Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2006 Proceedings

International Conference on Information Systems (ICIS)

December 2006

Measuring Coordination through Social Networks

Hossain Liaquat
University of Sydney

Andre Wu
IBM Australia

Byounggu Choi
University of Sydney

MEASURING COORDINATION THROUGH SOCIAL NETWORKS

Quantitative Research Methods

Liaquat Hossain
School of Information Technologies
University of Sydney
lhossain@it.usyd.edu.au

Andre` Wu
Global Business Consulting
IBM Australia
awu@ibm.com.au

Byounggu Choi
School of Information Technologies
University of Sydney
choi@it.usyd.edu.au

Abstract

In this study, we explore the correlation between actor centrality and project based coordination. By drawing from established coordination and organisational process theory, a text-mining tool is designed and implemented to measure coordination from a large dataset of emails. Here, we provide effective mechanisms for: (i) cataloguing coordination key phrases from an email corpus; (ii) the calculation of coordination score based on project scope; (iii) the construction of social network matrices using centrality measures, and (iv) approaches for exploring the association between network centrality and coordination score. We argue that actor centrality affects the ability of an individual to coordinate the actions of others. The following questions guide this study--What is the effect of network centrality on organisational coordination? How is the actor's ability to coordinate projects related to their structural position in the communications network? We developed multi-layered test designs to explore this relationship in a project-based (macro) and cross-project (micro) level. We suggest four major findings from the analysis of communication data from Enron email corpus. Firstly, it is concluded that centrally positioned actors show more coordinative activity. Secondly, it is found that betweenness index of centrality is the most potent predicate for coordination. Thirdly, the influence of an actor is associated with coordination more so than the actor's prominence. Lastly, results suggest that coordination ability is more closely correlated to actor centrality than the organisational position. It is therefore concluded that centrally 'well-connected' people are able to exercise greater coordination within the network structure.

Keywords: Social network analysis, coordination theory, longitudinal data analysis, cross-level analysis, clustering techniques

The Context of the Study: Enron Corporation

The Enron Corporation was an energy trading, natural gas, and electric utilities company based in Houston, Texas, which employed around 21,000 people by mid-2001. After a wave of accounting scandals, the company filed for bankruptcy on December 2, 2001. The US Justice Department investigated whether Enron defrauded investors by concealing information about its finances.

Enron was originally involved in the distribution of electricity and gas throughout the United States and the development and operation of power plants, pipelines, and other infrastructure worldwide. In the 1990s, Enron expanded its business front into energy trading and even securities trading. As a result, Enron was named "America's Most Innovative Company" by Fortune magazine for five consecutive years, from 1996 to 2000. Enron's global reputation was undermined, however, by persistent rumours of bribery and political pressure to secure contracts in Central and South America, in Africa, and in the Philippines (McLean and Elkind, 2003). On January 9, 2002, the United States Department of Justice announced it was going to pursue a criminal investigation of Enron and Congressional hearings began on January 24. After a series of scandals involving irregular accounting procedures bordering on fraud involving Enron and its accounting firm Arthur Andersen, it stood at the verge of undergoing the largest bankruptcy in history by mid-November 2001 (Wikipedia, 2005). During 2001, Enron shares fell from US$85 to US$0.30. As Enron was considered a blue chip stock, this was an unprecedented and disastrous event in the financial world. Enron's plunge occurred after it was revealed that much of its profits and revenue were the result of deals with special purpose entities (limited partnerships which it controlled). The result of this was that many of the losses that Enron suffered were not reported in its financial statements. The context of this study was focused on three of Enron's prominent undertakings. These three projects were: (i) the Dabhol Power Company, (ii) the Azurix Water Company, and (iii), the JEDI Partnership. These three projects were selected as the sample frame as they shared many employees that were common to all three. This was required for the cross-project study of variances in coordination and centrality. In addition, these projects were selected because of their sheer size and scope, along with their notoriety. Here we provide a brief background of the selected three projects.

