Collective Intelligence at Enron during the California Energy Crisis

uncovering Collaborative Innovation Networks using Social Network Analysis

by

Ilias Hamdouch

MBA, Instituto de Empresa Business School, Madrid, Spain, 2006
BS, Texas Tech University, Lubbock, TX, USA 1997

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the
Degree of Master of Science in Engineering and Management
at
Massachusetts Institute of Technology
May 2015
[June 2015]

Author:
Signature redacted
....
Ilias Hamdouch
System Design and Management Program
May 2015

Certified by:
Signature redacted
....
Dr. Peter Gloor
Thesis Supervisor
Research Scientist at the Center for Collective Intelligence
MIT Sloan School of Management

Accepted by:
Signature redacted
....
Patrick Hale
Director, System Design and Management Program
Senior Lecturer, Engineering Systems Division

THIS PAGE IS INTENTIONALLY LEFT BLANK

Collective Intelligence at Enron during the California Energy Crisis

By

Ilias Hamdouch

Submitted to the System Design and Management Program in May 2015 in Partial fulfillment of the Requirements for the Degree of Master of Science in Engineering and Management.

Abstract

As interaction takes place between individuals, relationships are formed and collaboration and innovation emerge. In this thesis I have applied Coolfarming (Gloor, 2011b), a social network analysis method using Condor, a software tool to quantify communication patterns based on various data sources. I analyzed the Enron email archive to see if communication patterns of convicted employees differ from ordinary ones. Toward that goal, I compared the dynamic semantic social network metrics of 17 Enron employees convicted in the criminal trial following Enron's implosion with a control group of ordinary employees. I focused on 17 mailboxes of 24 Enron executives that were convicted.

Identifying criminals based on email behaviors is possible depending on the sampling strategy. When sampling based on employees with comparable total emails, the statistical analysis of the Contribution Index (Ci) metric revealed that criminals were less active. When sampling based on employees with comparable total influence, the statistical analysis of Betweenness Centrality Oscillation (Bco) and Degree Centrality (Bc) metrics revealed that criminals were less connected to others and less creative.

Key Terms: Collaborative Innovation Networks (COINs), Collective Intelligence, Coolfarming, Social Network Analysis, Condor, California Energy Crisis, Enron, E-Mail

Thesis Supervisor: Dr. Peter Gloor

Title: Research Scientist at the Center for Collective Intelligence, MIT Sloan School of Management

THIS PAGE IS INTENTIONALLY LEFT BLANK

Dedicated to

My wife, Zineb and daughter Lilia

My grandparents for their inspiring life journey that shaped my values.

My parents for always believing in me and sacrificing so much for me.

My parents-in-law for being supportive in challenging times.

My family and friends for keeping me going and making it a fun journey.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor Dr. Peter Gloor for his guidance and for this introduction to the fascinating world of swarm creativity and network science. In leading by example, Dr. Gloor was available in person, over the telephone and over email to promptly field questions and to brainstorm ideas at various stages. Dr. Gloor showed immense understanding and patience for my thesis during difficult times especially when I was vacillating between topics early on or data exploration paths later on.

I wish to also thank SDM Executive Director, Patrick Hale for giving me the opportunity to be part of the program. Patrick Hale was also understanding when considering class substitution of academic requirements toward my fields of interest. It was a unique experience to learn from the most exceptional and the most humble teachers and students in a world class environment.

I wish to thank my wife Zineb and daughter Lilia for their love and their support from a distance and for keeping my spirits up.

Finally, I would like to thank my parents, my parents-in-law, my family and friends for their love, prayers and support through this journey.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Table of Contents

THIS PAGE IS INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE IS INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE IS INTENTIONALLY LEFT BLANK

Chapter 1: Introduction

Motivation for the Thesis

As part of the System Design and Management curriculum, I wanted to study collective intelligence. This tends to emerge from the collaborative innovation and the competitive efforts of many individuals particularly in consensus-based decision-making. The concept is used in sociology, biology, political science, mass communication and crowdsourcing applications.

The most comprehensive definition I have found was that of Pierre Levy who defined it as "a form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills" (Levy, 1987). I'll add the following indispensable characteristic to this definition: "The basis and goal of collective intelligence is mutual recognition and enrichment of individuals rather than the cult of fetishized or hypostatized communities." Levy sums it up as the capacity to enhance the collective knowledge by using information and communication technologies to simultaneously expand the extent of human interactions. Open source software such as the Linux kernel is one of the most powerful examples of collective intelligence I can think of. Linus Torvalds has been the Linux project's coordinator and chief architect since 1991. By 2006, Linus' coding contributions represented merely 2% of this publicly-distributed software development initiative. By the end of 2014, Linux has become a popular platform in both the server and the supercomputer environments where it commands respectively 35.9% (W3Techs, 2014) and 97% (Top500, 2014).

In 2013, I began reading up on collective intelligence research. By the fall of 2014, I attended a course offered by Dr. Peter A Gloor at the Sloan School of Management. The Collaborative Innovation Networks (COINs) course introduced me to the fundamentals of social network analysis and to hands-on practice using Condor, a software designed to analyze large datasets and to uncover cool collective ideas inside and outside the bounds of an organization.

As part of the COINs course, I worked on a project to analyze the communication patterns inside a global organization, Enron Corporation based on the publicly-available email archive (Klimt, 2004). The objective of this analysis was to consider alternate methods to identify the leaders of suspicious activities and to propose a complementary way to detect illicit activities.

Following this course, I asked Dr. Peter Gloor to become my thesis advisor. He kindly accepted.

It is important to take a step back from collective intelligence to define the broader field of network science, to highlight its evolution and to relate it back to the areas of social and dynamic network analysis.

Network Science

The United States National Research Council defines network science as "the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena." Network science is a multidisciplinary academic field which studies interaction complexity of human and machine networks at a large scale including telecommunication networks, computational networks, biological networks, cognitive and social networks (Wikipedia, 2015).

The field endeavors to understand complex interactions by drawing theories and methods across scientific branches including graph theory from mathematics, statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, and social structure from sociology. This discipline has emerged with the evolution of various disciplines that sought to analyze complex informational relationships.

In 1736, Leonhard Euler published Seven Bridges of Konigsberg, a pioneering work that laid out the foundations of graph theory (Euler, 1736), a branch of mathematics that studies the properties of network structures. In 1735, the Seven Bridges of Königsberg was an intricate mathematical problem that imagined the notion of topological representation. The City of Königsberg in Prussia was built around the Pregel River and was made of two islands which were connected to each other and the mainland by seven bridges. One of the challenges that was considered was to find a way to walk through the city and cross each bridge once and only once. Euler redefined the problem by imagining the topological representation of the land masses (nodes) and the bridges (edges) connected to them. Euler argued that the condition of traversing each bridge only once required that the graph be connected with zero or two nodes of odd degree. However, the graphical representation of the City of Königsberg had four nodes of odd degree. Euler concluded that this problem had no solution. German mathematican Carl Hierholzer would later prove this conclusion. In the history of mathematics, Euler's conclusion is considered to be the initial theory of networks. Moreover, the difference between the actual layout and the graphical representation was a precursor in the development of topology. The main idea was not about the objects and their shapes but about the network structure.

Figure 1 – Euler's abstraction of the City of Königsberg (Story of Mathematics)

The field of network science continued to evolve slowly with applications in the natural sciences particularly chemistry (Sylvester, 1878) and psychology (Moreno, 1930) and (Northway, 1940). In Networks of Scientific Papers (Price, 1965), de Solla Price endeavored to demystify the social dynamic in the scientific research community by analyzing the pattern of bibliographic references. To do this, de Solla Price mapped a network of links of each published paper to the other papers directly associated with it. In a given field in a single year where we assumed to have 100 historical papers, the 7 new papers in the field are linked to 10 papers in the field that are cited more than once and connected to 10 miscellaneous papers outside the field. Each group of new papers was closely tied to a small portion of scientific literature but connected in a weak and a random way to a much larger group.

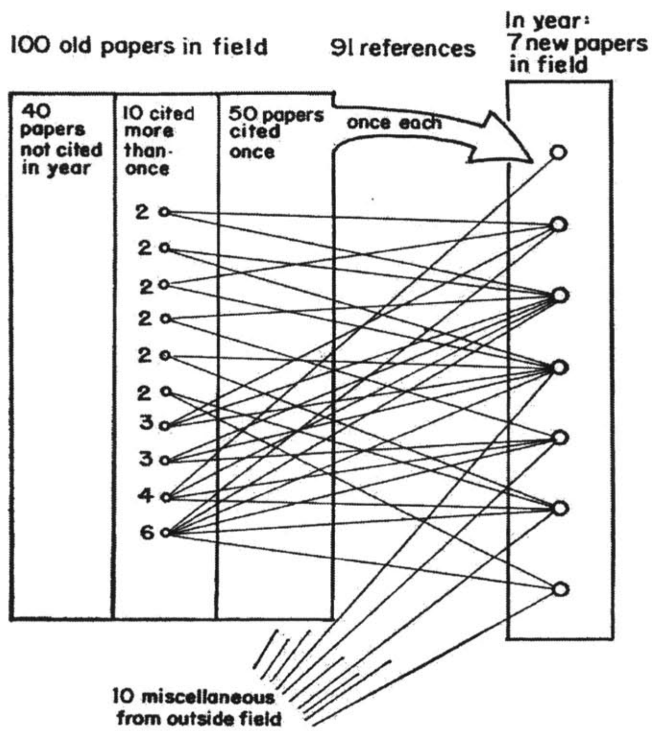

Figure 2 De Solla's Networks of Scientific Papers (1965)

This small part that is knitted together by the year of publication can also be referred to as the active research front. De Solla Price found that most papers through citations are linked closely together. Rather than a single row of knitting, the total research front of science was distributed network that I referred to as stitches and strips. From this study of citations, De Solla Price argued that these stitches and strips corresponds to the work of a 100 scientists and represented defined subjects whose descriptions may evolve year on year. De Solla Price concluded that the universe of 35000 journals was merely background noise and the knitted strips represented the central material like a piece of clothes on a body of science. More recently, network science research has endeavored to demonstrate mathematically different network topologies.

By the late 1990s, Duncan Watts demonstrated the existence of small-world networks (Watts, 1998) in which most nodes were not connected to one another, but most nodes could be connected to every other by small number of intermediate connections. In the context of a social network, this described the phenomenon of strangers being connected by a mutual contact following a six-degree separation pattern. It is also referred to in everyday languages as an affinity group in which people share a common interest or goal. This small-world network theory found direct applications in sociology, in earth sciences and in brain cognitive sciences.

A year later, Barabási and Albert mapped the topology of a portion of the World Wide Web to highlight the existence of what they called a scale-free network (Barabási, 1999). In this research, the phenomena of preferential attachment was demonstrated mathematically. Some nodes known as "hubs" had more connections than others and the network as a whole had a power-law distribution of the number of links connected to a node. This research built on Derek de Solla Price's network science work on the social dynamic of scientific citations. In 1965, Price had explained mathematically the occurrence of a Pareto distribution or power law when studying the "cumulative advantage" mechanism that existed in the scientific citations among published researchers.

Although many network topologies can be described with a Pareto distribution, a new concept best described by a frequency distribution (Mandelbrot, 1953) has reemerged in the past decade. Building on the groundwork of Mandelbrot, many researchers were able to map empirical data to the frequency distribution including applications in economic models of online business (Brynjolfsson, 2000), user-driven innovation (Von Hippel, 2005), micro-finance, viral marketing and social network mechanisms such as peer-to-peer and crowdsourcing phenomena.

By the late 2000, Erik Brynjolfsson and his colleagues researched the evolving distribution phenomena (Brynjolfsson, 2011). This research scientist found that in the Internet channel the decreasing costs of product information discovery could substantially increase the collective share of hard-to-find products, thereby creating a longer tail in the distribution of sales. This was counter-intuitive as the retail world was used to the 80/20 rule that described product sales in the traditional catalog channel. Brynjolfsson and his colleagues found that niche books accounted for approximately 37% of Amazon's sales and that ratio increased fivefold over the course of the period of 2000 to 2008. In their research methodology, it was identified that widely used power laws are a good first approximation for the rank-sales relationship. In fact, the slope may not be constant for all book ranks. The slope became steeper for the niche books.

With the advent of the Internet, network science theory evolved quickly in recent decades both in defining its various topological structure and in analyzing complex informational relationships across disciplines. With the Web 2.0, a branch of Network Science known as Social Network Analysis became increasingly popular to describe human interaction across online channels.

Social Network Analysis

Social Network Analysis (SNA) theory dates back to the early part of the 20th century when sociologists such as Georg Simmel and Emile Durkheim wrote about the importance of researching and understanding the relationship patterns between social actors. In the development of social network analysis (Freeman, 2004), the author argued that basic analytical methods were introduced by Jacob Moreno and Helen Jennings in the 1930s. However, it would be until the mid-1950s that the term would be used systematically to designate patterns of ties and to discuss common concepts in the general public and among social scientists. Some of these included a bounded group such as tribes and families. Others represented a social category based on gender, ethnicity and such. Over the recent decades many researchers have expanded its applications across academic disciplines ranging from economics to crime. Email interactions will be the relationship between individuals that we will be considering in this thesis.

SNA was traditionally mapped out using surveys and questionnaires directed towards individuals to determine their relationships with other individuals (Garton, 1997). The author and her colleagues highlighted the limitations of such an approach that can be biased given that one needs to rely on recollecting the participating individuals. Moreover, surveys and questionnaires relied on individual recalling their interactions and their behaviors. Online communication archives provided more precise representations of interactions and relationships between actors (Wellman, 2001). The author and his colleagues studied "computer-supported social networks (CSSNs) to understand human interactions and relationships through computer mediated communications such as email. Wellman et al. concluded that relationships formed in CSSNs go beyond real-world interactions by showing unique features specific to this communication context. This makes SNA of computer mediated communications complementary to traditional SNA. More importantly, computer mediated communications could reveal a lot about how human interaction and collaboration is shifting with the ubiquitous use of a variety of computing devices.

One of the most comprehensive SNA reference is Social Network Analysis – Methods and Applications (Wasserman, 1994) which defined the field as the study of relationships between individuals and groups as opposed to the study of the individuals themselves. The latter was prevalent in traditional social sciences. The structures of such relationships and the ways in which they evolved over time have been the essence of SNA. One important finding in the past decade was a Christmas card experiment (Dubar, 2002) that led to the conclusion that our neocortex can handle a maximum network size of 150 active contacts in our individual networks.

For those new to SNA, I will summarize some of the key structural metrics that will be used in the data analysis section of this thesis. The definitions provided here come from that SNA reference book (Wasserman, 1994).

Actors: These are the social entities such as individuals, organizations or companies that form the part of the social network to be analyzed. The other part is the relationships between these actors.

Relation: This is a set of ties between actors. In an interactive communication context like email, this a relation between two actors that is formed when one of them sends an email to the other.

Centrality and Prestige: This is one of the key measure in SNA to determine which actors are important and which are not based on relationships and the network structure. Wasserman and Faust use the following definitions:

- Prominence describes the importance of an actor.
- Centrality measures prominence using non-directional relationships.
- Prestige measures prominence using directional relationships.

These two metrics have different meaning when evaluating prominence. Consider looking at President Barack Obama's connections on Facebook. You can assume to a large extent that people connected to Obama are prominent. Now consider the difference between a situation where someone sent Obama an invitation and a situation where Obama sends someone an invitation. A measure of prominence based on simply being connected to Obama would be centrality. A measure based on who invited whom would be prestige. In this thesis, I will only be looking at various measures of centrality.

Degree Centrality: This is a local measure of direct relationships of an actor.

Betweenness Centrality: This is a more global measure compared to degree centrality. In communication networks such as email, betweenness centrality is a measure of an actor's ability to influence how and what information flows in the network by virtue of her/his position. This metric considers the shortest paths connecting every pair of actors in the social network. Mathematically speaking, betweenness centrality is the sum of the probabilities that the actor happens to be on the shortest path connecting any pair of actors other than himself. In general, actors have high betweenness centralities if they have relationships with two or more groups of actors who otherwise would have been disconnected from each other.

Throughout this thesis, I will be assessing degree and betweenness centrality as measures of prominence in the email communication networks. In addition, I will also be considering some of the temporal and the content-related metrics such as contribution index and total influence. I will discuss these further in the research method section of the Analysis chapter.

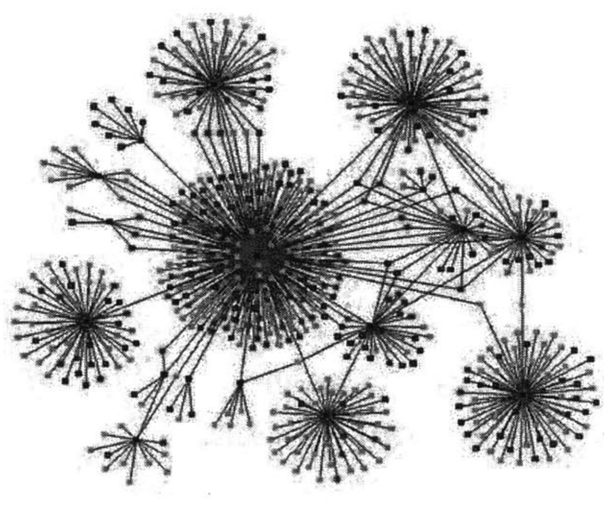

Figure 3 Little's illustration of a Dynamic Social Network (2008)

Dynamic Network Analysis

Recently, interesting findings have contributed to the emergence of a field known as Dynamic Network Analysis which is at the intersection of social network analysis (SNA) and multi-agent systems (MAS). This exciting branch of network science investigates larger, multi-mode and multi-plex biological and social networks among others. In the case of social networks, it explores how people influence each other's behavior which can be quite subtle when the network structure is not taken into account. When considering the network structure, however, the underlying mechanisms based on information and direct benefits can be observed both at the level of whole populations and at a local level in a crowd of friends or colleagues. Often, individuals care about aligning their response based on the prevalent behavioral tendency of their immediate neighbors or social group rather than the population as a whole.

When individuals have incentives to adopt the behavior of their neighbors in the network, there could be cascading effects, in which a new behavior starts in a small group of early adopters and then spreads radially outward through the network. This cascading effect could be also be referred to as a social contagion like a biological epidemic, (Centola, 2011). The Framingham Heart Study concluded that obesity spread through a social network (Christakis, 2008). In this research project, Christakis and Fowler evaluated 12,067 people regularly from 1971 and 2003. The statistical data showed that as people gained weight, their attitudes about what constituted an acceptable body weight shifted. In other words, a process of emulation took place where individuals started to be influenced by their friends and their neighbors which turned out to be more influential than their spouses or siblings in determining the acceptable body weight. Provided a certain threshold was reached, the authors contended that the behavior became an epidemic that spread quickly through the network and became hard to stop.

