Text Mining of Social Networks
Detecting patterns from unstructured data

THE AYASDI IMPACT

Ayasdi's Text Analysis Solution was able to quickly analyze the entire Enron email data set consisting of over 20,000+ individuals and over 500,000 emails. Using the analyses, our data scientists were able to identify 205 key individuals that were likely to have relevant information regarding Enron fraud. By understanding the flow of information through time, individuals, and subgroups, data scientists were able to reconstruct the events leading to Enron's collapse.

THE PROBLEM

The Enron Scandal in 2001 was one of the most complex bankruptcy cases in U.S. history. The question which investigators faced was: "How could Enron and its top accounting firm, Arthur Andersen, swindle the global financial system?"

The challenge was to efficiently interrogate the email network with over 500,000 emails that over 20,000 individuals had written over a period of years to determine the scope of the Enron accounting fraud.

THE SOLUTION

Ayasdi conducted an analysis of all email users as well as the text content of their emails over time. The image below shows this analysis.

Each individual sub group consists of email senders that have similar email content. The circled subgroup represents senders using the words "bankruptcy" and "crisis" more often than other email senders (red color indicated high amount of sent emails). Interestingly, the "bankruptcy" group has yet another subset of individuals who sent a higher number of emails to external accounts. Drilling further into this subgroup, Ayasdi distinguished 205 individuals who sent a high number of external emails with terms such as "discuss," "schedule," "product," and "risk."

Bankruptcy Group Sending High Volumes of External Emails
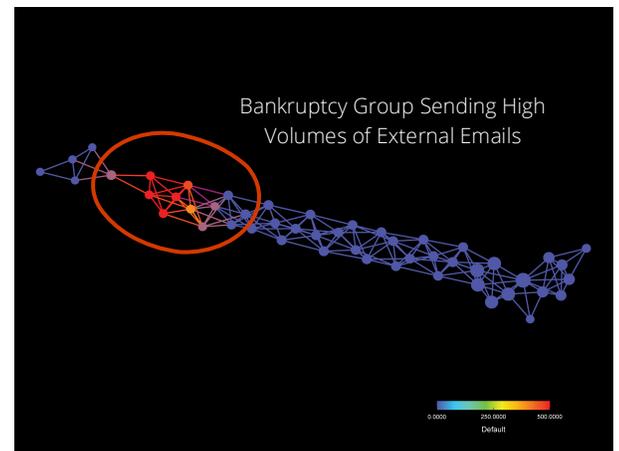
After a quick investigation, the email analysis leads to Vince Kaminski who repeatedly raised strong objections to the financial practices of Enron's CEO, designed to fraudulently conceal the company's burgeoning debt.