Dabhol Power Company

In 1997, Enron created the Dabhol Power Corporation (DPC) for the purchase and sale of electricity in Maharashtra, India. Enron International unveiled an energy plan that included a new Power Plant and pipeline from Dabhol to Hazira at an estimated cost of $2.8 billion. The project faced a major problem in that the newly elected state government threatened to cancel the deal because of its high price tag and the alleged corruption by the previous government that negotiated the project (Chatterjee, 1995). The project involved two phases; the construction of the power plant (740 megawatts), and the eventual expansion of its output capacity (1,444 megawatts). By May 1999, phase I was completed and the DPC reported profits of $42 million during the first year of its operations. However, phase II of the project stagnated and in December 2001, Enron filed for Chapter 11 bankruptcy before the project was completed.

Azurix Water Company

In 1998, Enron moved into the water sector, creating the Azurix Corporation, which it part-floated on the NYSE in June 1999. Azurix struck a major deal to operate the water and sewage for two regions of Argentina's Buenos Aires province. Azurix paid $439 million for the 30-year concessions, which served just under two million people. With operations in Argentina, England, and Mexico, Azurix was a globe-spanning company. Although Azurix was profitable, it wasn't living up to the majestic expectations of Enron. The company reported net income of $37.7 million in 1999, on revenues of $618 million. Overall, Azurix failed to break into the water utility market, and in April 2001, Enron announced its intention to break up Azurix and sell its assets.

JEDI Partnership

Enron had a partnership known as the Joint Energy Development Investors (JEDI). It was a 50-50 partnership between Enron and the public pension system of California. Starting in 1993 both partners committed $250 million over three years to JEDI to invest in natural gas projects (Fox, 2003). At the time, Enron did not include JEDI in its earnings since it did not own more than fifty percent of the partnership. Then in 1997, Enron created Chewco which bought 100% stake in JEDI. With the purchase, Enron essentially owned JEDI. However, since Enron had organised Chewco in such a way, the entire partnership was kept off of its books. Since Chewco was not included, the JEDI partnership, still apparently run by Enron and an outside group, also remained off the balance sheet. Consequently, the Chewco "partnership" made it possible for Enron to keep roughly $600 million off its books (PBS, 2005). This was a major revelation during the congressional hearings as part of the criminal investigation.

The Enron Dataset

Email logs from the Enron Corporation between 1997-2002 were made public by the FERC during their investigation. This became known as the Enron dataset, or the Enron corpus. The Enron corpus is unparalleled in terms of email datasets that can be used for research purposes. It is more extensive than any other research-friendly email corpus by several orders of magnitude (SGI, 2005). There are at least three versions of Enron dataset. The original was prepared by the CALO project and SRI International and had many duplicate and corrupt messages. A team at the University of Southern California (USC) cleaned it and created a MySQL database from the dataset to assist in the statistical analysis of the data. The MySQL version contains 252,759 messages belonging to 17,568 total users. As the corpus is a structured database, it allows extensible queries to be run on the dataset. It provides added flexibility such as partitioning the dataset into emails based on project scope. This dataset was used to extract evidence of coordination and to perform centrality measurements. Our study of coordination requires the position of each employee in the organisational hierarchy. To accommodate this, the role of each employee was added to the MySQL database. This modification was made to the database by the Social Network Reading Group at the University of Sydney. The final database schema used for this study can be seen in Figure 1.