When we observe a large population over time, we see how culture evolves and emerges in the form of new ideas, beliefs, ideals, opinions, innovations, technologies, products and social norms. The pace at which cultural change happens depends on the underlying incentive to innovate or to conform. Either way, peer pressure seems to lead us to behave according to others behaviors. Let's examine further how collaborative innovation occurs.

Chapter 2: Collaborative Innovation Networks

In the last 15 years, the wave of corporate restructuring in large organizations has promoted efficiency and flexibility. Human resources worked increasingly through informal communities rather than formal hierarchies. A sub-branch of dynamic network analysis research, organizational network analysis has shown that connectivity within these informal social networks within organizations can have a major impact on performance, influence, learning and innovation (Battilana, 2012).

To continue to understand the inner working of this social dynamic, I examined interactions within a social group inside or outside an organization. Groups of people, as well as communities, have a collective intelligence that is different from individual intelligence of each group member (Apicella, 2012). There are many contexts in which one could study the impact of collective intelligence from sales and marketing to political voting. In this study, we will focus on collective intelligence in the context of innovation.

Dr. Gloor dedicated an entire book to this phenomenon that he qualified as swarm creativity (Gloor, 2006). In his study, Gloor argued that this phenomenon present in nature was the most effective process of innovation. Whether it is the birds migrating together from one region to the other or a group of people shifting their thinking from one paradigm to another, swarm creativity is enabled by Collaborative Innovation Networks (COINs). As defined in the book, "a COIN is a cyber-team of self-motivated people with a collective vision, enabled by the Web to collaborate in achieving a common goal by sharing ideas, information and work".

It is important to distinguish swarms from crowds. As implied in the COIN definition above, swarms have a common goal, a common direction. This is not systematically the case of a crowd. Following the work of E. Bonabeau who first considered the phenomenon of "swarming" outside the world of arthropods, Gloor also researched human swarms and compared them to bee swarms which are trying to find a new site for their hives to describe the behavioral dynamics of a COIN. Gloor concluded that collective intelligence of a swarm was much greater than that of any single individual. Gloor highlighted the following characteristics as pre-requisites for innovation as part of a COIN:

- Openness, diversity and trust are the foundation of any collaboration;
- A common vision or goal is crucial for members of a team to collaborate;
- A common code of ethics must be adopted;
- Direct contact is the most effective way to communicate. In other words, no gatekeepers or hierarchies;
- Members of COINs engage in an open conversation to generate new ideas.

Several examples could highlight the power of swarm creativity and COINs. These include the ascent of the world-wide web, the unexpected success of Yammer and the impact of Twitter. I will expand on those examples in the next iteration to highlight and to conclude on the different benefits of swarm creativity and COINs within an organization and in society at large.

Coolhunting and Coolfarming

Coolhunting and coolfarming are two major categories where swarm creativity may happen and COINs come together. On the one hand, coolhunting refers to swarm creativity process outside an organization most often on the Web (Gloor, 2007). On the other hand, coolfarming refers to the swarm creativity process that happens with the confines of an organization (Gloor, 2011a). The use of the word cool is borrowed from music history when some of the early Jazz musicians used the term to qualify outstanding work. Hunting refers to the idea of tracking and identifying new concepts, trends, individuals that result from a social movement on the World Wide Web. Farming refers to the idea of detecting and nurturing COINs with an organization.

In this chapter, I will highlight in further details the process of coolhunting and that of coolfarming. This includes the research methods that are incorporated in the Condor software. These methods measure three types of social network analysis metrics: structural, temporal and content. Furthermore, the goal would be to compare networking variables with real-world events. One way of doing this is to measure changes between events based on various social network analysis metrics including (betweenness, sentiment, complexity, oscillation in betweenness, contribution index, etc.), by looking at a time interval between events and also by pre-filtering using specific keywords.

An in-depth look at Coolfarming

In the previous section, we learned the difference between coolhunting and coolfarming. On the one hand, coolhunting pertains to detecting cool trends and COINs outside an organization. On the other hand, coolfarming pertains to identifying cool ideas and COINS inside an organization.

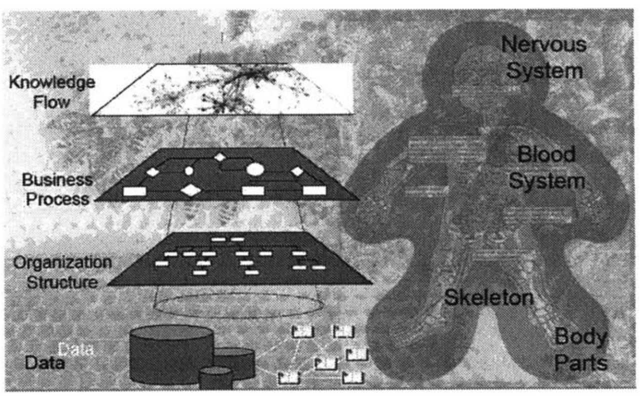

Figure 4 – Gloor's illustration of Knowledge Flow (Swarm Creativity, 2006)

As shown in figure 4, coolfarming is about taking a pulse on the nervous system of an organization. Organizational structure and business processes are the building blocks that enable collective intelligence to happen and that results in data generation and knowledge flow.

There are two applications of coolfarming. First, it is taking an idea from concept to reality using the creator. The process of dissemination follows the following framework from Collaborative Innovation Network (COIN) to Collaborative Learning Network (CLN) to Collaborative Interest Network (CIN) framework (Gloor, 2011b). Second, it is the process of nurturing existing COINs using Knowledge Flow Optimization (Gloor, 2012). Let's consider each one in detail.

Unlike project management, coolfarming is a bottom-up approach once a leader has defined an idea that engaged her/his peers. The table below summarizes the key differences with project management when it comes to motivation, management style, innovation type and measure of progress.

	Project Management	Coolfarming
Motivation	Extrinsic	Intrinsic
Management style	Supervised	Self-organized
Innovation type	Planned innovation	Disruptive Innovation
Measuring project progress	Fixed milestones	Dynamic development

Table 1 - Gloor's comparison of project management and coolfarming (Gloor, Coolfarming, 2010)

Once a team latches on the idea, and in their own time, the self-organized team-members of the COIN –build a product prototype. Using this group as an Alpha user, the COIN seeks feedback from friends and family of each member to test and to improve the product. Some members even decide the join the COIN. This immediate network circle acts as a Collaborative Learning Network or CLN, providing a source of Beta users, as well as external raving fans who can help the product's adoption grow to a tipping point at which point it reaches a word-of-mouth growth trend. Once the new product has outgrown this development stage, and is spoken about in mainstream media, it will be embraced by the Collaborative Interest Network (CIN). This is the commercialization phase of the growth phase, as CIN members will be willing to buy the product that they consider cool.

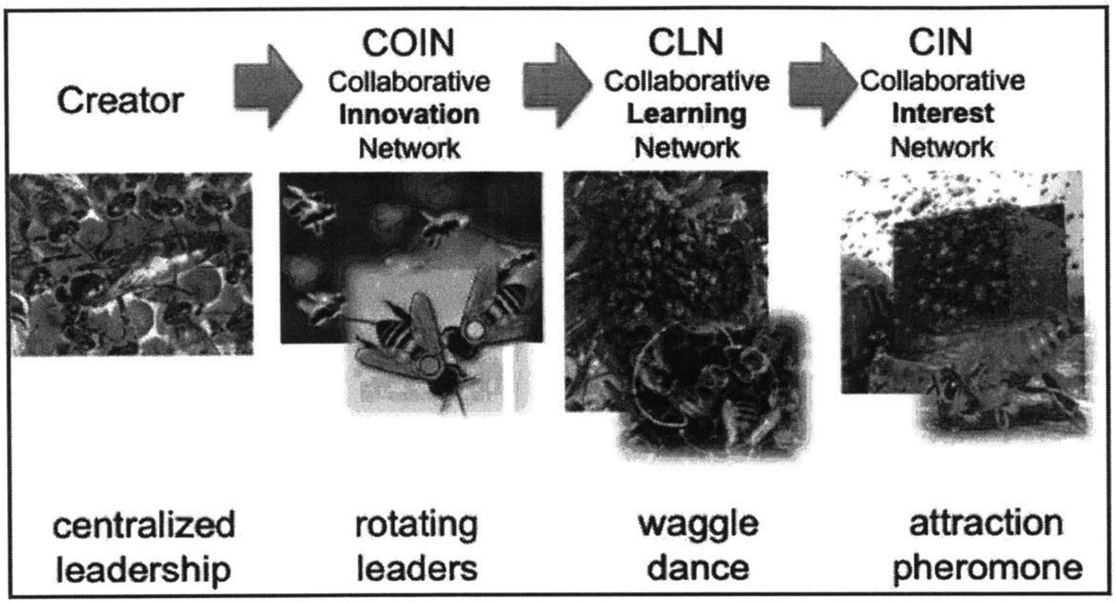

Figure 5 - Gloor's Coolfarming Framework (Coolfarming, 2010)

The second aspect of coolfarming involves nurturing of a COIN through a process called Knowledge Flow Optimization (KFO) (Gloor, 2012) The following figure shows the KFO process superimposed on the coolfarming stages in a typical case (Gloor, 2012).

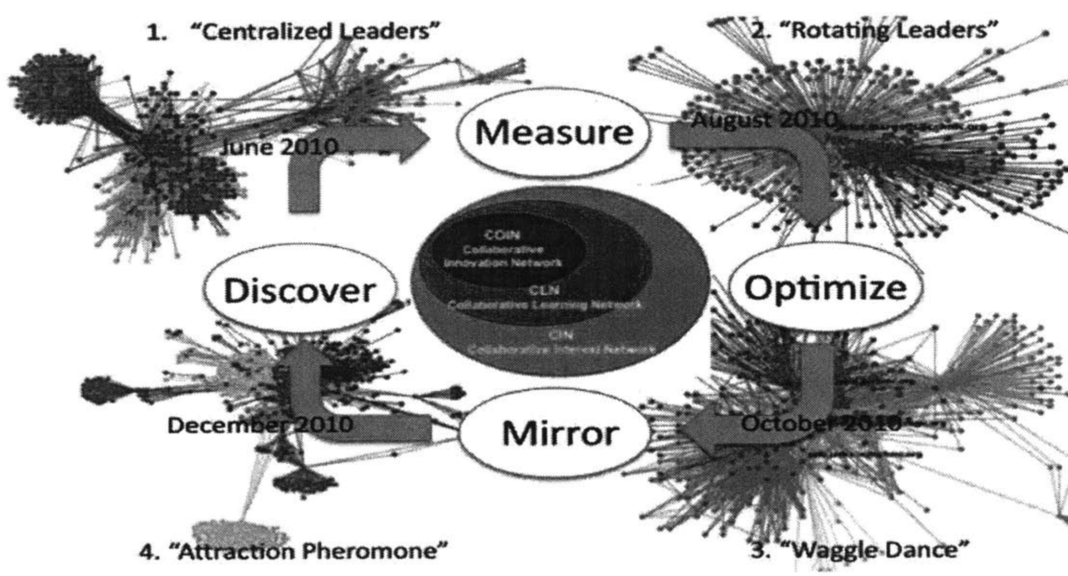

Figure 6 - Gloor's Knowledge Flow Optimization (Coolfarming, 2010)

The four steps of the optimization process (Gloor, 2009) are:

1. **Discover:** The goal of this step is to create a Social Network of people in the organization that can be analyzed.
2. **Measure:** The goal of this step is to analyze the structural and the temporal metrics of the social network defined in the previous step. As shown in figure 7, these metrics then help understand the pattern of communication and identify the improvements that can be made based on a project objective. In a creative context, it can be noticed that leadership is rotating, contribution is balanced among team members and emotions are fluctuating since the day-to-day activity is more random. In an operational context, it can be noticed that the structure of a team and the contribution among its members are less distributed. In addition, emotions are more balanced since the day-to-day is more deterministic.

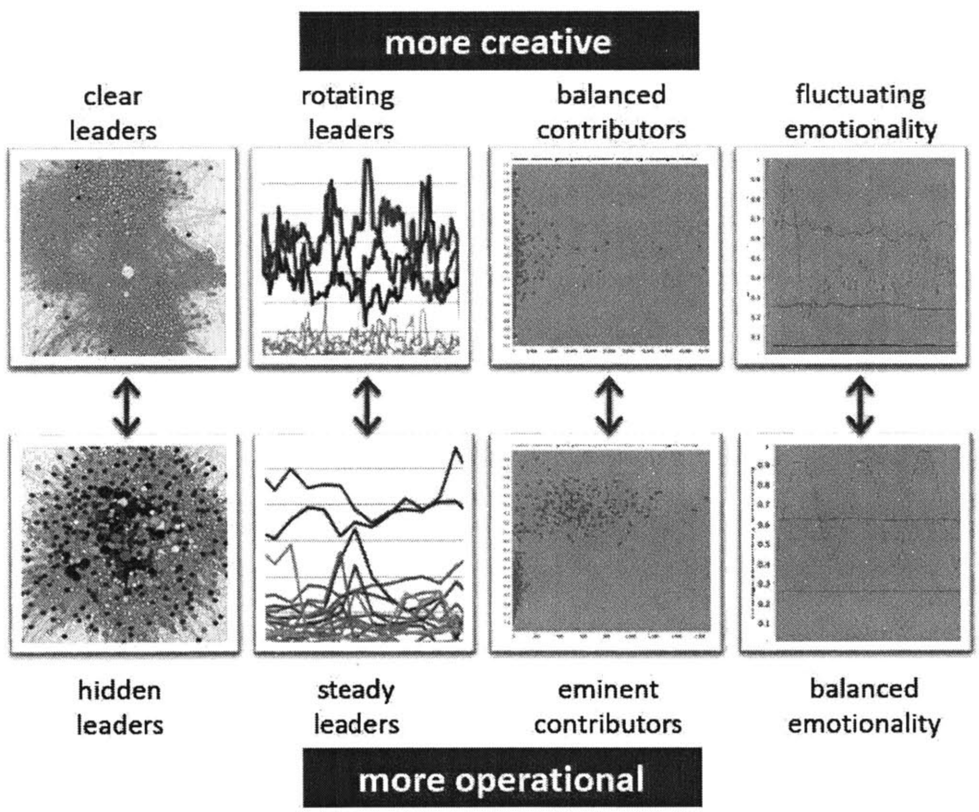

Figure 7 - Social signal metrics of a creative and an operational teams (Galaxy Advisors, Condor 2015)

3. **Optimize:** The goal of this step is to propose recommendations that would help a social network progress more effectively toward its objective.
4. **Mirror:** The goal of this step is to facilitate the improvement process by sharing with the team members' the patterns of communication that would lead to success. The actors can look at themselves in the mirror to become awareness of what they are doing right and what they can improve. This facilitation and fine-tuning process can be done on an ongoing basis.

Now that we have covered coolfarming in greater depth, I can focus on the applying of this methodology to the Enron email archive.

Coolfarming an Email Archive

The project of this thesis is to coolfarm the Enron public email archive. The Enron corpus has been widely used as a test bed for e-mail based SNA (Klimt, 2004) . I imported the Enron email archive into Condor, a SNA visualization software tool (Gloor, 2014). The dataset consists of 517,431 messages that belong to the 150 most active email users. We considered three methods to identify potential suspects:

(1) analyzing the overall social network and identifying individual closeness to suspect actors (2) comparing the convicted executive team members to the most influential executives in the social network (3) searching the collaboration clusters of suspicious activity according to a framework of three types of communities: Collaborative Innovative Networks (COINs), Collaborative Learning Networks (CLNs) and Collaborative Interest Networks (CINs).

Chapter 3: Market Power in an Electricity Wholesale Market

Since the 1990s, a restructuring of electricity markets in the United States has been aimed at driving down prices thanks to increased competition. This deregulation objective came with a trade-off between market power and efficiency.

The aim of this chapter is to analyze this trade-off by simulating the dynamic of strategic trading in a deregulated electricity wholesale market from the perspective a participating energy company. Further, this chapter introduces game theory and the application of market power in the context wholesale electricity. This was important element of Enron's role in the California Energy Crisis discussed in the next chapter.

According to Easley and Kleinberg "connectedness" of a complex socio-technical system is composed of two elements: first, the structural factor of interactions, and second, the interdependence in the behaviors of agents which are stakeholders or beneficiaries of the system. The first element was discussed in previous chapters using network science theory. The second element considers interconnected at the level of behavior using game theory (Easley, 2010). There are three ingredients to a game:

1. There is a set of participants, whom we call the players;
2. Each player has a set of options/strategies for how to behave/decide;
3. For each options/strategies, each player receives a benefit/profit that can depend on the strategies selected by other players.

In Energy Economic and Policy class at MIT, Professor Christopher Knittel introduced me to this market-driven social dynamic through the Electricity Strategy Game (ESG), an educational tool developed by Severin Borenstein and Jim Bushnell at the University of California at Berkeley. Class teams were randomly defined and competed with other class teams in a sequence of daily electricity spot markets. Spot market conditions varied from hour to hour and day to day. For the purpose of this game, there were four hours in each day. Teams had to develop strategies to deploy their assets over this sequence of spot markets while accounting for the cost structure of their portfolio, varying levels of hourly demand, and the strategies of other players. My team was composed of Sebastian Bastek and Obinna J. Ukwuani.

With a high capacity portfolio, my team was able to exercise market power during high demand hours. The idea of market power in economics and particularly in oligopolistic industrial market structure is the ability of a firm to profitably raise the market price of a good or service over marginal cost (MC).

In this simulation game, we first had to find out in which hours it would make sense to use our market power. We made use of it whenever the residual demand for our portfolio is higher than zero. In these hours our plants were required to fulfill the demand for electricity and we could set the price for electricity as high as we wanted because of the inelastic demand function. The residual demand for our portfolio was determined by calculating the difference between demand and rest of the world (RoW) supply/capacity.

