# Predict 452 Section 55: Web and Network Analytics

# Assignment 3: Option 1: Social Network Analysis (SNA)

Jack Doig

## Management Problem

Attrition of employees, particularly in a highly skilled organization is a huge expense to the organization. Company-specific knowledge is lost which imposes a training cost, the subsequent recruitment costs can be expensive as a replacement is found, and there are less tangible impacts on company reputation and morale.

This study makes an assumption that one important predictor of employee attrition is the social network, with a social network defined as one-to-one communication between employees (as opposed to impersonal distribution groups). With a strong social network, the fabric of the organization is stronger and more resistant to attrition. There is value, then, in identifying the employees who create high numbers of edges in this network, and giving those employees close management attention, for losing them would severely damage this network and contribute to this predictor of attrition.

## Research design and methods

We can use the Enron email network to develop the tools we need to identify these individuals. Firstly, we can extract the whole set of one-to-one communications from the Postgres database, in order to reduce the data size, we can group on the sender and receiver.

In order to then isolate the most important actors in this network, we can create a network object using the R Networkx package, and apply a number of techniques:

a) The actors with the highest *distinct* Degree measures would have the most distinct one-to-one connections and therefore be at least loosely connected with the largest number of people

b) The actors with the highest *total* degree in the one-to-one network would have the strongest individual social connections, although possibly with a limited number of people

c) Density of the one-to-one network gives a general measure of the strength of the total company social network. Alone, this is not useful, however useful to understand over time series.

d) Closeness Centrality, Betweenness Centrality and Eigenvector Centrality: these measures may indicate people who might be in a position to enhance the one-to-one social network easily, and therefore are important actors.

## Results/Findings

Tana Jones ("Executive"), with 432 distinct one-to-one connections, is the most distinctly connected actor in the network, followed by Vince Kaminski (MD, Research), with 412. Note the Average is 10.17. See the attached histogram for the distribution.

Vince Kaminski also has the highest number of total one-to-one emails (13,742) followed by Kay Mann (13,692) and Pete Davis (9,148). Surprisingly, the average is 54.16.

The density of the network over time show a degrading density over time (see graph) from 1999 to 2001. This might be normal with an expanding organization. It would be good to compare with other organizations, and work on a more granular scale.

Betweenness Centrality reveals Mary Hain (an in-house Lawyer) as the strongest. The Closeness winner Sara Shackleton (note: named in a paper as a "middle man") and finally the strongest Eigenvector centrality belongs to Tana Jones.

## Next Steps/Recommendations

Assuming agreement with the assumption stated at the outset of this study, the actors identified above are vital links in the social fabric of the firm. While retention of employees at the particular firm in this study is not realistic, these methods could be applied to any firm in order to identify the most important 'social' actors within it. Also studying the density of one-to-one email traffic over time can indicate how this particular predictor of attrition is evolving.