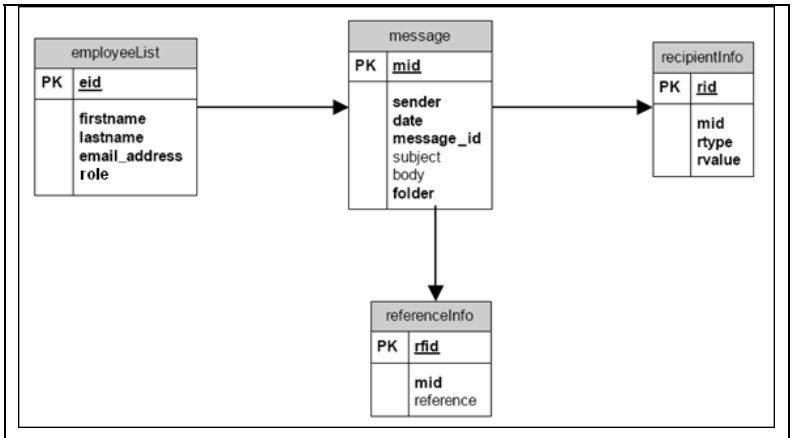


Figure 1. Enron Database Schema Used in This Study

Research Question

The purpose of this study is to determine the effects of actor centrality on coordination. The study builds upon established coordination theories to investigate differences in coordinative activity between individuals with high and low network centralisation. The central research question may thus be phrased as: Are centrally 'well-connected' people are able to exercise greater coordination within the network structure. Does an individual's 'potential for communicative activity' become reflected in its actual coordination score? Does the data reflect the theories of centrality and coordination?

The nature of this study serves to find evidence of association but not necessarily causality. The research claim is that high network centrality increases the ability of an individual to coordinate the actions of others. This is made possible through increased leadership and social influence, a power attributed to the structural position. It is claimed that the inverse should also hold true. Low network centrality decreases the ability to coordinate. We investigate the following sub-problems for this study.

(i) Investigate whether high network centrality increases the ability to coordinate.

(ii) Investigate whether low network centrality decreases the ability to coordinate.

If (i) and (ii) hold true; then:

(i) Determine which measure of centrality is the most potent predicate for coordination; degree, closeness or betweenness.

(ii) Determine whether in or out-centrality correlates closer to coordination.

(iii) Determine whether an actor's organisational position or network centrality has a greater effect on coordination.

Exploring the Relationship between Centrality and Coordination

With an understanding of what this study sets out to achieve, we continue with a look at how social network analyses have been performed in the past. Coordination theory is collated and synthesised in a way that allows the application for this study. This review of literature will discuss established concepts and mechanisms used in studies of centrality and coordination. We seek to answer questions such as: Does higher network centrality increase one's ability to coordinate? Does the inverse also apply? Which measurement of centrality is the most useful predictor for coordination ability? Is in or out-centrality more important? Is organisational position indicative of coordination ability?

Introduction to Centrality and Coordination

In a purely graph theoretical sense, centrality is the state or quality of being central in a network structure (Faust, 1997). Centrality is a structural attribute of nodes in a network, not an attribute of actors themselves, but of their structural position in the network. Measures of centrality are based such attributes as the closeness, degree or betweenness of a point. In a social network sense, centrality has been defined by leading social network researchers as a measure of the potential importance, influence, and prominence of an actor in a network (Freeman, 1979; Borgatti, 1998). Therefore, it can be argued that the potential importance, influence and prominence of an actor are important indicators for a strong coordinator. Measures of centrality and prominence were designed for identifying key individuals in a social network (Zemljic and Hlebec, 2005). Coordination is an abstract concept that is difficult to be measured quantitatively. It has been measured using a combination of other factors such as centrality and the strength of social ties. Coordination is primarily measured using qualitative methods, which identify the people who have the potential to lead and influence others.

Relationship between Centrality and Coordination

The traditional approach to coordination was to delegate more authority to a single actor. The idea was to secure coordination by control from the top, a 'coordination by command' approach (Donini and Niland, 1994). This notion has been a contentious issue mainly due to the difficulties in selecting a suitable governing body. Contemporary thinkers contend that it may be timely to consider whether an organisation should be reconceived as constituting a social network (Moore et al., 2003; Stephenson-Jr. 2004). Minear (2002) argued that "coordination is multilayered, involving the orchestration of relationships not only at headquarters but also at the regional, national and field levels". As a consequence, the top-down principal agent perspective of coordination needs to be reconsidered.