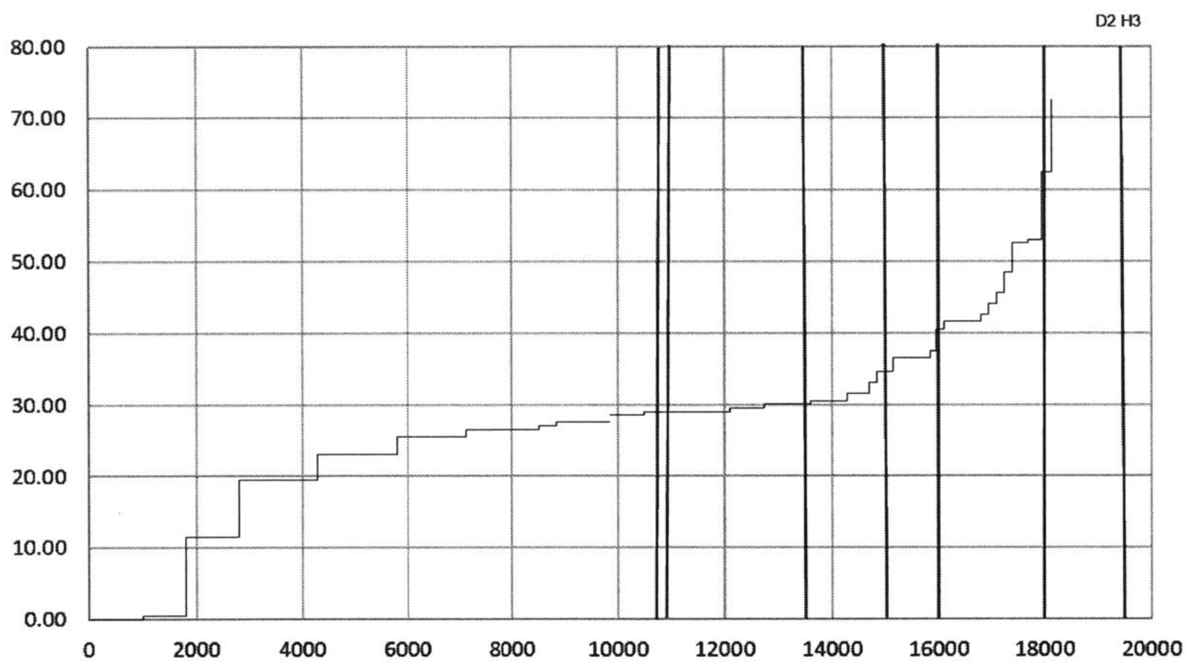

Figure 8 – Supply and Demand Chart for ROW Supply (Bastek et al., Energy Economics Class Spring 2014)

Figure 8 included the ROW-MC curves and the demand curves for each mean demand level. If there was no cross between the rest-of-the-world marginal cost curve and the demand curve for a given hour, then this meant that were price maker. Based on this graph, I could determine that in day 2 hour 3 there was definitely the potential for us to use our market power to our advantage. But even on days 1 and 3 in hour 3 it might be a good idea to speculate for market power to happen. Why was a demand of more than 18150 in these hours even possible? This was due to the fact that the demand data was not 100% certain of what the future demand would be. So, the demand in each given hour was a random variable that had a normal distribution with a mean equal to the forecast and a standard deviation equal to 3% of the forecast. This meant that there was a certain chance that we could use our market power during these two hours as well if the demand was more than 18150.

This game was a unique way to experience the concept of market power, the complexity of energy trading and the interdependence in the behaviors of individuals who were part of this marketplace. Energy traders spend their days forecasting future demand and anticipating other players' market moves on a daily basis to optimize the trade-off that exists between risk and profit for their firm.

In the next three chapters, I will endeavor to understand Enron's irrational exuberance in a period of flawed regulatory policies of a newly deregulated power market in California.

Chapter 4: The California Energy Crisis

California's experience with restructuring the electricity industry stigmatized its reputation when it comes to forward-looking energy policy ideas (FERC, 2001). The Electric Utility Industry Restructuring Act (Assembly Bill 1890) became law in 1996 in California. This was intended toward lowering electricity prices. Just four years later, the California electricity market was in crisis. The goals of the restructuring—lower prices for residential customers and more competitive prices for industrial customers—seemed farther away than ever. When the new wholesale power market started in March 1998, it worked relatively well for 18 months. By the summer of 2000, retail electricity prices in southern California reached a record level and generation capacity shortages resulted in power outages in northern California. The years of 2000 and 2001 were marked by a crisis situation that created a chain reaction from periodic blackouts to the demise of Enron Corporation. In this section, I will try to understand three major problems how they impacted Enron and the various stakeholders of the California Energy Crisis.

According to the Energy Information Agency (EIA, 2001), the situation can be highlighted in three interwoven problems:

1. **High Wholesale Electricity Prices:** Between June and July 2000, wholesale prices increased by 270% compared to the same period in 1998. By December 2000, prices cleared at $376.99 per MegaWatthour (MWh) which was 11 times higher than the average clearing price of $29.71 per MWh in December 1999.
2. **Intermittent Power Shortages:** Since 1999, California had experienced numerous rotating blackouts and involuntary curtailment of electricity. Stage 3 (less than 1.5% dispatch) emergency notifications, which may require rotating blackouts, increased from 1 in 2000 to 38 b the first half of 2001. Stage 1 (less than minimum operating reserves) and stage 2 (less than 5% dispatch) notifications increased from 91 in 2000 to 127 in the first half of 2001.
3. **Inconsistent Policy Designs:** The conjunction of high wholesale power prices in the newly privatized generation market and the imposition of retail price caps in the regulated distribution market cause severe financial distress for the incumbent utilities. Pacific Gas & Electric and Southern California Electric had to file for protection under Chapter 11 of the U.S. Bankrupcy Code. These utilities spent respectively $9 billion and $ 2.6 billion in unrecovered power costs.

As the deregulation process took its course in California, the supply-side market was partially privatized whereas the demand-side remained regulated (CBO, 2001). In the late 1990s, a total of 40% of the installed capacity was sold to private energy companies also referred to as independent power producers including Reliant, Dynergy and Enron. Incumbent utilities were still responsible for the electricity distribution. All players including the utilities were required to compete to buy electricity in the newly created day-ahead only market, the California Power Exchange (PX). Utilities were prevented from entering into longer-term agreements that would have allowed them to hedge their energy purchases and to mitigate day-to-day supply-side or demand-side fluctuations. By 2000, wholesale prices were deregulated. However, retail prices were regulated for incumbents. The only flexibility that incumbents managed to negotiate was the cost recovery of under-utilized capacity due to greater competition.

Federal and State policy makers expected that the price of electricity would decrease due to increased supply-side competition. Hence, the price of electricity was capped at the pre-deregulation level. When power demand rose, utilities had no financial incentive to expand production capacity as long term prices were capped. In addition, utilities were required by law to buy electricity from market exchange at uncapped prices when faced with imminent power shortages.

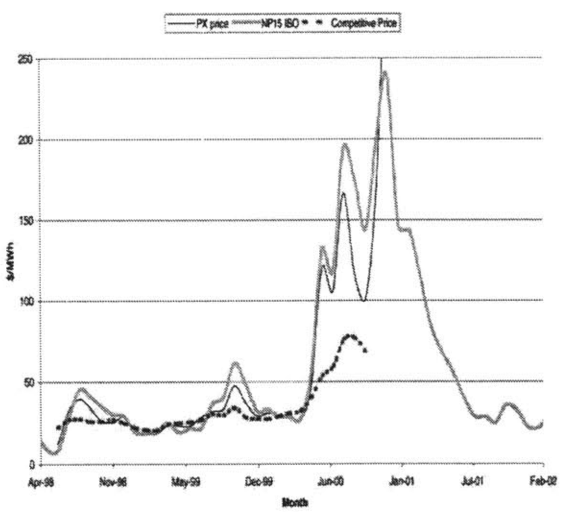

Figure 9 - California Electricity Prices (Bushnell, 2003)

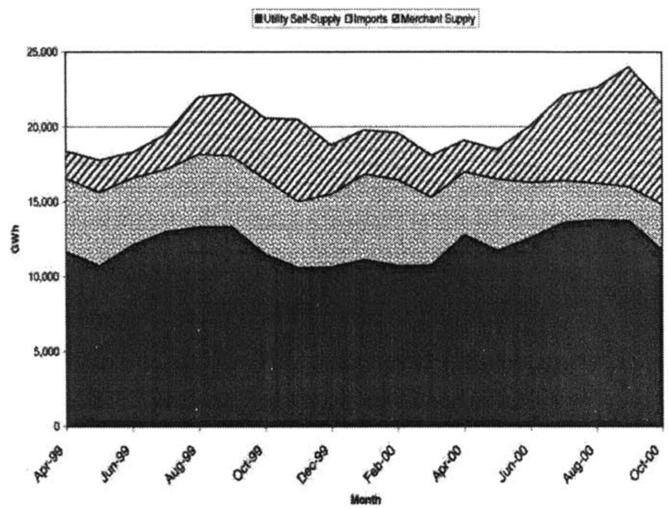

Figure 10 - California Supply Sources (Bushnell, 2003)

In figure 9, the Power Exchange (PX) price line represents the monthly average unconstrained PX price through the 32 months of the PX's existence. Over the course of 1998 and 1999, substantial hydro resources and significant imports contributed to an annual average price hovering respectively around $26/MWh and $28/MWh. By May 2000, energy prices climbed to record levels averaging for the year at $110/MWh. Incumbent utilities Pacific Gas & Electric and Southern California were caught in the rate-freeze policy trap of $60/MWh. They were losing $50 for each MWh they carried to their distribution customers during this period. The resulting financial distress engendered a physical supply crisis as shown in Figure 10.

Researchers (Bushnell, 2004) investigated the extent of supplier market power before the crisis. In figure 9, the dashed line estimates the "perfectly competitive" price and measures supplier market power which can be expressed as a margin over a counter-factual "perfectly competitive" price.

When expressed in percentage the mark-up over competitive levels, prices during August 1998 were exhibiting similar levels of market power to that experienced during the crisis. As the price increased, the difference between 1998 and 2000 can be highlighted in figure 10 where we see much lower imports due in part to lack of long-term supply arrangements.

Enron's management can be blamed for much of this crisis as they exploited the intertwined problems above to hijack the market. However, when thinking broadly about this situation to consider both the economic distress and the physical supply shocks, Bushnell argues that Federal and State regulators also contributed to this extreme scenario with inconsistent policies and slow interventions to contain the emergency situation. One may argue that many of these were present in other power markets and none of which has experienced the extent of the Californian crisis. Bushnell found that the only factor that was not present in other power markets was the lack of incentives to seek contracts or other forms of long-term supply arrangements a transition mechanism. Widespread long-term contracts and high-volume trading in the spot market are not mutually exclusive.

The conjunction of these deregulation policies and the continued regulation policies in neighboring States provided less flexibility to incumbent utilities. This opened the way for an extreme scenario where wholesalers particularly Enron artificially created power shortages in the daily spot markets to make short-term gains. Enron's stock benefited well from this period of 1998 to 2000.

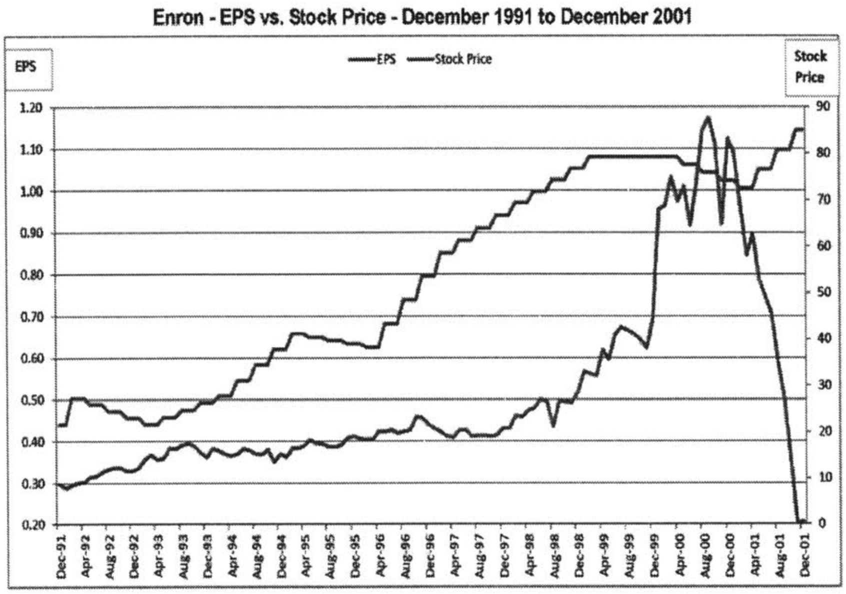

Figure 11 - Enron Earning-per-Share (EPS) and Stock Price (SeekingAlpha.com, 2012)

One can notice that Enron's trouble started when San Diego Gas & Electric Company (SDG & E) filed in August 2000 a complaint alleging manipulation of the markets. However, it took until September 2001 before prices normalized. In the meantime, Enron collapsed filing for bankruptcy protection in December 2001. By February 2002, the Federal Energy Regulatory Commission (FERC) started an investigation of Enron's involvement. FERC later referred to this manipulation as megawatt laundering where Enron bought electricity in-state at below cap price to sell it out-of-state, creating shortages. Another approach was to overload the bottlenecks in the power transmission network in order to prevent a faster power transfer from one region to the other. Later that year, the Enron Tapes of trader conversations were material evidence that market manipulation took place. 24 members of Enron's executive team (Eichenwald, 2003) were convicted over the course of 2003. The next section studies the dynamic of social interactions among the 24 Enron's executive that were convicted (NYTimes, 2003).

Chapter 5: Enron's Social Network Analysis

The Data

The Enron dataset was made public during the Federal Energy Regulatory Commission investigation. This raw corpus contains approximately 160 users including Enron's senior management. The dataset contains about 619,556 messages. A paper about "The Enron Corpus: A New Dataset for Email Classification Research", was published in the proceedings of the 2004 European Conference on Machine Learning (ECML).

During its early research use at SRI and at MIT around 2004, numerous integrity problems were corrected. Melinda Gervasio at SIR and Leslie Kaelbling at MIT can be credited for correcting these issues. In the past decade, the dataset become a great a learning platform for the scientific community across fields from machine learning to social network analysis. By 2011, Kimmie Farrington and her colleagues used the Enron email dataset to study crowdsourcing human- vs. computer- generated classification explanation. A paper on their work was published as part of the proceedings of the 2012 Association for the Advancement of Artificial Intelligence Spring Symposium on the Wisdom of the Crowd.

My interest in examining this dataset is to explore senior management interaction in the last two years of the company when the stock price peaked in 2000 during the initial period of the California Energy Crisis and collapsed by the end of 2001. It is interesting to highlight that in the 1990s, Fortune had named Enron "America's Most Innovative Company" for six consecutive years.

Research Objectives

Coolfarming projects have indicated that its distributed form of management leads to increased creativity, better performance and self-maintained/sustainable project development and management. Kidane and Gloor have shown that oscillation in communication patterns indicates rotating leadership which translates into higher levels of creativity (Kidane, 2007). In addition, stability in emotionality indicates centralized leadership which implies operational excellence.

In another coolfarming project related to the Enron case, Zhao and Gloor used Abstract iQuest, a novel software system to understand organizational behaviors in greater granularity (Gloor, 2004). By mining communication archives and actors discussion topics by semantic social network analysis, new insights could be gathered when correlating organizational performance and creativity with key social metrics.

In similar research, Gloor et al. have used Condor, a predictive search and analysis software to test these social metrics outside the organization (Gloor, 2009). This coolhunting study consisted in identifying trends on the web 2.0 through semantic social network analysis. By mining communication archives on the Web, blogs and online forums, Condor could map the social network, computer the key structural and temporal social metrics. The findings confirmed that cool trends around a brand, a product or a politician could be correlated to better performance whether it is in winning a political election or in witnessing increased stock prices.

Regarding the Enron case, numerous papers have been published. The Enron email corpus has become a benchmark among researchers because it represents a rich temporal archive of internal communication within a real-world global organization facing extreme situations where profits are maximized in 2000 and financial collapse is reached a year later. In an interesting paper in 2005, Diesner et al. could relate the frequency and the direction of email communication to various organizational situations. In a crisis context, this study revealed that (a) communication among actors becomes more diverse (b) previously disconnected actors begin to engage in mutual communication, and (c) social interaction intensifies and spreads more widely throughout the network (Diesner, 2005).

In 2007, Davis et al. offered further insights on communication patterns and structures during the period of collapse of Enron in 2001. Based on network disintegration theory, this study found that (a) communication network becomes increasingly operational. (b) the centralization of the network gravitates toward senior management as the company moves toward bankruptcy ((Davis, 2007).

In 2011, Michalski et al. were able to match organizational structure and social network based on two corporate cases including Enron and a manufacturing company. The insight behind this study was to consider social network metrics as another possible way to align company management structure and to optimize company performance (Michalski, 2011).

Based on these prior research findings, I set forth the following hypotheses to be validated through the social network analysis of the Enron 2000 and 2001 email archive.

Hypothesis I: The most active people are likely those convicted.

Hypothesis II: Convicted individuals have high total influence in the social network.

Hypothesis III: The more the communication patterns oscillate between centralized and rotating leadership, the more distributed the creativity, in this case the fraudulent activity.

Hypothesis IV: The communication pattern signals a role in the California Energy Crisis.

Research Methods

Galaxy Advisor's Condor (Gloor, Nov. 6. 2004) is a unique Coolfarming software tool that helps us study the graphical representation of the social network and computes the structural and the temporal social metrics of the network. Besides the structural metrics that were discussed in the introduction, it is important to note that there are two other categories of SNA metrics: temporal metrics and content metrics. Temporal metrics help us measure the social interaction over time in terms of leadership rotation, activity over time, and response time. Content metrics help us understand individual and collective contribution.

In the Enron case, I will consider measuring the following:

Structural Metrics

- **Betweenness Centrality:** refer to the Social Network Analysis section of the Introduction Chapter.
- **Degree Centrality:** refer to the Social Network Analysis section of the Introduction Chapter.

Temporal Metrics

- **Oscillation Frequency of Betweenness Centrality:** This is a temporal measure of an actor's betweenness centrality.

- **Oscillation Frequency of Contribution Index:** This measures how an actor's Contribution Index changes over time.
- **Average # of nudges before an actor responds:** This measures an actor's responsiveness. It is the average number of messages an actor needs to receive from another actor before he/she responds to that actor.
- **Average response time:** This is another measure of an actor's responsiveness. It is the absolute time elapsed between an actor receiving a message and he/she responding to it.
- **Average number of messages received by an actor from others:** This measures the quantity of email received by an actor from each of the other actors in his/her immediate network.

Content Metrics

- **Contribution Index:** This measures of how many messages an actor sends in relation to those she/he receives. In other words, it is a measure of level of interaction of a given actor. Possible values fluctuate between -1 and +1. In the situation where the actor only receives messages without sending any, the contribution index tends toward -1. The other extreme of +1 would signal the opposite when an actor only sends messages without receiving any (Gloor, et al., 2007). For most people, the contribution index tends to be between those two extreme situations.

These metrics formed the set of explanatory variables used in the correlation analysis with the independent variable, suspect activity. In other words, we want to see whether or not interacting with criminals leads to becoming a convicted actor.

