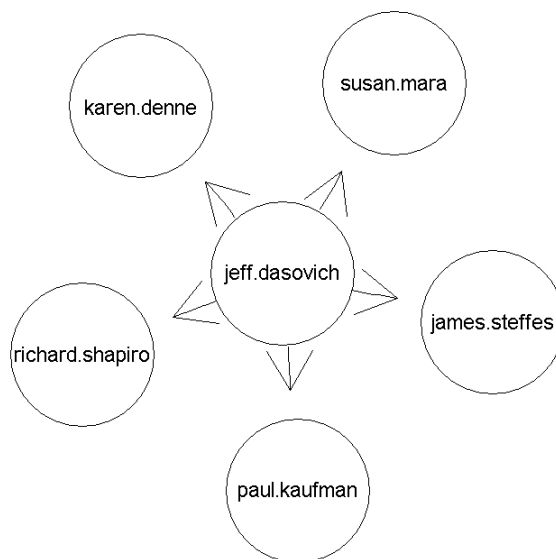


FIGURE 1. A small network of who Jeff emails

Which gives us the same information, but visually, and without information about the weights.

However, with network graphs it is easy to see who people Jeff emailed were communicating with. The following is a graph of the top three people Jeff was emailing, as well as the top three people his top three were emailing.

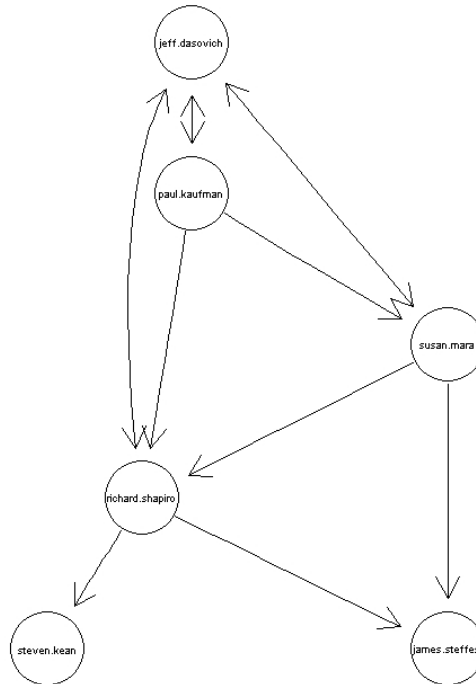


FIGURE 2. A small network of who Jeff emails, and who they email.

This shows that there is a small group, some of which email each other frequently.

0.1. Who is Bccing? We may be interested in who is Bccing people, and who they are Bccing. We find the following: It is interesting to see that Pete Davis Bcced a very large number of

Rank	People	Frequency
1	pete.davis@enron.com	205896
2	jeff.dasovich@enron.com	24000
3	vince.kaminski@enron.com	19170
4	kay.mann@enron.com	13536
5	kay.chapman@enron.com	13530

TABLE 3. The top five people that were Bccing.

people. In fact, if we look at all the emails grouped together (that is, disregarding if they were in the From, Bcc, or Cc field), we see that Pete sent out the most amount of emails as well.

Rank	Sender	Frequency
1	pete.davis@enron.com	215046
2	jeff.dasovich@enron.com	155302
3	veronica.espinosa@enron.com	111745

TABLE 4. The top three people that sent out any type of email.

Lets take a look at who Pete was emailing;

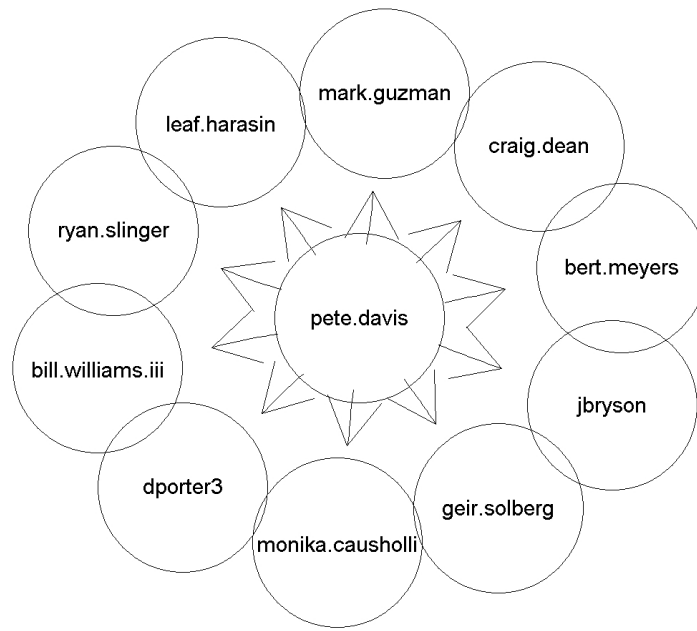


FIGURE 3. A network of who Pete emails.