Krackhardt and Hanson (1993) suggested that social networks provide a better insight into organisational behaviour than formal structure. A reason for this is that central nodes can exert more influence by virtue of being linked with a large number of people in the network. They are more likely to be connected with other powerful actors in the network, potentially receiving information of higher quality (Carley et al., 2003). Network centrality can be viewed as a source of informal power. Like formal authority, it can translate into a high level of access to various resources (Burt, 1982). The difference between formal and informal sources of influence [power] is that informal power stems from an actor's position in communication patterns and interaction whereas formal sources are defined by position in the organisational hierarchy (Monge & Eisenberg, 1987).

It has been argued that centrality measures are an explanatory variable in studies of actor influence and actor control (Ove, 2002). Network structures in an organisation form a social network and the participants should be encouraged to support a culture which promotes knowledge sharing and consensus building. This term has been coined 'collective sense making' (Weick and Sutcliffe, 2001), and may exist across successive instances of joint engagement. Weick and Sutcliffe (2001) further suggest that maintaining social ties between collaborative work

groups is critical for knowledge sharing and improved coordination. A central node is strategically located on the communication paths linking pairs of others. A person in such a position can influence the group by withholding or distorting information in transmission. Shimbel (1953) stressed the responsibility of persons occupying such positions for the maintenance of communication links and emphasised their potential as coordinators of group processes.

Networks that have few or weak connections, or where some actors are connected only by pathways of great length may display slow or ineffective response to stimuli. This is due to the reduced potential for social influence. Networks that have more and stronger connections with shorter paths among actors may be more robust and more able to respond quickly and effectively (Kapucu, 2005). Measuring the number and lengths of pathways among the actors in a network allow us to index these important tendencies of whole network (Hanneman, 2001; Wasserman and Faust, 1994). In a study of inter-organisational relations, Galaskiewicz (1985) stated that hierarchical models of resource allocation and coordination emphasise the power and dependencies that develop during inter-organisational transactions. Research on resource dependency has shown a strong association between an organisations' network centrality and their supposed influence in community affairs. The central and more visible actors in the network are more likely to be potential allies with other powerful actors, thereby appearing even more powerful.

Centrality and its Implications on Actor Coordination

Freeman (1979) defined three measures of centrality and explained their structural implications. The three measures identified were degree, betweenness and closeness (see Figure 2). Freeman stated the degree of a point and seemed to be an index of that position's potential for activity in the network. Betweenness is the extent to which a point falls between others on the shortest paths connecting them. It was taken to be an index of potential for control of communication. Closeness measures the distance of a point to all others. This was viewed as a measure of independence from control. According to Freeman these kinds of centrality imply three competing 'theories' of how centrality might affect group processes.

Measure	Social Implication
Betweenness	Control
Degree	Activity
Closeness	Independence

Figure 2. Measures of Centrality and Their Social Implications

The formal properties of centrality linking structural position and coordination of small groups were first documented in an MIT experiment in the 1950s (Bavelas, 1950; Leavitt, 1951). In the late 1970s, Freeman, Roeder and Mulholland (1980) replicated the MIT experiment with groups of five subjects based on four different structural forms. The visual and auditory communication paths were restricted on the structural forms. In performing the group task with controlled communication links, the researchers were able to measure the level of perceived control, independence and activity. The centralisation of each position in the structural form was measured based on betweenness (control), independence (closeness) and activity (degree). In a post-experimental questionnaire, the participants were asked to identify the position of the person who acted as the leader. The majority participants identified the most central position as the leader. The study found evidence that centrality is associated with leadership. We extend on this finding to test whether centrally positioned actors show more coordinative activity.

Hypothesis 1: Centrally positioned actors show more coordinative activity.

In the study by Freeman et al (1980), Centrality did emerge as an important structural variable, but not the traditional kind of centrality based on closeness (independence). Instead, the experimentally important kinds of centrality are based on the potentials for activity (degree) and control (betweenness). The level of certainty in determining the group leader was calculated using an analysis of variance. It was found that the levels of certainty rate in the same order as predicted by the control-based measure of centralisation (betweenness). This supports the argument that betweenness is the key to leadership. Since betweenness is based on potential for control of

communication, this outcome makes intuitive sense; perceived leadership is related to what is termed 'control potential'.