When conducting a preliminary social network analysis on the data using Condor, I measured the activity level and the word usage pattern in the 2000 and the 2001 graphs below. It becomes evident that California and trading are top keywords in the email communication among the top 175 most active actors. In addition, Jeff and Vince appear to be the most cited actors in the email interaction.

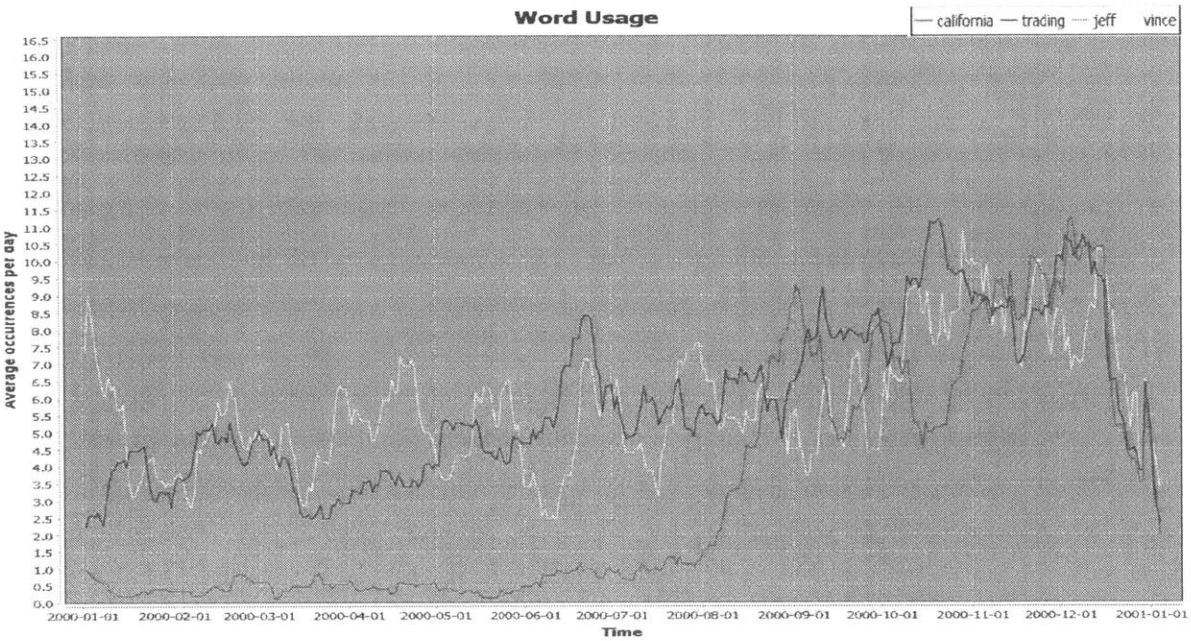

Figure 12 - Top Common Words in 2000 in the Enron dataset

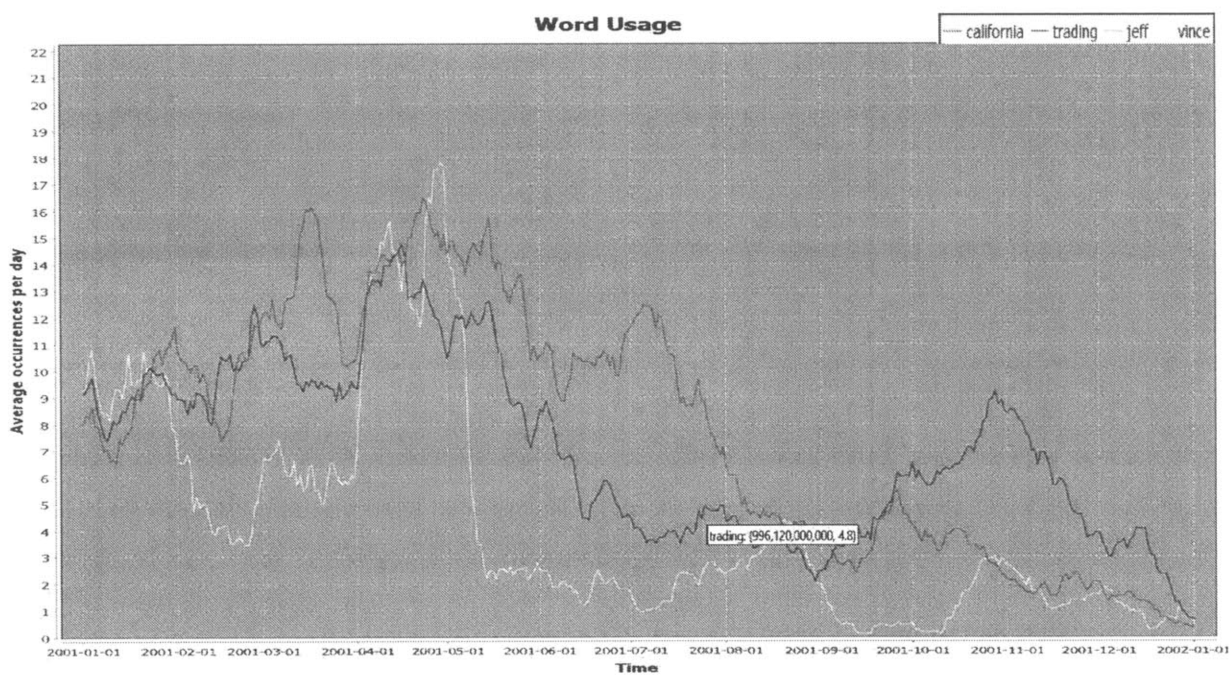

Figure 13 - Top Common Words in 2001 in the Enron dataset

In the period of 2000 to 2001, several sample groups were selected:

- **Outgroup**: 25344 actors including the most active by total messages. These actors were employed either inside or outside the Enron organization.
- **Ingroups:** 140 suspect actors. This included three samples: one experimental group that included those convicted actors, one control group based on the most active actors by total messages and one control group based on actors selected randomly convicted. These actors were employed either inside or outside the Enron organization.
- **In/Ingroup:** 24 which is primarily composed of the actors that were convicted. 17 of these actors were found to be connected with more than 10 email interactions in the social network. We will only consider these 17 for the purpose of this analysis.

T-test and correlation analysis were performed using both Microsoft's Excel and IBM'SPSS, a predictive analytics software. Pearson coefficients and two-tailed significance values were computed from correlations of suspect activity with each of the explanatory variables. Correlations were considered significant for p-values <= 0.05. Please note that outliers and email duplicates were removed from the samples.

Results of Coolfarming on email archive

Figure 14 show the email communication network of the 25344 actors. Internal interaction is at the center.

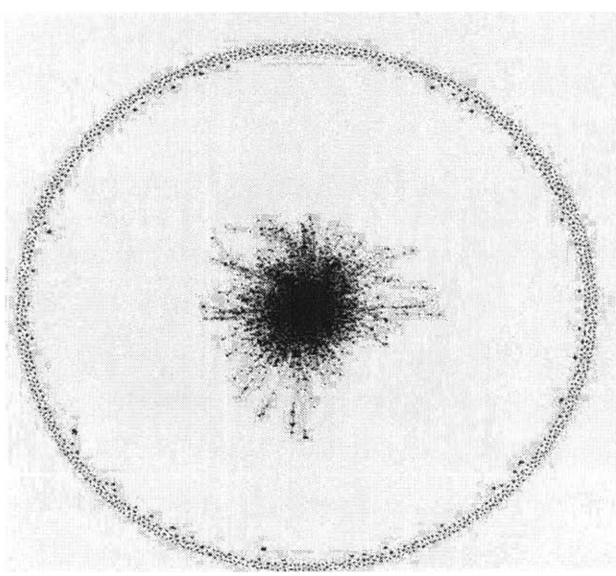

Figure 14 - Static view in the Email Network among 25344 actors

Figure 15 highlights the email communication network primarily among Enron staff. This core network is composed of 4667 actors. The yellow dots represent 3277 actors that were Enron employees.

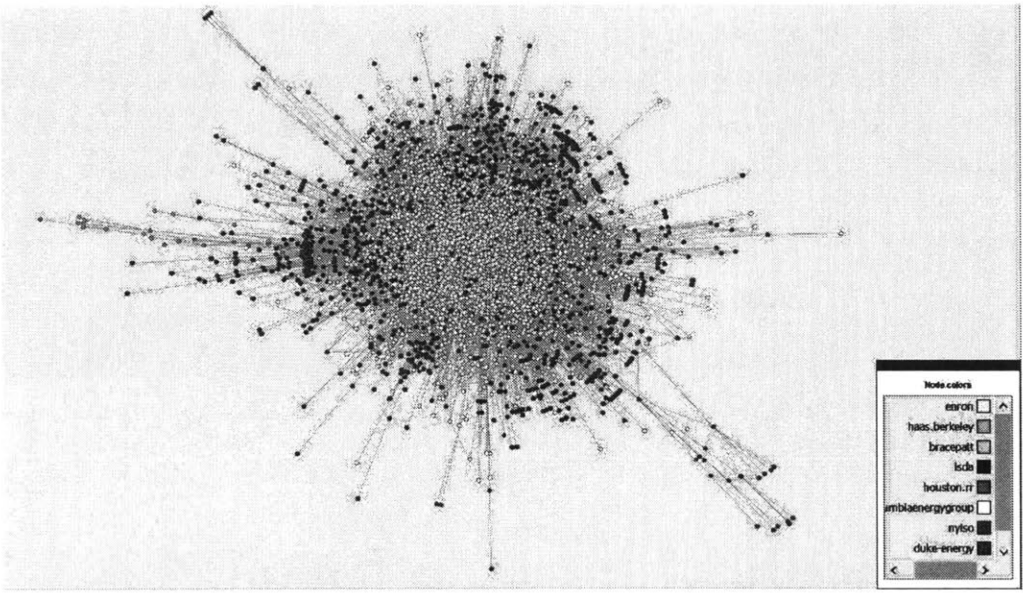

Figure 15 - Static view in the Email Network among 4667 actors

Figure 16 focuses on a shortlist of 140 suspects. The hubs are highlighted according to total messages.

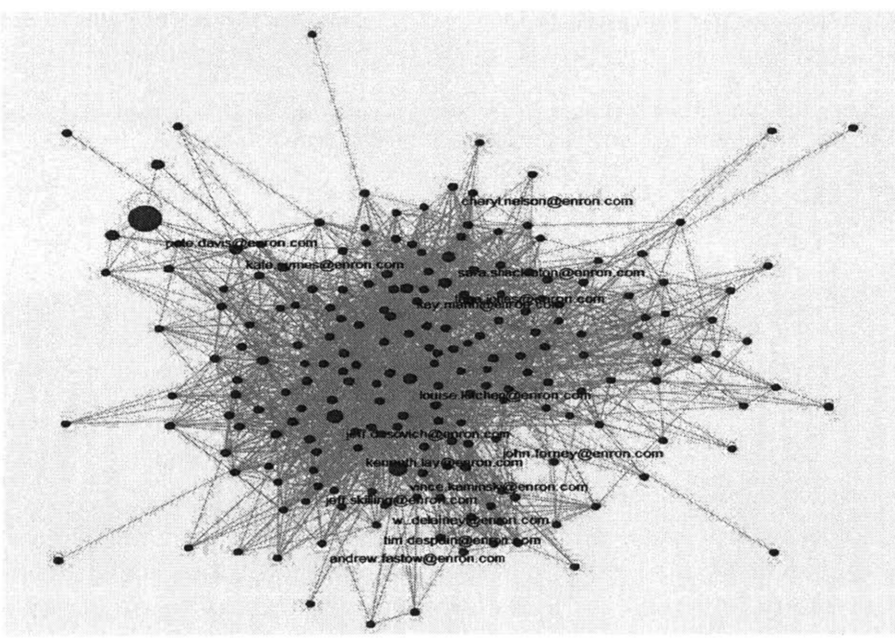

Figure 16 - Static view in the Email Network among 140 actors by total message

Figure 17 focuses on a shortlist of 140 suspects. The hubs are highlighted according to total influence. It is interesting to note that those convicted are not the most influential.

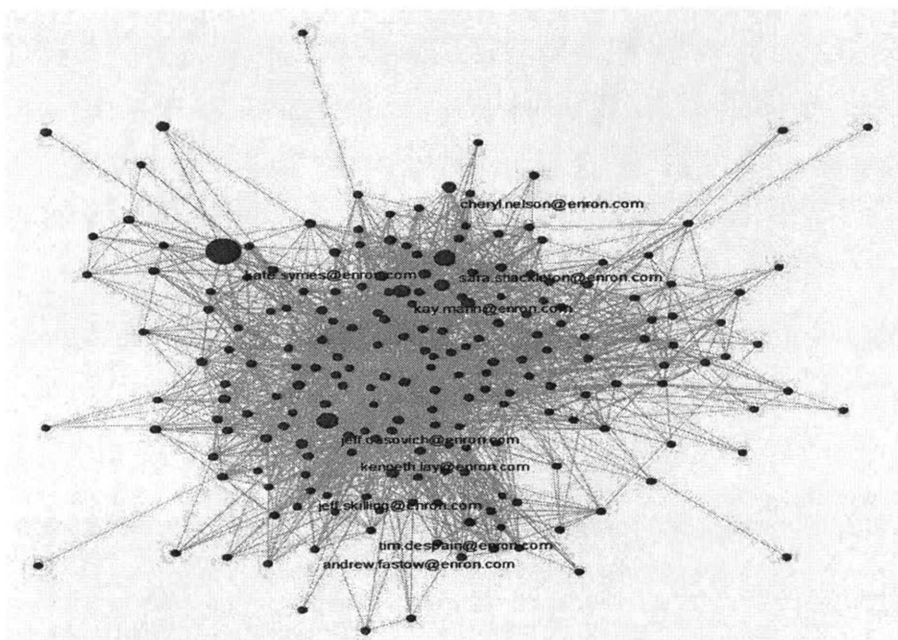

Figure 17 - Static view in the Email Network among 140 actors by total influence

Condor was used to measure the social network metrics and interactive dynamics as described in the Research Methods section earlier.

The following two graphs show the Contribution Index of all the actors in each of the experimental groups.

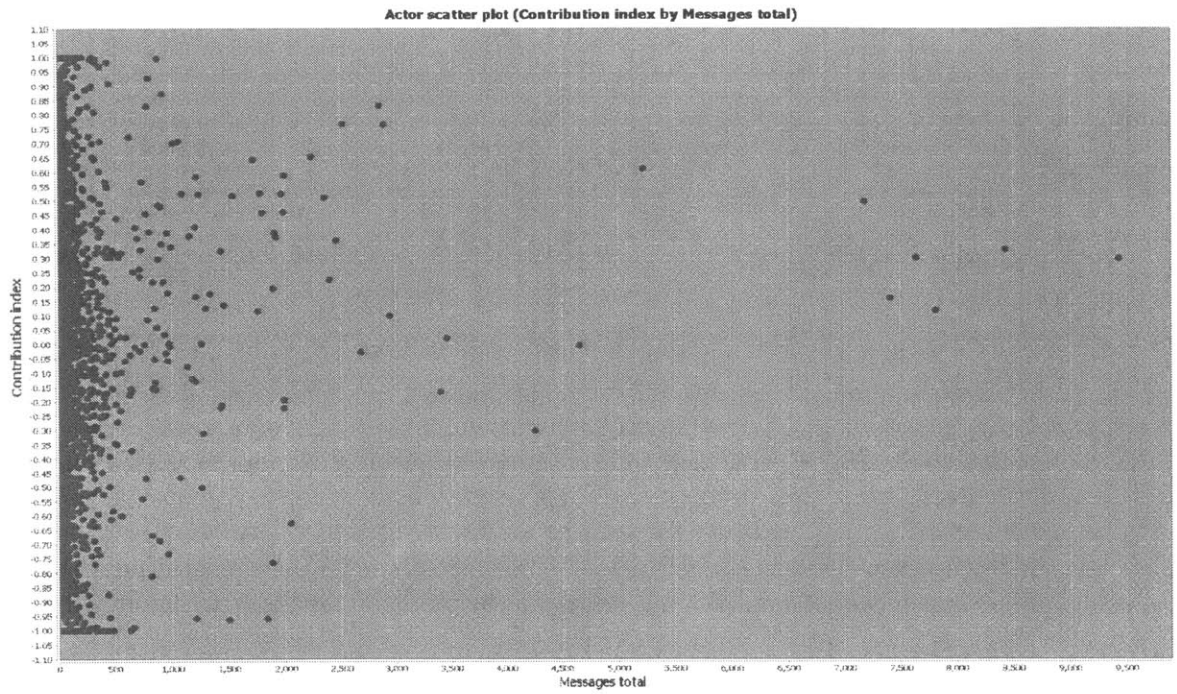

Figure 18 - Contribution Index in the Email Communication

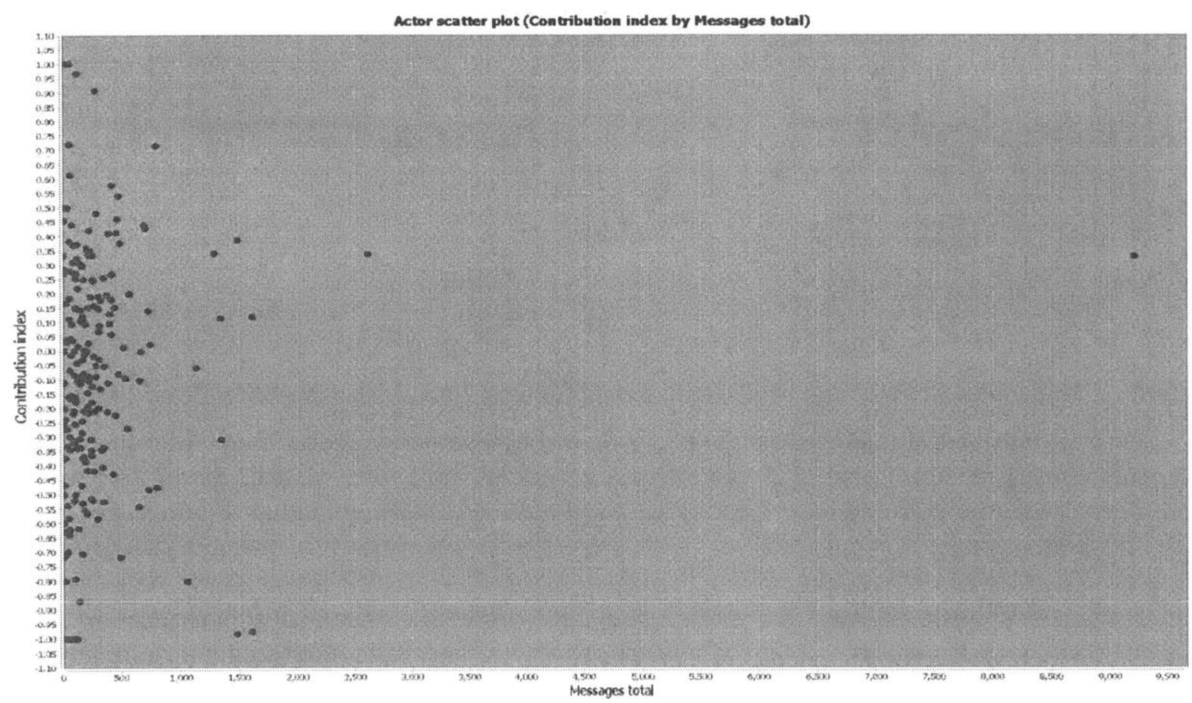

Figure 19 - Contribution Index in the Email Communication among 140 actors

In the figures 18 and 19, it can be determined that both sample's contribution index are well-balanced. This implies that Enron was a more creative than operational environment.

