Memorandum

Date: October 24, 2022

To: U.S. Securities and Exchange Commission (SEC)

From: Abigail McDonald

Subject: Analysis of the Enron Email Corpus

The purpose of this memo is to analyze relationships and social structure found within the Enron email corpus. Various Enron executives are accused of hiding failed investments using shady accounting practices, artificially inflating stock and insider trading. As a business social network can indicate key relationships and players, this project aims to analyze the connections within the Enron email corpus.

A social network was created by working off a list of emailers and their recipients taken from a sample of the Enron email corpus (see Figure 1). Many gatekeeper nodes characterize this network, that is, people who only have contact with one person from the network. Many of these are peripheral clusters and indicate someone in the organization sending information to those much lower in the hierarchical company structure or, more often, outside the organization (Enron). Gatekeeper nodes (persons) are labeled in blue, and those with only a single email from/to (usually from) are labeled in black. The network structure is hierarchical; emails flow from the central cluster to the periphery and rarely go in the other direction. The exception is inside the central clusters, where emails flow in a bi-directional format. Three separate clusters make up the central cluster in the colors purple, light blue, and green (see Figure 2).

The central network nodes appear to be at the top of the hierarchy. Many names in this group can be identified directly from the Enron Organization chart. Figure 3 shows how these nodes are tightly connected regarding information dissemination. Steven Kean (purple cluster), chief of staff for Ken Lay, according to the organization chart for Enron, is the closest to all other nodes (high betweenness centrality). This high betweeness centrality is likely because of his role in the organization, disseminating information. Most of Kean's email traffic is back and forth between higher-level executives. See Figure 3 for the people with whom he most often communicates. Next in direct email traffic size is Gary S. Fergus (green cluster), a lawyer for the outside firm Brobeck, Phleger, & Harrison LLP. None of his emails appear personal, and most of his back-and-forth communication is with Richard Sanders and Alan Comnes. The third node by connectedness is Alan Comnes (green cluster). From his email, it appears that Comnes deals with government regulations for Enron. His email is professional (no appearances of spam or lunch deals), and much of it is back and forth with people outside of the company.

The following group is still centrally located but smaller in connection size. First is Mary Hain (green cluster), a lawyer for Enron out of the Portland office. She appears to be a prolific emailer with the rest of the legal team. Hain seems to be involved in many of the deals and trades for Enron from a legal standpoint. She has little to no personal email, and what little she does

concerns her leaving the company in April 2001 for a new position. Next is Richard Sanders (green cluster), a high-ranking member of the legal team who remained with Enron until its complete bankruptcy collapse. He is very fond of one-line responses to others' emails. His correspondence is primarily professional and with other members of the internal legal team and counsel outside of Enron. Jeff Dasovich (light blue cluster) has the most significant number of emails in the collection and appears to be a government relations executive for Enron. He also has a generous amount of personal email in the corpus. Finally, the last of the major players in the center clusters is Susan Mara (light blue cluster). She was the Enron director for regulatory affairs and appeared to remain with the company until 2002. Her emails are primarily group emails to/from the government relations and legal sections of Enron to people like Richard Sanders and Jeff Dasovich concerning energy regulatory concerns.

The final group of well-connected nodes connects the center nodes to the outer clusters. Most of this group are gatekeeper nodes, passing on messages to an extensive network of underlings or outside parties with vested interests in Enron. For example, most of John Shelk's (light blue cluster) messages appear to be political regarding congressional matters in Washington, D.C. It can be inferred from the email that Shelk is a lobbyist working for Enron (he has an internal email). His messages reach out across a wide swath of internal and external emails, but most internal messages are directed to individuals in the government and legal affairs sections. Next, Miyung Buster (teal cluster) appears to be a media or congressional liaison. He often emails articles to long lists of otherwise unconnected individuals (they are not emailing others in the corpus), making him a gatekeeper node. Following this, Vince J Kaminski (orange cluster) has two Enron email addresses and even emails himself back and forth between them. Kaminski is a lead analyst and the gatekeeper for a team of analysts otherwise not connected to the network. Kaminski appears to run a Stanford alumni club and is heavily involved in college recruiting for internship/associate positions. Then, Michelle Cash (gray cluster) is an employment lawyer responsible for hiring management resources. From her emails, she appears to handle the employment contracts and HR legal policies. She has a great deal of personal email mixed in with the business. Finally, Frank Wolak (orange cluster), an economics professor at Stanford University, is the California ISO cluster's gatekeeper and a member of Vince Kaminiski's Stanford Alumni/recruitment club.