With an analysis of 30 years of major social psychological journals, Mullen et al. (1991) performed a meta-analysis on the effects of centrality. They studied three aspects of individual behaviour as part of a group task; leadership, satisfaction, and messages sent. It was concluded that the betweenness index is the most powerful independent predictor of the effects of centrality. The individual in the most centralised position in terms of betweenness is likely to emerge as the leader, to be more satisfied, and to participate more in the task solution. Moreover, this indicates that the potential for the control of communication seems to be uniquely important in the development of leadership in communication networks. These behavioural characteristics support the findings of the MIT study over 50 years earlier (Leavitt, 1951) and the study by Freeman et al. (1980).

Hypothesis 2: Betweenness is the most potent predicate for coordination in undirected graphs.

Communication links are often skewed towards one direction. The nodes can be adjacent to or adjacent from another node depending on the direction of the relation. Directed networks differentiate between messages sent and messages received. Using these directed networks, it makes sense to distinguish between the 'in-centrality' and the 'out-centrality' of the various points (Knoke and Burt, 1983). In-centrality is measured much like regular centrality but uses only the messages received. In social network applications, it has been argued that out-centrality measures the expansiveness and influence of the actor, whereas in-centrality can be regarded as an indicator for the prominence or popularity of an actor (Wasserman and Faust, 1994).

Hypothesis 3: Out-centrality measures are the most potent predicate for coordination in directed graphs.

Organisational Position and Centrality and their Effects on Coordination

Carley et al. (2003) studied the relative contributions of individual role characteristics and actor centrality on individual performance. Measuring the characteristics from email archives, it was suggested that an employee's functional role, organisational position, and communication role have a direct influence on individual performance, and an indirect influence through actor centrality. The study is pertinent to coordination due to the selected measurement of performance. Performance was measured based on the influence of the actor. The direct effect of functional role, status, and communication role were not supported consistently by the data. There was clear support, however, of a mediating effect of actor centrality in all these relationships (Carley et al., 2003). It was found that the effects of network centrality (structural characteristics) surpassed that of the individual role characteristics.

Hypothesis 4: Network centrality correlates more closely to coordination ability than does organisational position.

Using simulated experiments on a network of email contacts within the HP Labs, Adamic and Adar (2005) tested assumptions regarding the structure of social networks and the closeness of actors. The study measured the distance and ease of communication with people that are connected in various ways. The overall distance (number of hops) that people are connected by is critical to effective coordination. According to Chwe (2000), minimising distance between actors increases an individual's social influence. This increases the ability of an individual to coordinate the actions of others. The simulated search experiments measured the closeness of people in the network using three corresponding properties of the nodes: degree, position in the organisational hierarchy, and physical location.