I also decided to observe the behavior of the betweenness oscillation at two particular extreme conditions: when the company is experience top market valuation and when the company is going through bankruptcy.

In the 2000, the graphic below shows that the betweenness oscillation appears to be more distributed hence more creative resulting in the company maximizing in market capitalization. In 2001, betweenness oscillation is less distributed which implies more operational activity as the company prepares for bankruptcy. The correlation analysis between both years also confirmed it marginally significant with a Pearson correlation coefficient of - 0.135 with a p-value of 0.077.

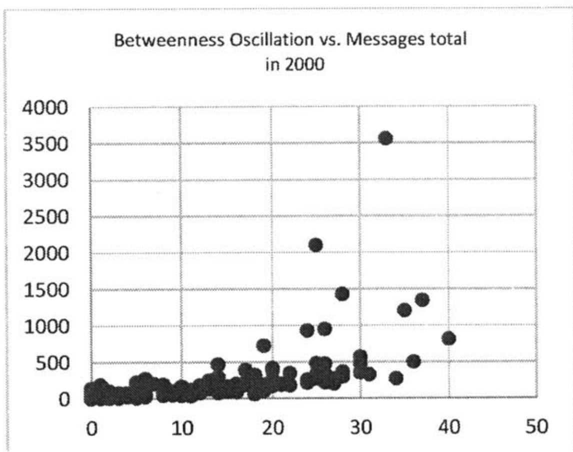

Figure 20 - Betweenness Oscillation in 2000 among total message of 140 actors

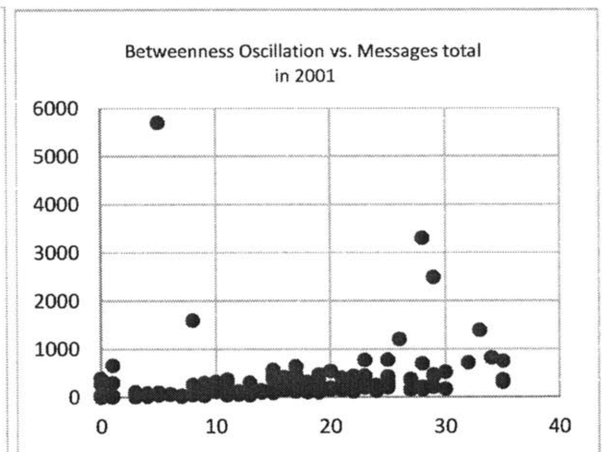

Figure 21 - Betweenness Oscillation in 2001 among total message of 140 actors

Using the social metrics computations from Condor, I performed statistical analysis to test the likelihood of suspect activity and to determine the potential correlations that might exist between social metrics of the three samples of 25344 actors, 140 actors and 24 actors.

Correlation

First, I considered the descriptive statistics and the correlation that exists between the social metrics for each sample on the condition that the p-value < 0.05.

Descriptive Statistics

Variable	Outgroup (N=25344) M	SD	Ingroup (N = 140) M	SD	Ingroup - Most active (N=140) M	SD	Ingroup - Random (N=140) M	SD	In/Ingroup (N = 17) M	SD
Betweenness centrality oscillation	1.36	5.29	24.92	18.43	27.50	18.05	11.56	13.71	2.31	4.45
Messages sent	10.08	104.73	252.54	643.17	275.55	647.56	76.44	161.50	9.19	15.61
Avg sentiment	0.54	0.12	0.51	0.06	0.51	0.06	0.52	0.10	7.50	0.10
Betweenness centrality	34371.28	348665.31	79.74	99.14	75.04	93.24	92.25	137.16	7.50	10.25
Degree centrality	4.57	18.82	17.29	10.25	18.23	10.01	9.35	6.93	4.06	2.74
Avg complexity	3.46	3.21	5.87	1.31	5.96	1.12	5.96	1.48	5.41	2.38
Total influence	1.12	13.99	20.93	37.95	24.09	42.54	7.71	22.44	0.30	0.55
Messages received	10.08	70.84	252.54	451.92	275.55	448.81	76.44	114.56	9.60	10.40
Messages total	20.17	166.60	505.09	1050.92	551.09	1048.26	152.88	255.81	18.38	19.67
Contribution index	-0.18	0.84	-0.09	0.40	-0.07	0.39	-0.08	0.44	-0.08	0.60
Avg emotionality	0.36	0.13	0.24	0.05	0.23	0.04	0.24	0.06	0.27	0.10
Average influence per message	0.05	0.11	0.09	0.07	0.10	0.07	0.10	0.18	0.03	0.06

Table 2 – Descriptive Statistics for the Outgroup, the Ingroups and the In/Ingroup

For the population of 25344 actors, the following correlations were observed:

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.49**	1.00										
3 Avg Sentiment	-0.02	-0.01	1.00									
4 Betweenness centrality	0.55**	0.77**	-0.01	1.00								
5 Degree centrality	0.74**	0.77**	-0.01	0.91**	1.00							
6 Avg complexity	0.19	0.08	0.32	0.07	0.12	1.00						
7 Total influence	0.35	0.76**	-0.03	0.65**	0.63	0.00	1.00					
8 Messages received	0.61**	0.79**	-0.03	0.75**	0.79**	0.08	0.72**	1.00				
9 Messages total	0.57**	0.97**	-0.02	0.80**	0.82**	0.08	0.78**	0.92**	1.00			
10 Contribution index	0.04	0.05	0.34	0.03	0.04	0.78**	-0.04	-0.02	0.03	1.00		
11 Avg emotionality	-0.21	-0.08	-0.25	-0.08	-0.14	(0.89)**	-0.02	-0.09	-0.09	(0.75)**	1.00	
12 Avg influence per message	0.06	0.02	-0.06	0.01	0.04	-0.03	0.07	0.05	0.03	-0.22	-0.03	1.00

p Values *p < .05, **p < .01

Table 3 – Pearson's correlation matrix of the measured variables in the Outgroup (N = 25344)

- Message metrics' particularly total messages had a very significant correlation with betweenness centrality, degree centrality and total influence. This may imply that the Enron convicts were among the most active actors in the social network acting as hubs through which an epidemic could spread.

- Total influence's correlation is also significant with the structural social metrics particularly betweenness centrality and degree centrality. This may mean that if the Enron convicts had a high total influence, they must exhibit high betweenness centrality and degree centrality in the network.

For the experimental group of 140 actors, the following correlations were observed:

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.25	1.00										
3 Avg sentiment	-0.03	-0.08	1.00									
4 Betweenness centrality	0.62**	0.22	-0.04	1.00								
5 Degree centrality	0.82**	0.22	-0.02	0.77**	1.00							
6 Avg complexity	0.29	0.22	0.12	0.16	0.26	1.00						
7 Total influence	0.52**	0.48	-0.07	0.29	0.45	0.24	1.00					
8 Messages received	0.31	0.84**	-0.14	0.19	0.24	0.19	0.44	1.00				
9 Messages total	0.28	0.97**	-0.11	0.22	0.24	0.22	0.48	0.94**	1.00			
10 Contribution index	0.07	0.23	0.13	0.13	0.12	0.38	0.22	-0.08	0.11	1.00		
11 Avg emotionality	-0.16	0.12	0.02	-0.06	-0.16	-0.45	0.03	0.09	0.11	-0.30	1.00	
12 Average influence per message	0.31	0.04	-0.04	0.04	0.21	0.25	0.43	0.17	0.10	-0.17	0.05	1.00

p Values *p < .05, **p < .01

Table 4 – Pearson's correlation matrix of the measured variables in the Ingroup – sample (N = 140)

- Degree centrality's correlation with betweenness centrality and betweenness centrality oscillation is very significant. This implies that the convicts acted as hubs to infect others with the Enron "culture". The illicit behaviors spread and resulted in the formation of Collaborative Innovation Networks (COINs) within the organization.

- Total influence's correlation with betweenness centrality oscillation is somewhat significant. This may mean provided the convicts were influential, their illicit activities were distributed in the network.

For the sample of 140 most active actors, the following correlations were observed:

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.22	1.00										
3 Avg sentiment	0.08	-0.07	1.00									
4 Betweenness centrality	0.49**	0.16	-0.01	1.00								
5 Degree centrality	0.80**	0.20	0.05	0.69**	1.00							
6 Avg complexity	0.32	0.25	0.16	0.13	0.24	1.00						
7 Total influence	0.44	0.45**	-0.10	0.17	0.37	0.25	1.00					
8 Messages received	0.27	0.82**	-0.13	0.10	0.19	0.22	0.40**	1.00				
9 Messages total	0.25	0.97**	-0.10	0.14	0.21	0.25	0.45**	0.94**	1.00			
10 Contribution index	-0.01	0.24	0.09	0.10	0.08	0.28	0.22	-0.13	0.09	1.00		
11 Avg emotionality	-0.05	0.20	0.04	0.02	-0.04	-0.18	0.07	0.18	0.20	-0.19	1.00	
12 Average influence per message	0.23	0.04	-0.18	-0.01	0.12	0.25	0.55**	0.17	0.10	-0.10	0.07	1.00

p Values *p < .05, **p < .01

Table 5 – Pearson's correlation matrix of the measured variables in the Ingroup - most active (N = 140)

- Degree centrality's correlation with betweenness centrality and betweenness centrality oscillation is very significant. This implies that the convicts acted as hubs to infect others with the Enron "culture". The illicit behaviors spread and resulted in the formation of COINs within the organization.

- Total influence's correlation with betweenness centrality oscillation and total messages is somewhat significant. This may mean provided the convicts were influential, they were among the most active communicators and they nurtured COINs engaged in illicit activities.

For the sample of 140 random actors, the following correlations were observed:

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.70**											
3 Avg sentiment	-0.01	-0.04	1.00									
4 Betweenness centrality	0.69**	0.54**	0.03	1.00								
5 Degree centrality	0.85**	0.71**	0.04	0.80**	1.00							
6 Avg complexity	0.20	0.13	0.01	0.15	0.17	1.00						
7 Total influence	0.59**	0.92**	0.00	0.47	0.63**	0.10	1.00					
8 Messages received	0.75**	0.71**	-0.08	0.57**	0.69**	0.11	0.77**	1.00				
9 Messages total	0.78**	0.95**	-0.06	0.59**	0.76**	0.13	0.92**	0.89**	1.00			
10 Contribution index	0.12	0.27	-0.03	0.13	0.14	0.27	0.11	-0.14	0.11	1.00		
11 Avg emotionality	-0.14	-0.09	0.04	-0.10	-0.20	(0.44)**	-0.03	-0.04	-0.07	-0.35	1.00	
12 Average influence per message	0.06	0.03	0.17	-0.01	0.04	0.07	0.13	0.14	0.08	-0.25	-0.08	1.00
p Values *p < .05, **p < .01												

Table 6 – Pearson's correlation matrix of the measured variables in the Ingroup - random sample (N = 140)

- The message metrics' particularly total messages has a very significant correlation with betweenness centrality, degree centrality and total influence. This may imply that the Enron convicts were among the most active actors in the social network acting as hubs through which an epidemic could spread.

- Total influence's correlation with messages total and messages sent is very significant. This may mean that those convicted were influential and the most active communicators.

- Degree centrality's correlation with betweenness centrality and betweenness centrality oscillation is significant. This may imply that those convicted were nurturing COINs engaged in illicit activities.

When considering the 17 convicted actors in the Outgroup (N = 25 344), the Ingroup (N = 140) and In/Ingroup (N = 17), the following statistics and correlations were observed:

Descriptive Statistics

	17 convicted in Outgroup (N = 25344)		17 convicted in Ingroup (N = 140)		17 convicted in In/Ingroup (N = 17)	
Variable	M	SD	M	SD	M	SD
Betweenness centrality oscillation	24.06	19.09	16.81	16.44	2.31	4.45
Messages sent	110.56	232.34	58.13	97.00	9.19	15.61
Avg sentiment	0.54	0.09	0.53	0.09	7.50	0.10
Betweenness centrality	1914746.41	3919364.10	93.63	180.69	7.50	10.25
Degree centrality	117.38	192.79	18.06	13.39	4.06	2.74
Avg complexity	5.66	1.66	6.12	0.98	5.41	2.38
Total influence	5.28	10.18	2.78	5.47	0.30	0.55
Messages received	234.69	360.83	96.88	132.77	9.60	10.40
Messages total	345.25	462.10	155.00	209.44	18.38	19.67
Contribution index	-0.34	0.46	-0.24	0.35	-0.08	0.60
Avg emotionality	0.05	0.04	0.06	0.07	0.27	0.10
Average influence per message	0.05	19.09	1.20	0.45	0.03	0.06

Table 7 – Descriptive Statistics for those *convicted* actors (N = 17) in the Outgroup, the Ingroup and the In/Ingroup

Convicted in the Outgroup

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.66**	1.00										
3 Avg sentiment	0.17	0.01	1.00									
4 Betweenness centrality	0.63**	0.09	0.21	1.00								
5 Degree centrality	0.68**	0.17	0.22	0.99**	1.00							
6 Avg complexity	0.30	0.23	0.22	0.05	0.08	1.00						
7 Total influence	0.69**	0.98**	0.07	0.14	0.21	0.31	1.00					
8 Messages received	0.76**	0.18	0.11	0.94**	0.93**	0.09	0.19	1.00				
9 Messages total	0.93**	0.64**	0.09	0.77**	0.81**	0.19	0.65**	0.87**	1.00			
10 Contribution index	0.03	0.52**	0.08	-0.40	-0.36	0.51**	0.44	-0.38	-0.04	1.00		
11 Avg emotionality	-0.30	-0.02	-0.03	-0.13	-0.15	(0.73)**	0.32	-0.17	-0.15	-0.25	1.00	
12 Average influence per message	-0.09	-0.12	0.48**	0.04	0.03	0.34	0.02	-0.05	-0.10	-0.07	0.18	1.00

p Values *p < .05, **p < .01

Table 8 – Pearson's correlation matrix of the measured variables for those convicted actors (N = 17) in the Outgroup

- Message metrics' correlation with betweenness centrality oscillation, betweenness centrality, degree centrality is very significant. This may imply that the Enron convicts were among the most active actors in the social network acting as hubs through which an epidemic could spread.

- Total influence's correlation with messages total and message sent is very significant. This may mean that those convicted were influential and the most active communicators.

Convicted in the Ingroup

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.74**	1.00										
3 Avg sentiment	-0.10	-0.05	1.00									
4 Betweenness centrality	0.91**	0.53**	-0.15	1.00								
5 Degree centrality	0.93**	0.57**	-0.04	0.90**	1.00							
6 Avg complexity	0.12	0.19	0.25	0.11	0.10	1.00						
7 Total influence	0.71**	0.98**	-0.04	0.50**	0.54**	0.18	1.00					
8 Messages received	0.97**	0.65**	-0.16	0.97**	0.92**	0.13	0.61**	1.00				
9 Messages total	0.96**	0.88**	-0.12	0.86**	0.85**	0.17	0.84**	0.94**	1.00			
10 Contribution index	-0.23	0.29	0.13	-0.28	-0.43	0.26	0.30	-0.24	-0.02	1.00		
11 Avg emotionality	0.22	0.22	0.21	0.21	0.21	0.94**	0.18	0.22	0.24	0.19	1.00	
12 Average influence per message	0.06	0.07	0.09	0.03	0.01	-0.01	0.22	0.02	0.05	0.18	-0.19	1.00

p Values *p < .05, **p < .01

Table 9 – Pearson's correlation matrix of the measured variables for those convicted actors (N = 17) in the Ingroup

- Message metrics' correlation with betweenness centrality oscillation, betweenness centrality, degree centrality is very significant. This may imply that the Enron convicts were among the most active actors in the social network acting as hubs through which an epidemic could spread.

- Total influence's correlation with messages total and message sent is very significant. This may mean that those convicted were influential and the most active communicators.

Those convicted - In/Ingroup

	1	2	3	4	5	6	7	8	9	10	11	12
1 Betweenness centrality oscillation	1.00											
2 Messages sent	0.89**	1.00										
3 Avg sentiment	-0.04	0.16	1.00									
4 Betweenness centrality	0.76**	0.59**	-0.11	1.00								
5 Degree centrality	0.76**	0.55**	0.01	0.79**	1.00							
6 Avg complexity	0.29	0.24	0.28	0.34	0.44	1.00						
7 Total influence	0.52	0.59**	0.14	0.63**	0.67**	0.21	1.00					
8 Messages received	0.40	0.11	-0.21	0.61**	0.59**	0.33	0.22	1.00				
9 Messages total	0.91**	0.85**	0.02	0.79**	0.75**	0.37	0.60**	0.61**	1.00			
10 Contribution index	0.23	0.46	0.50**	0.00	0.10	0.56**	0.09	-0.45	0.13	1.00		
11 Avg emotionality	-0.07	-0.13	0.00	-0.20	-0.28	(0.68)**	0.13	-0.19	-0.21	-0.47	1.00	
12 Average influence per message	0.04	0.03	0.16	0.40	0.50	0.23	0.81**	0.30	0.19	-0.20	0.02	1.00

p Values *p < .05, **p < .01

Table 10 – Pearson's correlation matrix of the measured variables for those convicted actors (N = 17) in the In/Ingroup

- Message metrics' correlation with betweenness centrality oscillation, betweenness centrality, degree centrality is very significant. This may imply that the Enron convicts were among the most active actors in the social network acting as hubs through which an epidemic could spread.

- Total influence's correlation with messages total and message sent is very significant. This may mean that those convicted were influential and the most active communicators.