Each colored cluster appears to contain specific work groups, which is most apparent in the many gatekeeper nodes with a fan cluster or even a set of consecutive clusters (dandelion appearance of the outside clusters). First, the central purple cluster primarily comprises Steven Kean's network. Kean has an extensive dissemination network; thus, this cluster is globally influential but not internally as influential (much of the network is terminal nodes). The purple cluster appears to be primarily connected to the green cluster. The green cluster mainly comprises Enron's legal team and external legal counsel. This cluster includes corporate lawyers Alan Comnes and Mary Haines and external attorney Gary S. Fergus. Rounding out the internal clusters, light blue is composed of government liaisons and regulatory affairs executives (Dasovich, Mara, Shelk, etc.). The teal cluster runs through the central network but has a terminal end with an unconnected team. This cluster appears to be composed of directors and commodities traders. Next, the orange cluster comprises Vince Kaminski's Stanford alumni club,

many of whom do not directly work for Enron. Attached to the orange cluster by way of Frank Wolak is a cluster composed of employees of California ISO (which runs California's power system). This cluster is relevant to the investigation of the artificial pricing scheme. After that, the pink cluster comprises small energy companies related to Enron and the marketing department. Finally, the gray clusters compose all other departments and external companies communicating through specific liaisons.

The central clusters are the densest and, when filtering all node connections of degree 20 (connected to 20 or less nodes), we are left with the most important nodes of this network (see Figure 3). Of these nodes, Steven Kean has the highest degree and highest closeness centrality (he has the shortest path to the bulk of the network). Kean, John Shelk, Susan Mara, Alan Comnes, Mary Hain and Jeff Dasovich have the highest eigenvector centrality, in other words, they send the most emails to others who are also highly connected. The legal (green cluster) is the densest and has the most clique like shape. This group is well connected to each other and sends many emails back and forth, however, their emails contain far less personal connections (dinners, discussions about family members, etc.).

Relationships in the email data are apparent in both the content and number of interactions. For example, Jeff Dasovich sent the most emails overall and, while most of them are work related, he does send many personal emails to his family and close friends outside of the company. He has the strongest personal email correspondence (tie strength) with Susan Mara (792 emails) and does occasionally share thoughts and opinions about coworkers and outside contacts with her. He does this far less frequently with John Shelk (47 emails) who is also in his department. Steven Kean and Jeff Dasovich also have a strong tie with over 700 emails, however, most of this are logistic or formal in nature. Cluster, or department, however, does not fully indicate tie strength as Alan Comnes and Gary Fergus have almost no direct contact. Miyung and Steffes have a smaller amount of correspondence (72 emails), but their emails imply that they work out of close offices and talk in person during the day, thus their tie strength is greater due to the personal connection (content rather than number).

In conclusion, the most important ideas from the Enron email corpus are that the highest-ranking individuals in the hierarchy (e.g. the CEO) are not necessarily the most well connected or prolific in emails sent. The corpus is complex, and the relationship (tie) strength is indicated in both the content and the amount of emails sent. Additionally, clusters within the email indicate relationships between departments and centrality of those disseminating the most information (Kean is a good example). As for clusters, regulatory affairs and the legal team seem to be the most central departments and affect the flow of information.