Final thoughts on correlations

From this statistical analysis, I can conclude across samples that:

- Message metrics particularly total messages have a very significant correlation with betweenness centrality, degree centrality and total influence. The messages metrics can reveal an actor tendency to be a hub in Enron's social network through which the corporate culture gets shaped to influence the behaviors and the activities. As such, they can act as a leader or creator of COINs to spread the creativity and the productivity within the organization.

- Total influence in particular correlates significantly with total message. In such case, two meanings can be considered. First, those actors that are hubs are able to shape the culture. Second, this may mean following the herd in a criminal activity in the case of a random actor that is actively interacting with colleagues and upper management

Predicting an Actor's Influence

One way to understand actor's behavior is by analyzing the dependence between some of the correlated social metrics.

t-Test Analysis

I compared the sample group of 17 convicted individuals in the Outgroup (N = 25344) and the Ingroup (N= 140) with a control group for which I selected 17 actors based a comparable number of total messages and then based on a comparable number of total influence. Table 11 below summarizes the test performed:

	17 actors based on a comparable number of total messages	17 actors based on a comparable total influence
17 convicted within Outgroup (N = 25344)	Test I	Test II
17 convicted within Ingroup (N = 140)	Test III	Test IV

Table 11 – Summary of t-tests performed

Test I – Outgroup: 17 convicted vs. 17 comparable by total message

Using t-test of two samples assuming unequal variances, I analyzed the social metrics of the 17 convicted individuals in the Outgroup (N = 25344) with those of an experimental group of 17 individuals with an equivalent number of total messages. Based on the results of the t-test summarized in table 12, we could not reject the Null hypothesis for most of the social metrics at the exception of the contribution index for which the p value is less than 0.05%. In other words, the contribution index t-test refutes the hypothesis that a comparable suspect could become convicted.

$\alpha = 0.05$	Mean	Variance	df	t Stat	P(T<=t) two-tail
Betweenness centrality oscillation	25.67	346.38	23	2.01	0.06
Betweenness centrality oscillation - Comparable	14.33	129.52			
Messages sent	194.73	9.42E+04	26	0.77	0.45
Messages sent - Comparable	117.93	5.69E+04			
Average sentiment	0.54	0.01	27	1.31	0.20
Average sentiment - Comparable	0.50	0.01			
Betweenness centrality	2.04E+06	1.62E+13	14	1.59	0.13
Betweenness centrality - Comparable	3.64E+05	4.99E+11			
Degree centrality	125.07	3.88E+04	16	1.53	0.14
Degree centrality - Comparable	43.47	3.62E+03			
Average complexity	6.03	0.52	27	-0.71	0.49
Average complexity - Comparable	6.23	0.62			
Total Influence	5.28	103.56	26	-0.01	1.00
Total Influence - Comparable	5.30	60.99			
Messages received	250.20	1.35E+05	14	2.03	0.06
Messages received - Comparable	54.40	4.54E+03			
Messages total	368.13	2.20E+05	26	0.77	0.45
Messages total - Comparable	249.13	1.34E+05			
Contribution index	-0.34	0.21	26	-2.57	0.02
Contribution index - Comparable	0.15	0.34			
Average emotionality	0.23	0.001	26	-1.48	0.15
Average emotionality - Comparable	0.24	0.001			
Average influence per message	0.05	0.002	24	-0.37	0.71
Average influence per message - Comparable	0.06	0.004			

Table 12 – T-test with two samples assuming unequal variances – (N = 17) convicted sample vs. comparable total messages sample - Outgroup

Test II – Outgroup: 17 convicted vs. 17 comparable by total influence

I then analyzed the social metrics of the 17 convicted individuals in the Outgroup with those of an experimental sample of 17 individuals with a comparable total influence measure. Based on the results of the t-test summarized in table 13, we could not reject the Null hypothesis for many of the social metrics however not for the betweenness centrality oscillation, the degree centrality, the messages received, the total messages, the average influence per message for which the p value was less than 0.05%. In other words, the t-test for the latter metrics refute the hypothesis that a comparable suspect could become convicted.

α = 0.05	Mean	Variance	df	t Stat	P(T<=t) two-tail
Betweenness centrality oscillation	8.73	74.21	19	-3.20	0.00
Betweenness centrality oscillation - Comparable	25.67	346.38			
Messages sent	30.07	922.50	14	-1.42	0.18
Messages sent - Comparable	117.93	5.69E+04			
Average sentiment	0.52	0.01	27	1.31	0.20
Average sentiment - Comparable	0.54	0.01			
Betweenness centrality	8.68E+04	1.60E+10	14	-1.88	0.08
Betweenness centrality - Comparable	2.04E+06	1.62E+13			
Degree centrality	16.60	151.40	14	-2.13	0.05
Degree centrality - Comparable	125.07	3.88E+04			
Average complexity	6.65	1.23	24	1.81	0.08
Average complexity - Comparable	6.03	0.52			
Total Influence	5.47	102.06	28	0.05	0.96
Total Influence - Comparable	5.28	103.56			
Messages received	34.53	1.11E+03	14	-2.26	0.04
Messages received - Comparable	250.20	1.35E+05			
Messages total	64.60	368.13	14	-2.49	0.03
Messages total - Comparable	3.76E+03	2.20E+05			
Contribution index	-0.04	0.19	27	1.62	0.12
Contribution index - Comparable	-0.29	0.19			
Average emotionality	0.25	0.00	24	1.84	0.08
Average emotionality - Comparable	0.23	0.00			
Average influence per message	0.15	0.02	16	2.52	0.02
Average influence per message - Comparable	0.05	0.00			

Table 13 – T-test with two samples assuming unequal variances – (N = 17) convicted sample vs. comparable total influence sample - Outgroup

Test III – Ingroup: 17 convicted vs. 17 comparable by total message

I then analyzed the social metrics of the 17 convicted individuals in the Ingroup with those of an experimental sample of 17 individuals with an equivalent number of total messages. Based on the results of the t-test summarized in table 14, we could not reject the Null hypothesis for most of the social metrics at the exception of the contribution index for which the p value is less than 0.05%. In other words, the contribution index t-test refute the hypothesis that a comparable suspect could become convicted.

t-Test: Two-Sample Assuming Unequal Variances α 0.05

	Mean	Variance	df	t Stat	P(T<=t) two-tail
Betweenness centrality oscillation	17.87	270.55	23	1.26	0.22
Betweenness centrality oscillation - Comparable	11.53	109.12			
Messages sent	61.53	9.88E+03	27	-0.15	0.88
Messages sent - Comparable	66.67	8.64E+03			
Average sentiment	0.53	0.01	20	1.07	0.30
Average sentiment - Comparable	0.50	0.00			
Betweenness centrality	99.59	3.44E+04	15	1.28	0.22
Betweenness centrality - Comparable	36.50	1.92E+03			
Degree centrality	18.93	179.21	19	2.04	0.06
Degree centrality - Comparable	11.13	39.98			
Average Complexity	6.03	0.91	25	2.12	0.04
Average Complexity - Comparable	5.13	1.79			
Total Influence	2.96	2.90	26	0.03	0.97
Total Influence - Comparable	31.42	18.16			
Messages received	103.07	1.82E+04	17	1.47	0.16
Messages received - Comparable	48.87	2.08E+03			
Messages total	164.60	4.54E+04	23	0.76	0.46
Messages total - Comparable	115.53	1.78E+04			
Contribution index	-0.28	0.11	27	-2.80	0.01
Contribution index - Comparable	0.06	0.11			
Average emotionality	0.23	0.00	24	1.31	0.20
Average emotionality - Comparable	0.21	0.00			
Average influence per message	0.06	0.01	18	0.89	0.38
Average influence per message - Comparable	0.04	0.00			

Table 14 – T-test with two samples assuming unequal variances – (N = 17) convicted sample vs. comparable sample in the Ingroup

Test IV – Ingroup: 17 convicted vs. 17 comparable by total influence

Finally, I analyzed the social metrics of the 17 convicted individuals in the Ingroup with those of experimental sample of 17 individuals with a comparable total influence measure. Based on the results of the t-test summarized in table 15, we could not reject the Null hypothesis for many of the social metrics however not for the betweenness centrality oscillation, the degree centrality, the messages received, the total messages, the average influence per message for which the p value was less than 0.05%. In other words, the t-test for the latter metrics refute the hypothesis that a comparable suspect could become convicted.

α = 0.05	Mean	Variance	df	t Stat	P(T<=t) two-tail
Betweenness centrality oscillation	8.73	74.21	19	-3.20	0.00
Betweenness centrality oscillation - Comparable	25.67	346.38			
Messages sent	30.07	922.50	14	-1.42	0.18
Messages sent - Comparable	117.93	5.69E+04			
Average sentiment	0.52	0.01	27	1.31	0.20
Average sentiment - Comparable	0.54	0.01			
Betweenness centrality	8.68E+04	1.60E+10	14	-1.88	0.08
Betweenness centrality - Comparable	2.04E+06	1.62E+13			
Degree centrality	16.60	151.40	14	-2.13	0.05
Degree centrality - Comparable	125.07	3.88E+04			
Average complexity	6.65	1.23	24	1.81	0.08
Average complexity - Comparable	6.03	0.52			
Total Influence	5.47	102.06	28	0.05	0.96
Total Influence - Comparable	5.28	103.56			
Messages received	34.53	1.11E+03	14	-2.26	0.04
Messages received - Comparable	250.20	1.35E+05			
Messages total	64.60	368.13	14	-2.49	0.03
Messages total - Comparable	3.76E+03	2.20E+05			
Contribution index	-0.04	0.19	27	1.62	0.12
Contribution index - Comparable	-0.29	0.19			
Average emotionality	0.25	0.00	24	1.84	0.08
Average emotionality - Comparable	0.23	0.00			
Average influence per message	0.15	0.02	16	2.52	0.02
Average influence per message - Comparable	0.05	0.00			

Table 15 – T-test with two samples assuming unequal variances – (N = 17) convicted sample vs. comparable total influence sample - Ingroup

Final thoughts on t-tests

For both the Outgroup and the Ingroup, the analysis of the metrics of 17 convicted employees was tested against a comparable group of 17 employees. The comparable criteria was an equivalent number of total messages in one case and an equivalent total influence in another case. Identifying criminals based on email behaviors is possible depending on the sampling strategy. When sampling based on employees with comparable total emails, the t-test results of Contribution Index (Ci) revealed for both the Outgroup and the Ingroup that criminals were less active. When selecting a sample based on employees with comparable total influence, the t-test results of Betweenness Centrality Oscillation (Bco) and Degree Centrality (Bc) revealed for both Outgroup and Ingroup that criminals were less connected to others and less creative.

The table below summarizes the test results obtained for these indicators:

	17 actors based on a comparable number of total messages	**17 actors based on a comparable total influence**
17 convicted within Outgroup (N = 25344)	Test I Bco (p = .06), DC (p = .14) and CI (p = .02)	Test II Bco (p = .00), DC (p = .05) and CI (p = .12)
17 convicted within Ingroup (N = 140)	Test III Bco (p = .02), DC (p = .06) and CI (p = .01)	Test IV Bco (p = .00), DC (p = .05) and CI (p = .12)

Table 16 – Summary of insights on t-test results

Chapter 6: Discussion

In this analysis, I aimed to detect communication patterns associated with unlawful activities inside Enron and the role this organization had on the California Energy Crisis. I set forth the following hypotheses to be validated through the social network analysis of the Enron email archive between 2000 and 2001.

- **Hypothesis I:** The most active people are likely those convicted.
- **Hypothesis II:** Convicted individuals have high total influence in the social network.
- **Hypothesis III:** The more the communication patterns oscillate between centralized and rotating leadership, the more distributed the creativity, in this case the fraudulent activity.
- **Hypothesis IV:** The communication pattern signals a role in the California Energy Crisis.

Before discussing the results, I would like to remind the four steps of the Knowledge Flow Optimization (Coolfarming, 2010) framework. In chapter 3, I described this framework that can be track new COIN formations and nurture existing ones in an organization. This framework can also be used to improve productivity and to nurture creativity inside an organization. In this chapter, I will apply this framework to make recommendations on how we can track creative and illicit activity inside Enron. The four steps of this framework are:

1. **Discover:** collect data and/or analyze a dataset. In this case, it is the Enron email archive.
2. **Measure:** compute and analyze various social network metrics to quantify the communications patterns and determine which ones are correlated with success. In this case, analyzing what
3. **Optimize:** improve the social network structure and communication patterns that were identifying when measuring the social signals in the previous step.
4. **Mirror:** share the results of the analysis with the members of the social network, which would help them become more aware of their communication patterns and improve them.

Having completed the analysis of the social network of Enron's email archive of the Knowledge Flow Optimization framework. I have also performed correlation, t-test and regression analysis using the social network metrics as the explanatory variables. Several significant correlations have been observed. The next step step is to interpret these correlations and relate them back to communication patterns and suspicious activities at Enron.

Discussion on correlations

Several significant correlations were observed in the social network analysis of Enron's email archive. The following are the key take-aways from these correlations:

Being the most active does not mean engaging in illicit activity

Many of the convicted group of actors were not present in the sample (N = 140) of the most active actors. It was worth noting that the correlating of total influence and total messages were significant. However, total influence also depends on other variables including betweenness centrality oscillation, degree centrality and betweenness centrality. The combination of total messages and betweenness centrality oscillation provided a satisfactory regression models to predict total influence of an actor. These findings do not support Hypothesis I that the most active are the most likely to be convicted.

Those convicted are not the most influential

When analyzing the social metrics of the convicted actors in the various samples, those convicted did not always have the highest total influence. Some of the gatekeepers including executive assistants and chiefs of staff can be quite influential. Furthermore, email is only one form of communication and may not represent the full range of communication including one-on-one and telephone. Email communication does however provide signals as to the communication patterns. These were clearly identified in the correlations between the structural metrics and the volume of email messages. However, these findings do not support Hypothesis II that those convicted are the most influential.

The more oscillation the more creative and the more influential

It was observed throughout the samples that total influence was significantly correlated with the total messages and the structural social metrics. The regression analysis across samples revealed that total influence depend on both the betweenness centrality oscillation and the total messages. Hypothesis III: The more the communication patterns oscillate between centralized and rotating leadership, the more distributed the creativity, in this case the fraudulent activity.

Having interpreted the correlations found during the "measure" step of the Knowledge Flow Optimization framework, the next step is to "optimize" communication patterns to improve collaboration.

Discussion on t-test analysis

For both the Outgroup and the Ingroup, the analysis of the metrics of 17 convicted employees was tested against a comparable group of 17 employees. The comparable criteria was an equivalent number of total messages in one case and an equivalent total influence in another case. Identifying criminals based on email behaviors is possible depending on the sampling strategy. When sampling based on employees with comparable total emails, the t-test results of Contribution Index (Ci) revealed for both the Outgroup and the Ingroup that criminals were less active. When selecting a sample based on employees with comparable total influence, the t-test results of Betweenness Centrality Oscillation (Bco) and Degree Centrality (Bc) revealed for both Outgroup and Ingroup that criminals were less connected to others and less creative.

Recommendation

Based on these findings, I would recommend the following to that organization if it still existed:

- Use social networks more broadly as an effective method for effective innovation provided these activities do not lead to unlawful creativity or financial bankruptcy.
- Use a social network analysis tool like Condor to generate social signal metrics to nurture productivity in more creative groups within the organization and to monitor performance in more operational groups. As part of monitoring performance, risk managers could also track irregular activities to prevent a bankruptcy or to limit its impact if external forces were contributing.

The final step in the Knowledge Flow Optimization process is mirroring the results back to the actors in the social network. As a first step toward it, I presented to Dr. Gloor the honest signals of my interactions with other team members in a project context to learn to measure and to harness the collective intelligence within a COIN.

In general, the hands-on journey of this project enabled me to learn about the social dynamic of a COIN and to distinguish between a creative and an operational context. In a creative context, I notice that leadership is rotating, contribution is balanced among team members and emotions are fluctuating since the day-to-day activity is more random. In an operational context, I notice that the structure of a team and the contribution among its members are less distributed. In addition, emotions are more balanced since the day-to-day is more deterministic.

Chapter 7: Conclusion

I aimed to detect communication patterns associated with illicit activities inside a global organization and the role this organization had on the California Energy Crisis. From the previous two chapters, the social network analysis using Condor and SPSS revealed several communication patterns that did not link suspect actors with criminal activity.

First, the most active actors were not necessarily engaged illicit activity. Most of the convicted group of actors were not present in the sample (N = 140) of the most active actors. These findings do not support Hypothesis I that the most active are the most likely to be convicted.

Second, the most influential actors are not involved in illicit activity. When analyzing the social metrics of the convicted actors in the various samples, those convicted did not always have the highest total influence. Some of the gatekeepers including executive assistants and chiefs of staff were the most influential. Furthermore, email is only one form of communication and may not represent the full range of communication including one-on-one and telephone. Email communication does however provide signals as to the communication patterns. These were clearly identified in the correlations between the structural metrics and the volume of email messages. However, these findings do not support Hypothesis II that those convicted are the most influential.

Third, the more oscillation the more creative and the more influential actors would be. It was observed throughout the samples that total influence was significantly correlated with the total messages and the structural social metrics. This supports Hypothesis III that the more the communication patterns oscillate between centralized and rotating leadership, the more distributed the creativity, in this case the fraudulent activity.

Finally, the link between Enron and California Energy Crisis could be observed only through qualitative observation in two ways. First, in the evolution of the California Energy crisis Enron collapse begins when California energy prices normalize. Enron artificially created power shortages in the daily spot markets to make short-term gains that were maximized at the peak of the crisis. Second, in the Condor the email activity data in the archive appear to follow a similar pattern as the stock price movement in the last two years. This would be an interesting future analysis.

Limitation of the Analysis

One of the limitations of this analysis was that the sample dataset did not include a comprehensive email history of one third of those convicted. There were some data integrity issues that were resolved over the course of research projects. However, additional email history was not available. Another limitation is that email is not the only form of communication inside an organization. As we have witnessed in this case, the Enron Tapes supported some of the evidence that authorities were gathering at the time of this investigation. Finally, it remains a challenge to track an actor's intent over the course of their communication patterns particularly when illicit activities are involved.

Future Work

When we look at the performance of Enron's stock price, we observe a similar pattern in the email activity that went on over the course of the last two years of the organization's existence. This would an interesting correlation to explore.

Another area to explore would to study the California Energy Crisis from the standpoint of the State and the Federal regulators to see what special interests may have influenced the policies during this restructuring process. In addition, it would be interesting to explore methods and tools that could drastically improve the crisis response time hence saving the stakeholders billions of dollars in the process.

Finally, it would be to investigate the collective intelligence between energy companies and State and Federal legislators in an era of climate change (NASA, 2015). Last December, a New York Times social network analysis has determined that there is a hidden coalition between a majority of Republican District Attorneys and energy companies to push back on Obama's regulatory agenda (Lipton, 2014).

THIS PAGE IS INTENTIONALLY LEFT BLANK

Bibliography

Apicella, C. L. a. F. W. M. J. H. F. N. A. C., 2012. Social networks and cooperation in hunter-gatherers society. *Nature,* p. 497–501.

Barabási, A.-L. A. R. a. J. H., 1999. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications,* p. 69–77.

Battilana, J. a. T. C., 2012. Change Agents, Networks and Institutions. *Academy of Management Journal 55, no. 2.*

Brynjolfsson, E., 2000. Frictionless Commerce: A Comparison of Internet and Conventional Retailers. *Management Science,* p. 563–585.

Brynjolfsson, E., 2011. Goodbye Pareto Principle, Hello Long Tail. *Management Science,* p. 1373–1386.

Bushnell, J., 2004. California's electricity crisis: a market apart?. *Energy Policy,* p. 1045–1052.

CBO, C. B. O., 2001. *http://www.cbo.gov/sites/default/files/californiaenergy.pdf.* [Online].

Centola, D., 2011. The Spread of Behavior in an Online Social Network Experiment. *Science* , pp. 909-912.

Christakis, N. A. a. J. H. F., 2008. Dynamic spread of happiness in a large social network. *BMJ.*

Davis, J. H. L. a. M. S. H., 2007. *Social Network Analysis and Organizational Disintegration: The Case of Enron Corporation.* Sydney, International Conference on Information Systems.

Diesner, J. a. C. K. M., 2005. *Exploration of Communication Networks from the Enron Email Corpuse.* Queen's University, Workshop on Link Analysis, Counterterrorism and Security.

Dubar, R. A. H. a. R., 2002. Social Network Size in Humans. *Human Nature, Vol. 14, No. 1,* pp. pp. 53-72.

Easley, D. a. K. J., 2010. *Networks, Crowds, and Markets.* New York, NY: Cambridge University Press.

EIA, 2001. *http://www.eia.gov/electricity/policies/legislation/california/subsequentevents.html.* [Online] [Accessed 2015].

Eichenwald, K., 2003. *http://www.nytimes.com/2003/05/02/business/us-indicts-11-former-enron-executives.html.* [Online].

Euler, L., 1736. *Seven Bridges of Konigsberg,* s.l.: s.n.

FERC, 2001. *https://www.ferc.gov/industries/electric/indus-act/wec/chron/chronology.pdf.* [Online].

Freeman, L. C., 2004. *The development of social network analysis.* Vancouver, B.C.: Empirical Press.

Garton, C. H. B. W. a. L., 1997. Studying Online Social Networks. *Computer-Mediated Communication.*

Gloor, P. A. ,. O. D. R. O. ,. P. A. a. S. D., 2009. The Virtual Mirror - Reflecting on your social and psychological self to increase organizational creativity. *International Studies of Management and Organization.*

Gloor, P. A., 2006. *Swarm Creativity.* New York, NY: Oxford University Press.

Gloor, P. A., 2007. *Coolhunting.* New York, NY: AMACOM.

Gloor, P. A., 2011a. *Coolfarming.* New York, NY: AMACOM.

Gloor, P. A., 2011b. Coolfarming: discovering the surprising power of social networks. *Ernst & Young Performance Journal.*

Gloor, P. A., 2014. *Condor Software - Courtesy licence to conduct this project.* Cambridge: Galaxy Advisors.

Gloor, P. A. a. Z. Y., 2004. *TeCFlow – A Temporal Communication Flow Visualizer.* Cambridge, MA, ICKN.

Gloor, P. A. M. P. S. M. a. D. G., 2012. *Coolfarming - Lessons from the beehive to increase organizational creativity,* Cambridge, MA: ICKN technical report.

Gloor, P. K. J. N. S. F. K. a. S. D., 2009. *Web Science 2.0: Identifying Trends through Semantic Social Network Analysis.* s.l., International Conference on Computational Science and Engineering.

Gloor, P. Z. Y., Nov. 6. 2004. *TeCFlow - A Temporal Communication Flow Visualizer for Social Networks Analysis.* Chicago, ACM CSCW Workshop on Social Networks. ACM CSCW Conference.

Kidane, Y. H. a. G. P., 2007. Correlating temporal communication patterns of the Eclipse open source community with performance and creativity. *Computational and Mathematical Organization Theory,* pp. 17-27.

Klimt, B. a. Y. Y., 2004. *Introducing the Enron Corpus.* s.l., CEAS conference.

Levy, P., 1987. Collective intelligence mankind's emerging world in cyberspace.

Lipton, E., 2014. *http://www.nytimes.com/2014/12/07/us/politics/energy-firms-in-secretive-alliance-with-attorneys-general.html?_r=0.* [Online]
[Accessed 2015].

Mandelbrot, B., 1953. An informational theory of the statistical structure of language. *Communication Theory,* pp. 486-501.

Michalski, R. P. S. a. K. P., 2011. *Matching Organizational Structure and Social Network Extracted from Email Communication.* Poznań, Poland, 14th International Conference, Business Information Systems.

NASA, 2015. *http://climate.nasa.gov/evidence/.* [Online].

NYTimes, 2003. *http://www.nytimes.com/ref/us/20061023_ENRON_GRAPHIC.html.* [Online].

Price, D. J. d. S., 1965. Networks of Scientific Papers. *Science,* pp. 510-515.

Top500, 2014. *https://www.linux.com/news/enterprise/high-performance/147-high-performance/778179--linux-nears-total-domination-of-the-top500-supercomputers.* [Online].

Von Hippel, E., 2005. *Democratizing Innovation.* Cambridge: MIT Press.

W3Techs, 2014. *http://w3techs.com/technologies/details/os-unix/all/all.* [Online].

Wasserman, K. F. a. S., 1994. *Social Network Analysis - Methods and Applications.* New York, NY: Cambridge University Press.

Watts, S. H. S. a. D. J., 1998. Collective dynamics of 'small-world' networks. *Nature,* pp. 440-442.

Wellman, B., 2001. Computer Networks As Social Networks. *Computers and Science,* pp. 2031-2034.

Wikipedia, 2015. *Network Science.* [Online].

Appendix

Electricity Strategy Game Reference Data

Daily and Hourly Wholesale Electricity Data Provided

		Hour 1	Hour 2	Hour 3	Hour 4
Day 1		10800	13500	18000	16000
Day 2		11000	15000	19500	15000
Day 3		10800	13500	18000	16000
Day 4		11000	15000	19500	15000
Day 5		10800	13500	18000	16000
Day 6		11000	15000	19500	15000

Table 17 – Daily and Hourly Wholesale Electricity Data Provided

Portfolio Data

	Capacity MW	Fuel Cost $/MWH	Var O&M $/MWH	Total Var Cost $/MWH	CO2 Emissions lbs./MWH	O&M/Day ($)
(=Bay_Views)						
Ostrom Portfolio						
MORRO BAY 1&2	335	26.50	0.50	27.00	1200	$2,000
MORRO BAY 3&4	665	25.00	0.50	25.50	1200	$4,000
MOSS LANDING 6	750	21.50	1.50	23.00	1500	$8,000
MOSS LANDING 7	750	21.50	1.50	23.00	1500	$8,000
OAKLAND	150	42.00	0.50	42.50	2150	$0
Totals	***2650***					***$22,000***
(=Big_Coal)						
Modigliani Portfolio						
FOUR CORNERS	1900	17.50	1.50	19.00	2600	$8,000
ALAMITOS 7	250	50.00	1.50	51.50	2000	$0
HUNTINGTON BEACH 1&2	300	27.00	1.50	28.50	1200	$2,000

HUNTINGTON BEACH 5	150	45.00	1.50	46.50	2000	$2,000
REDONDO 5&6	350	28.00	1.50	29.50	1200	$3,000
REDONDO 7&8	950	28.00	1.50	29.50	1200	$5,000
Totals	***3900***					***$20,000***
(=**Fossil_Light)**						
Vickrey Portfolio						
HUMBOLDT	150	32.50	0.50	33.00	2000	$0
HELMS	800	0.00	0.50	0.50	0	$15,000
HUNTERS POINT 1&2	150	33.00	1.50	34.50	2000	$1,000
HUNTERS POINT 4	250	51.50	1.50	53.00	2000	$1,000
DIABLO CANYON 1	1000	7.50	4.00	11.50	0	$20,000
Totals	***2350***					***$37,000***
(=Beachfront)						
Phelps Portfolio						
COOLWATER	650	29.00	0.50	29.50	1200	$2,000
ETIWANDA 1-4	850	28.50	1.50	30.00	1200	$8,000
ETIWANDA 5	150	42.50	1.50	44.00	1500	$1,000
ELLWOOD	300	52.00	0.50	52.50	2000	$0
MANDALAY 1&2	300	26.00	1.50	27.50	1100	$1,000
MANDALAY 3	150	35.00	1.50	36.50	1400	$1,000
ORMOND BEACH 1	700	26.00	0.50	26.50	1100	$7,000
ORMOND BEACH 2	700	26.00	0.50	26.50	1100	$7,000
Totals	***3800***					***$27,000***
(=Old_Timers)						
Tobin Portfolio						
BIG CREEK	1000	0.00	0.00	0.00	0	$15,000
MOHAVE 1	750	15.00	4.50	19.50	2600	$15,000
MOHAVE 2	750	15.00	4.50	19.50	2600	$15,000
HIGHGROVE	150	34.00	0.50	34.50	1400	$0
SAN BERNADINO	100	37.00	0.50	37.50	1400	$0
Totals	***2750***					***$45,000***
(=Big_Gas)						
Nash Portfolio						
EL SEGUNDO 1&2	400	30.00	1.50	31.50	1400	$1,000
EL SEGUNDO 3&4	650	27.50	1.50	29.00	1300	$1,000
LONG BEACH	550	36.00	0.50	36.50	1400	$2,000
NORTH ISLAND	150	45.00	0.50	45.50	2000	$0
ENCINA	950	28.50	0.50	29.00	1300	$2,000
KEARNY	200	62.00	0.50	62.50	2000	$0
SOUTH BAY	700	30.00	0.50	30.50	1400	$2,000

Totals	***3600***					***$8,000***
(=East_Bay)						
Samuelson Portfolio						
PITTSBURGH 1-4	650	28.00	0.50	28.50	1200	$2,500
PITTSBURGH 5&6	650	25.00	0.50	25.50	1200	$2,500
PITTSBURGH 7	700	41.00	0.50	41.50	1500	$4,000
CONTRA COSTA 4&5	150	40.00	0.50	40.50	2000	$1,000
CONTRA COSTA 6&7	700	27.00	0.50	27.50	1100	$6,000
POTRERO HILL	150	48.00	0.50	48.50	2000	$0
Totals	***3000***					***$16,000***

Table 18 – Electricity Portfolio Data Provided

Table showing the data that was required to plot a MC step chart for round 2

	Capacity	Fuel Cost	Var O&M	Total Var Cost	CO2 Emissions	O&M/Day ($)	Cumulative Supply	Difference in MC
	MW	$/MWH	$/MWH	$/MWH	lbs./MWH			
BIG CREEK	1000	0.00	0.00	0	0	15000.00	25.00	1000
HELMS	800	0.00	0.50	0.5	0	15000.00	13.00	1800
DIABLO CANYON 1	1000	7.50	4.00	11.5	0	20000.00	16.00	2800
ORMOND BEACH 1	700	26.00	0.50	78.88095238	1100	7000.00	23.00	3500
ORMOND BEACH 2	700	26.00	0.50	78.88095238	1100	7000.00	24.00	4200
MANDALAY 1&2	300	26.00	1.50	79.88095238	1100	1000.00	21.00	4500
CONTRA COSTA 6&7	700	27.00	0.50	79.88095238	1100	6000.00	41.00	5200
MORRO BAY 3&4	665	25.00	0.50	82.64285714	1200	4000.00	2.00	5865
PITTSBURGH 5&6	650	25.00	0.50	82.64285714	1200	2500.00	38.00	6515
MORRO BAY 1&2	335	26.50	0.50	84.14285714	1200	2000.00	1.00	6850
HUNTINGTON BEACH 1&2	300	27.00	1.50	85.64285714	1200	2000.00	8.00	7150
PITTSBURGH 1-4	650	28.00	0.50	85.64285714	1200	2500.00	37.00	7800
REDONDO 5&6	350	28.00	1.50	86.64285714	1200	3000.00	10.00	8150
REDONDO 7&8	950	28.00	1.50	86.64285714	1200	5000.00	11.00	9100
COOLWATER	650	29.00	0.50	86.64285714	1200	2000.00	17.00	9750
ETIWANDA 1-4	850	28.50	1.50	87.14285714	1200	8000.00	18.00	10600
EL SEGUNDO 3&4	650	27.50	1.50	90.9047619	1300	1000.00	31.00	11250
ENCINA	950	28.50	0.50	90.9047619	1300	2000.00	34.00	12200
MOSS LANDING 6	750	21.50	1.50	94.42857143	1500	8000.00	3.00	12950
MOSS LANDING 7	750	21.50	1.50	94.42857143	1500	8000.00	4.00	13700
SOUTH BAY	700	30.00	0.50	97.16666667	1400	2000.00	36.00	14400
EL SEGUNDO 1&2	400	30.00	1.50	98.16666667	1400	1000.00	30.00	14800
HIGHGROVE	150	34.00	0.50	101.1666667	1400	0.00	28.00	14950

MANDALAY 3	150	35.00	1.50	103.1666667	1400	1000.00	22.00	15100
LONG BEACH	550	36.00	0.50	103.1666667	1400	2000.00	32.00	15650
SAN BERNADINO	100	37.00	0.50	104.1666667	1400	0.00	29.00	15750
PITTSBURGH 7	700	41.00	0.50	112.9285714	1500	4000.00	39.00	16450
ETIWANDA 5	150	42.50	1.50	115.4285714	1500	1000.00	19.00	16600
HUMBOLDT	150	32.50	0.50	128.2380952	2000	0.00	12.00	16750
HUNTERS POINT 1&2	150	33.00	1.50	129.7380952	2000	1000.00	14.00	16900
CONTRA COSTA 4&5	150	40.00	0.50	135.7380952	2000	1000.00	40.00	17050
NORTH ISLAND	150	45.00	0.50	140.7380952	2000	0.00	33.00	17200
HUNTINGTON BEACH 5	150	45.00	1.50	141.7380952	2000	2000.00	9.00	17350
FOUR CORNERS	1900	17.50	1.50	142.8095238	2600	8000.00	6.00	19250
MOHAVE 1	750	15.00	4.50	143.3095238	2600	15000.00	26.00	20000
MOHAVE 2	750	15.00	4.50	143.3095238	2600	15000.00	27.00	20750
POTRERO HILL	150	48.00	0.50	143.7380952	2000	0.00	42.00	20900
OAKLAND	150	42.00	0.50	144.8809524	2150	0.00	5.00	21050
ALAMITOS 7	250	50.00	1.50	146.7380952	2000	0.00	7.00	21300
ELLWOOD	300	52.00	0.50	147.7380952	2000	0.00	20.00	21600
HUNTERS POINT 4	250	51.50	1.50	148.2380952	2000	1000.00	15.00	21850
KEARNY	200	62.00	0.50	157.7380952	2000	0.00	35.00	22050

Table 19 – Data to plot a MC step chart for round 2

Experiences during Round 1 of the ESG

First, all participant teams had to bid for generation portfolios and then use it to compete in a sequence of daily electricity spot markets. During the first portfolio auction we managed to buy the Modigliani Portfolio which has the highest base load capacity of all portfolios in the market. Due to its high capacity this portfolio is very well suited for exercising market power during high demand hours. The idea of market power in economics and particularly in oligopolistic industrial market structure is the ability of a firm to profitably raise the market price of a good or service over marginal cost (MC).

Our bidding strategy was to make use of our market power whenever this was feasible and to act as if there was perfect competition in all other hours. We first had to find out in which hours of the three days it would make sense to use our market power. Therefore, we have to look at what market power means in this simulation game.

We can make use of market power whenever the residual demand for our portfolio is higher than zero. In these hours our plants are required to fulfill the demand for electricity and we can set the price for electricity as high as we want because of the inelastic demand function. The residual demand for our portfolio is determined by calculating the difference between demand and rest of the world (RoW) supply/capacity. So, whenever demand exceeds 22050-3900 = 18150 we can utilize our market power. A different way to look at it is to analyze the demand supply graph in figure 22.

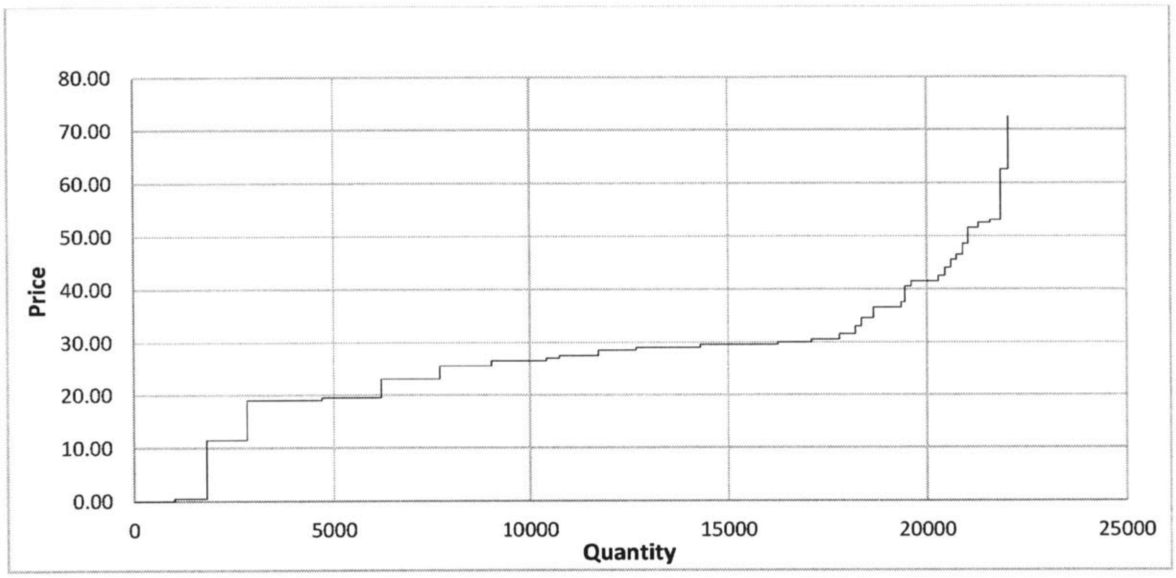

Figure 22 - MC Step Chart for round 1

Figure 22 includes the ROW-MC curves and the demand curves for each mean demand level. If there is no cross between the rest-of-the-world marginal cost curve and the demand curve for a given hour, then this means that we can set the price. Based on this graph we can assume that in day 2 hour 3 there is definitely the potential for us to use our market power to our advantage. But even on days 1 and 3 in hour 3 it might be a good idea to speculate for market power to happen. Why is a demand of more than 18150 in these hours even possible? This is due to the fact that the demand data is not 100% certain of what the future demand will be. So, the demand in each given hour is a random variable that has a normal distribution with a mean equal to the forecast and a standard deviation equal to 3% of the forecast. This means that there is a certain chance that we can use our market power during these two hours as well if the demand is more than 18150.