FIGURES AND CHARTS

Figure 1 *Social Network of Enron Emails (sample)*

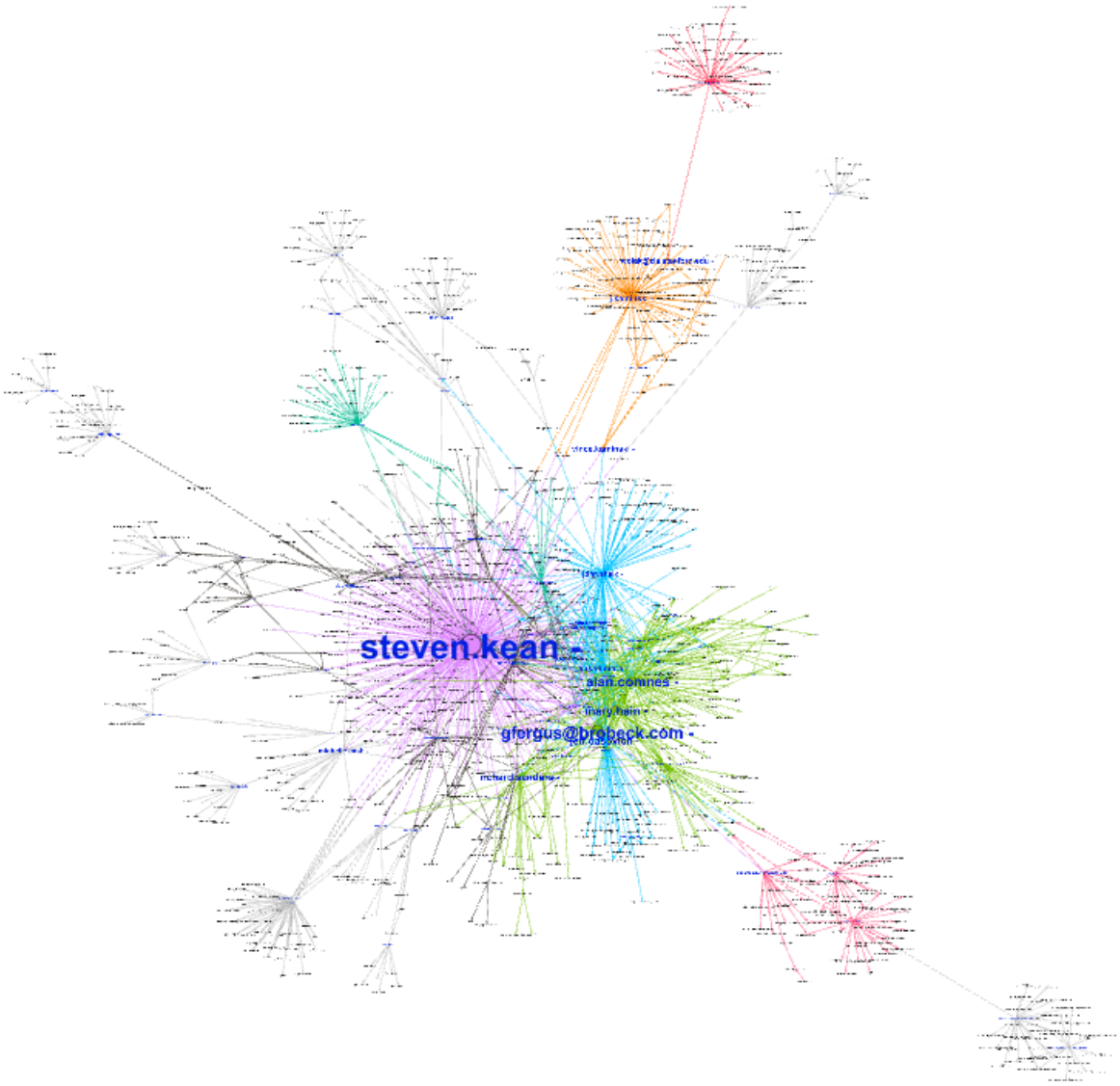

Figure 2 *Central clusters of sample Enron email network*

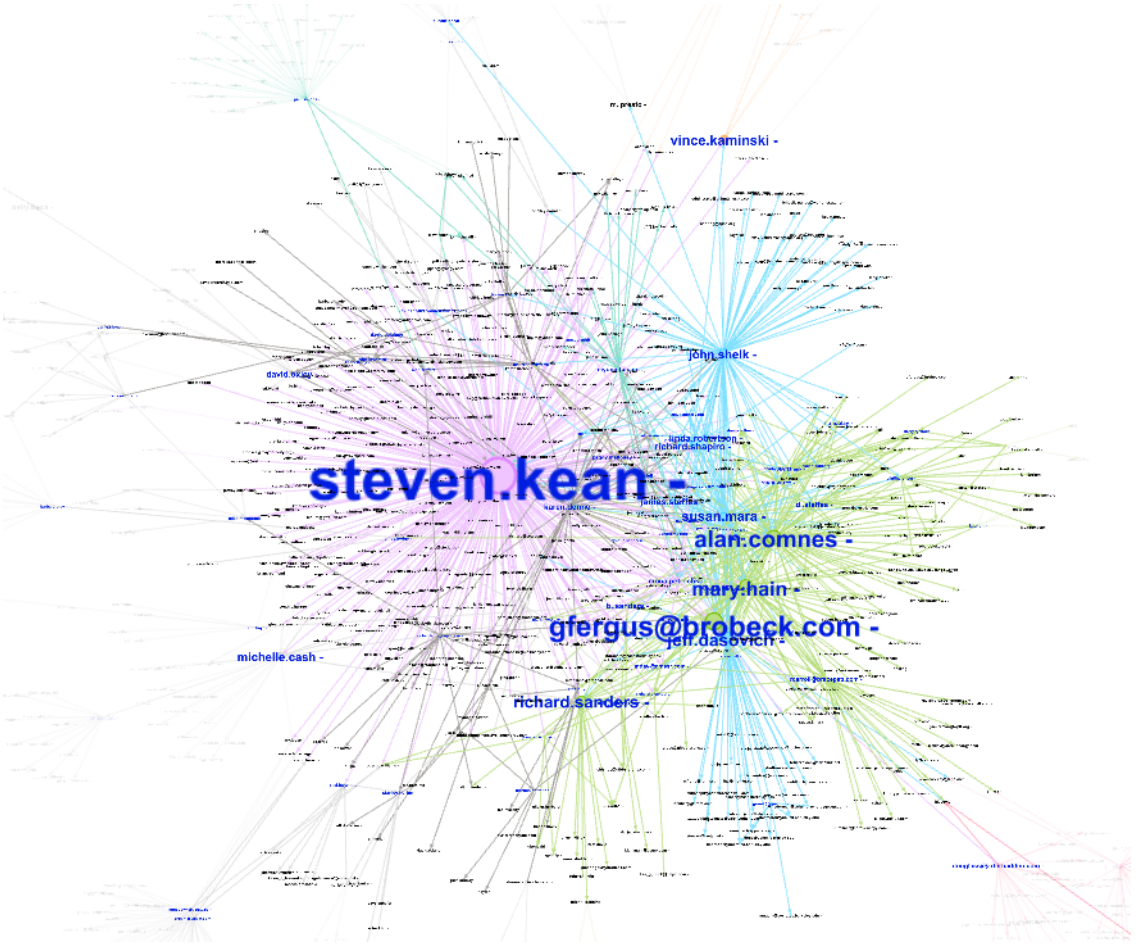

Figure 3 *Nodes that are the most connected*