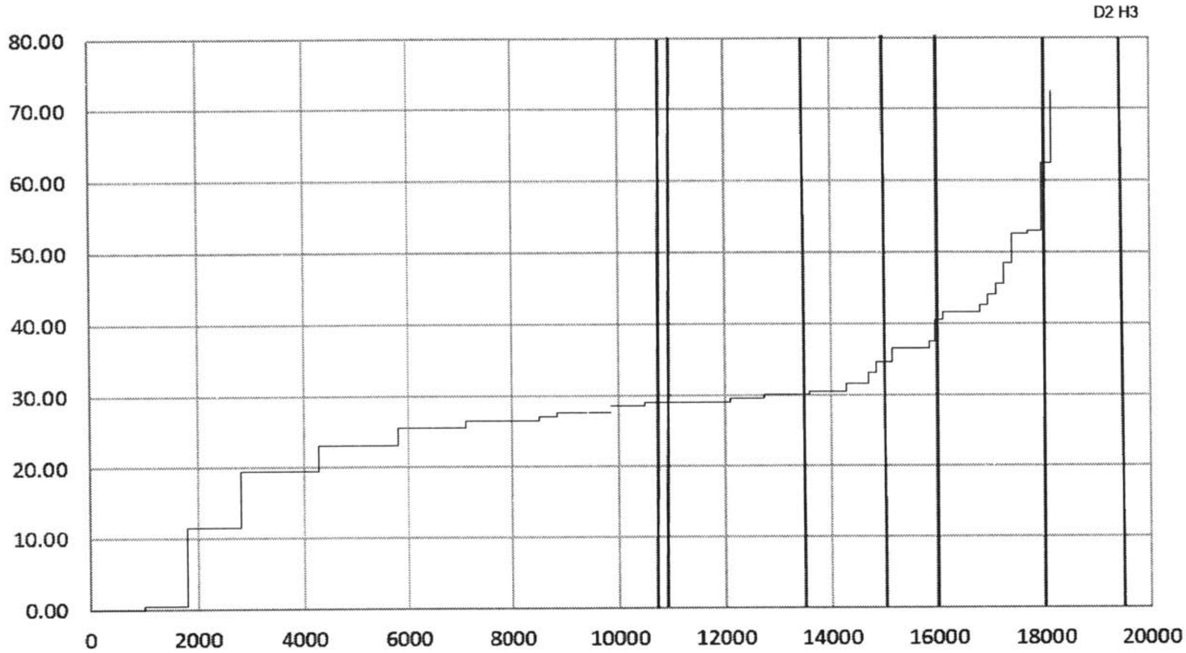

Figure 23 - Supply and Demand Chart for ROW Supply in Part 1 of the game

Let's now calculate if we want to take this risk of bidding high during these hours. If demand does not exceed 18150 and we bid around \$500 for every one of our plants, this means that the rest of the world will simply satisfy demand and we will produce 0, which results in a loss of \$20,000 for us due to fixed O&M costs. So, for bidding high to make sense the potential loss has to be more than offset by the potential gain from using market power. We can calculate the expected profit with the following equation: $E_{profit} = \int_{-inf}^{inf} p(Q) * Profit(Q)dQ$

P(Q) in this formula is the probability for the demand to reach the quantity Q and Profit(Q) is the potential profit for quantity Q in the case that every competitor bids their respective marginal cost curves. $P(Q<18150) = 0.75$ which means that there is a $P(18150)^2 = 55.88\%$ chance of us losing money. Therefore, we would only want to go down the route of bidding about \$500 if the expected return of doing so is very high. When calculating the above equation with Matlab and using a step profit function, we get an expected profit of $5.15*10^4$. The mean expected profit, which was calculated in the same way as described in the first memo, for the case in which we bid marginal costs is $2.22*10^4$. Because the expected profit for the risky case is not significantly higher, we did not want to take the risk of bidding about \$500 in period 3 of days 1 and 3. Instead we decided to only use our market power on day 2 as here the probability of making huge profits is a lot higher (~85%).

For the peak hour in period 2 we had to think about what the other competitors who might want to make use of their market power will probably do. We knew that we had paid more, relatively speaking, for our portfolio than the group that managed to acquire Beachfront. This group paid less than the value we calculated for this portfolio assuming everyone bid marginal costs and no use of market power. Unfortunately we paid more than the value for this basic case but still less than the value this portfolio had when market power was used by the company operating it. So, we had a higher incentive of getting the highest possible price during this peak hour than the competing group. This meant that we could not afford to bid a lot lower than $500 in an attempt to underbid this group. We simply wanted to make sure that we would underbid a competitor that simply entered a bid of $500 for its most expensive plant and 1ct less for each less expensive plant. So, we bid $499.93 for the most expensive plant and 499.89 for the least expensive one. Finally, we had exactly the amount of market power that we expected us to have as we effectively set the price to $499.89 for this hour.

After period 2 Professor Knittel announced that it would make sense to bid high in the peak hour of period 3 as well. Knowing that we thought that one of our competitors in the market would potentially try to bid high and make use of their market power. In order to increase the chances of them being successful in setting a high market price, we decided to enter bids of $500 for our two highest MC plants which would not contribute a lot of profit anyways. These plants, however, were useful to increase the probability of our competitors to have market power because entering bids of $500 had a similar effect as simply taking these plants out of the market. This strategy was successful as the demand in the last peak hour of the first round would not have been high enough to give one of our competitors market power if we hadn't entered such high bids for the two plants. In the end, our low MC plants were producing at full capacity at a very high price at the expense of our competitor that set the price for us.

Strategy and Experiences in Round 2 of the ESG

The second round started with the auction of the generation portfolios in a first price sealed bid auction. This means that the highest bidder wins and pays the exact price that was given by her/his bid. After a participant's bid was accepted all of her/his other bids are deleted and don't affect the auction anymore.

So, we have to come up with the value of the individual portfolios first. One way to look at the problem of deciding on a portfolio value is to calculate the Net Present Value (NPV) of profits that we can get from a certain generation portfolio. The NPV of an investment equals the sum of the present values of its cash flows. This means that all future expected cash flows are discounted at the specific interest rate and then summed up. So, we can write down the formula for the NPV of an investment in the following way:

$$\mathrm{NPV}(i, N) = \sum_{t=0}^{N} \frac{R_t}{(1+i)^t}$$

$i = interest\ rate$

$R_t = net\ cash\ flow\ for\ period\ t$

$N = total\ number\ of\ periods$

Setting the valuation/bid higher than the NPV would mean that we are willing to accept losses, which would not be rational. What if we bid less than the NPV? In this case the chances might be smaller to win, but the payoff is also higher. So, there is in fact an optimal strategy for the auction according to the Nash Equilibrium which states that the ideal bid is equal to:

$$b = \frac{n-1}{n} V; b = bid; n = \#of\ bidders$$

Only looking at the NPV and using this value as the value V in the auction, however, assumes that we only have to decide between owning no portfolio and owning one at the reservation price. Unfortunately, this would be too simple for this game. This is due to the fact, that we don't have the choice to simply not buy a portfolio. If we fail to acquire one of the first six portfolios during the auction, we will inevitably be assigned the last portfolio at a fixed price of $40,000. So, we have to compare the value of each portfolio to this last portfolio in order to decide how much more or less than $40,000 this portfolio would be worth.

As we want to use the NPV of each portfolio to determine our reservation value, we use the same approach as the one that was described in the previous memo. Every plant potentially contributes cash flows in each period. These can be calculated in the following way:

$$hourly\ cash\ flow_{plant_i} = (Clearing\ Price - MC) * Quantity_{plant_i} - hourly\ fixed\ cost$$

$$hourly\ fixed\ cost = fixed\ O\&M\ cost\ per\ day/\ 4\ hours\ per\ day$$

$$MC = marginal\ fuel\ cost + variable\ O\&M\ cost + Opportunity\ cost$$

$$Opportunity\ cost = \frac{carbon\ emissions\ in\ \frac{lbs.}{MWH}}{2000 \frac{lbs}{t}} * NPV(\$100\ after\ day\ 6)$$

Each portfolio is given a number of free permits that add value in form of permit price multiplied by number of permits to each portfolio. These permits can either be sold or used for producing electricity. Therefore, there is an opportunity cost associated with using these permits. In order to determine the associated opportunity cost and added value caused by these permits, we have to know the value of them. After the last period every permit that exceeds the carbon cap costs $100. Because of that we assumed that every group in need of permits would be willing to pay up to the net present value of $100 for the permits. This only holds if there are less free permits than the world actually needs. This was the case because the carbon cap was based on the amount of carbon emissions during the first round (90% of these). During this first round the overall demand was lower than the one which was expected for round 2. Because of that, the plants would supply more electricity overall which would lead to a similar amount of carbon emissions. So, some companies would necessarily exceed their permits. So, we calculated the value of each portfolio as the sum of the NPV of the hourly cash flows and the NPV of the free permits of each portfolio (number of permits * NPV[$100 after day 6]).

So, it all comes down to the clearing price and the quantity that each plant is expected to achieve. The word expected here is very important as we saw during the first portfolio auction that only our group and one other group had factored market power into their portfolio valuations. During this first price sealed bid auction we therefore wanted to adapt our valuation ideas to those of the other groups in order to avoid paying far more than the average group in our world. So, based on the previous bids during the first round, it was fair to assume that the Nash Equilibrium for Bertrand competition should dictate the value of each portfolio. This means that the market clearing price equals MC. If that is the case, the supply curve looks like the one that can be found below.

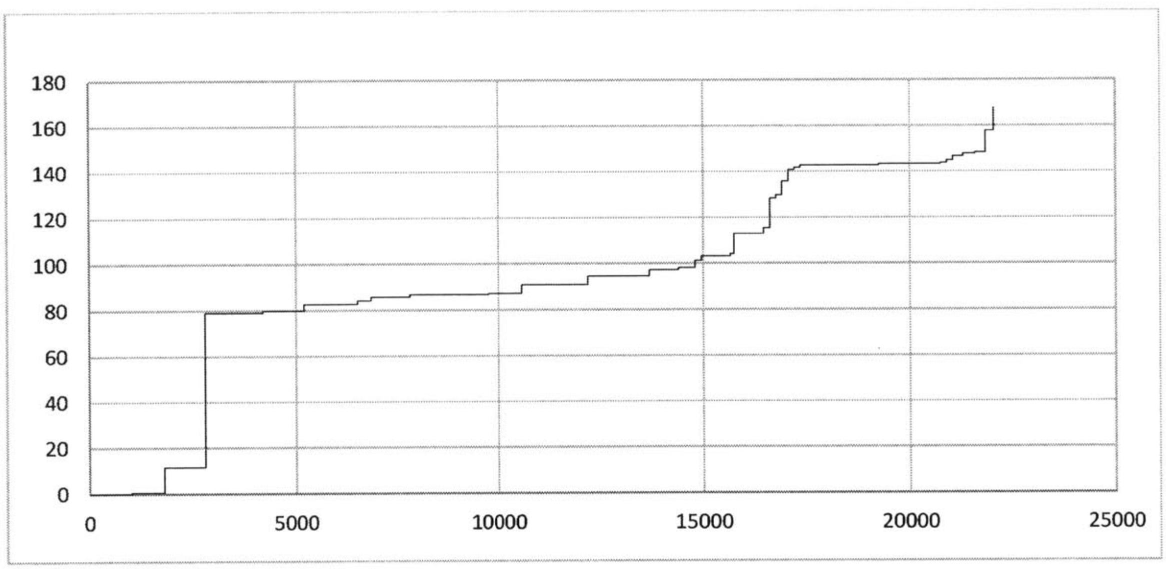

Figure 24 - MC Step Chart for round 2

In this scenario, the supply curve would change from day to day as the opportunity cost has to be changed due to the discount rate that is included in its calculation. The market clearing price would then be derived by intersecting the vertical demand curve and the respective cost step curve for each hour of the three days.

Doing that, we calculated all cash flows for the respective portfolios as seen in the table below and their NPV.

	MC	Worst Alternative	Free Permits
Portfolio 1	1,252,211.20	-1,184,023.95	9943
Portfolio 2	1,383,164.40	-1,053,070.75	12698
Portfolio 3	1,998,249.49	-437,985.66	1184
Portfolio 4	1,735,302.66	-700,932.48	11660
Portfolio 5	2,040,822.34	-395,412.81	10892
Portfolio 6	1,730,934.58	-705,300.57	14941
Portfolio 7	2,476,235.15	40,000.00	22354

Table 20 –EEE Portfolio Valuations using the wrong number of free permits

We then added the NPV of the free permits to this number to come up with a first valuation of the seven portfolios. Now, we had to factor in that portfolio 7 was sold at a fixed price of $40,000. So, we got the valuations that are shown in the table below.

	MC	Worst Alternative		Free Permits
Portfolio 1	1,705,544.54	395,023.67		14941
Portfolio 2	1,383,194.63	72,673.77		12698.33333
Portfolio 3	1,998,249.49	687,728.63		1184
Portfolio 4	1,665,642.80	355,121.94		10892
Portfolio 5	2,204,631.86	894,111.00		12698
Portfolio 6	1,433,338.21	122,817.35		11660
Portfolio 7	1,350,520.86	40,000.00		9943

Table 21 –EEE Portfolio Valuations using the right number of permits

Here we have to mention that we first calculated the values of each portfolio using the wrong number of free permits because we simply copied these permit numbers from the official spreadsheet without checking whether they were in the right order. Because of that portfolio 7 was highly overvalued in our calculations. We now changed this and got different values for the seven portfolios. During the portfolio auction it seemed like we were not the only group that made this mistake because there were other groups that bid negative values for some of the portfolios, too. Finally, we bought portfolio 7 for $40,000. This meant that we could not rely on market power to increase our profits, but the chance that some other group would increase the market price to around $500 during hour 3 was very high based on the good experiences our world made with market power during the first round.

In the end, implementing a cap-and-trade system theoretically meant that plants with higher CO_2 emissions would have higher marginal costs due to the added opportunity cost of the required permits. This should raise the overall price level of electricity in the market as demand was still inelastic and all marginal cost curves were moved upwards. That, however, was not the case in our market during the first few hours as the prices remained as low as before. Nevertheless, we did not change our bidding strategy of bidding marginal cost because it made no sense to sell electricity at around $40 when the opportunity cost of permits was higher than that. The low prices for electricity showed us that there might be irrational behavior in the market.

After the second day of round 2 we approximated that we would need a minimum of 480 more permits if we produced similar amounts of electricity as before. So, we decided to buy them at $93.25 which was lower than the NPV of having to buy the permits from the regulator. In the end, we sold far less electricity than calculated which meant that we tried to sell 200 of our permits after the last round. Even though we were in selling negotiations with another group, we did not manage to complete the deal by the end of the game.

Final Thoughts

In a deregulated market, there are two goals of the auction: (1): Make sure plants dispatched based on MC merit order (2): Price at marginal cost. Regulators also face with a choice between "uniform" and "discriminatory" (pay-as-bid) auctions. Uniform seems to be more popular, although some markets such as California have used both at some point in time. As in ESG, the challenge of uniform auction is that the price is set by the bid of the last used unit. In discriminatory auction, each plant is paid its bid. While this can be appealing, it is not necessarily the preferred choice. Indeed, discriminatory auctions force infra-marginal units to guess what the market clearing price will be between uniform-pricing and pay-as-bid in wholesale electricity markets. The figure illustrates the difference between supplier bids and costs in the Pay-as-Bid Auctions.

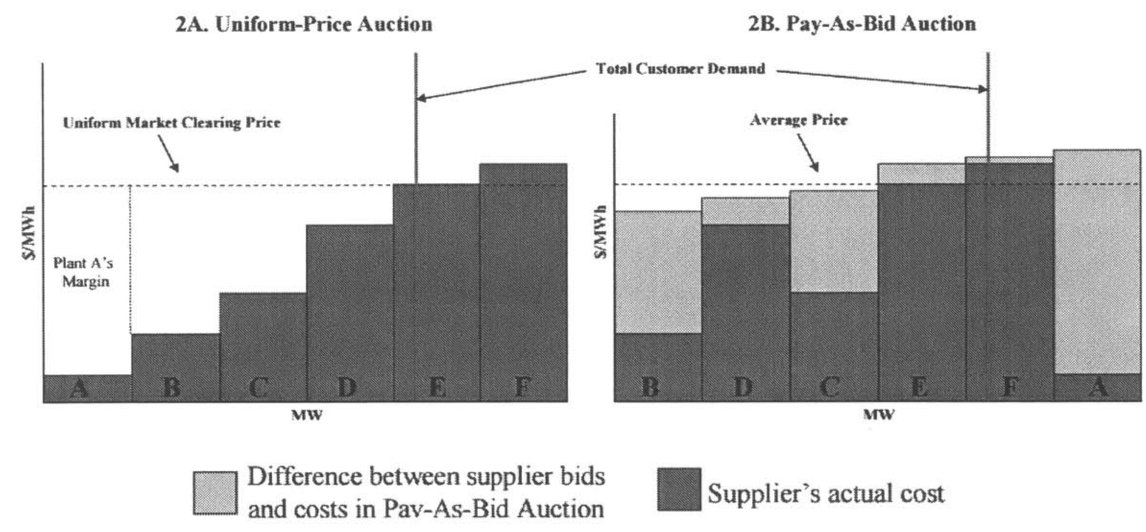

Figure 25 - Difference in two auction approaches - *Source: Electricity restructuring (C.R. Knittel Lecture Notes – Spring 2014)*

Under perfect competition, the dominant strategy is to bid marginal cost. Bidding marginal cost in a discriminatory auction implies fixed costs will never be recouped. With discriminatory auctions, if you have multiple units, you can no longer guarantee your "infra-marginal units" will run, and try to set the price with your "marginal units". This makes exercising market power more difficult under discriminatory auctions. The risks associated with "guessing" wrong are larger. But, uniform auctions might make tacit collusion easier. Discriminatory auctions increase information costs and likely puts small firms at disadvantage.

In brief, restructuring is not necessarily about short-term price effects. The real benefits arise from lower costs due to increased efficiency in the short- to medium-run and better investment decisions in long-run.