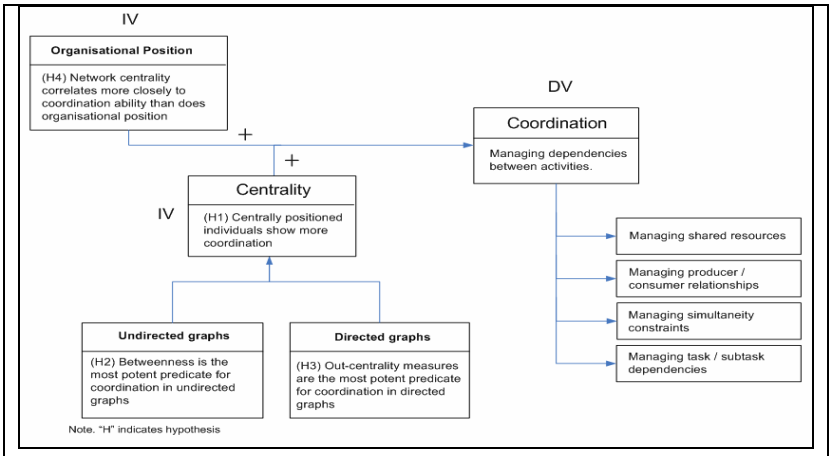


Figure 3. Research Model

The search based on organisational hierarchy and geography both outperformed that based on the degree of one's contacts. This paper shows that, as far as email communication is concerned, local degree centrality is not a feasible way of minimising distance. However, the experimental findings of Freeman et al. (1980) found that centrality is not necessarily based on closeness. In fact, it was found that closeness had the weakest effect on coordination. Instead, the important kinds of centrality are based on the potentials for activity (degree) and control (betweenness). A limitation of this study is that the authors looked only at the degree measurement of centrality. The study neglected to test betweenness centrality, which has been shown [in the experiment by Freeman et al. (1980) and Mullen et al. (1991)] to outperform degree centrality in coordination ability. Using betweenness centrality, perhaps the results may have been significantly different than what was presented in this study. The authors also failed to capitalise on the directional analysis of the dataset. Given that it was an email dataset, the authors could have tested the effect of routing messages through actors with high in or out-centrality. This has practical implications as traditional organisations gradually shift towards network centrality as a deciding factor in coordination strategies. In this section, we presented a review of established measures of centrality and their main uses and implications. Studies of coordination theory are introduced with a focus on measuring coordination and social influence. We discussed the relationship between centrality and coordination. The research hypothesis is presented to outline the origins of the propositions. The conceptual framework drawn from the aforementioned review of literature is presented below in Figure 3.

Furthermore, a summary of the research findings are shown in Figure 4. Each idea is listed with the hypothesis that inspired it and the test which will operationalise the hypothesis. With an understanding of the relationship between centrality and coordination, the next section describes how the hypotheses of the study are tested. We first discuss social network analysis (SNA) as a methodology for studying coordination theory. This study builds on the underlying assumptions of coordination theory to study organisational processes. These assumptions involve the creation, dissemination and processing of information. We implement these concepts, in addition to text mining, statistical analysis and coordination theory to present the process-action approach. To measure the effect of centrality on these coordination processes, data is collected on the actors which enact these coordination processes and their relative centralities is further analysed to determine if a correlation exists.

Overview of Related Literature	Influenced Hypothesis	Tested By
1) In a study of organisational coordination, it was found that higher levels of centrality correlated to higher involvement, stemming from greater internal coordination (Moore et al. 2003).	H1	T1, T2
2) In a research experiment of coordination in different structural forms, it was found that the central nodes were most often identified as the leader and coordinator (Freeman et al., 1980)	H1	T1, T2
3) A meta-analysis was conducted on major social psychological journals, studying the effects of centrality. It was concluded that the betweenness [the control-based measure of centralisation] is the most powerful independent predictor of the effects of centrality. (Mullen et al., 1991; Freeman et al., 1980)	H2	T1
4) In a study on the effect of clique structure on coordination, it was argued that weak ties and the dispersion of social influence relate to the effectiveness of betweenness centrality in predicating the strong coordinators. (Chwe, 2000)	H2	T1
5) It was argued that out-centrality measures the expansiveness and influence of the actor, whereas in-centrality can be regarded as an indicator for the prominence or popularity of an actor. (Wasserman and Faust, 1994)	H3	T2
6) It was found that the effects of network centrality [structural characteristics] surpassed that of the individual role characteristics. (Carley et al, 2003)	H4	T3
7) A study on email communication patterns found that people are more closely tied to those they are analogous to in terms of organisational position rather than those tied through network centrality. (Adamic and Adar, 2005)	H4	T3

Figure 4. Literature Review with Related Hypothesis

Social Network Analysis (SNA) of Enron Email Dataset

This study investigates the effect of structural network centrality on the individual and on the group level. We present an innovative approach combining SNA, text mining, and statistical analysis for exploring the correlation between actor centrality and coordination. The 'process-action approach' is proposed as a new method for measuring coordination. This method is the first to apply SNA techniques for studying coordination from an email dataset. This analysis illustrates patterns of communication and information flows among actors. The framework for this approach can be applied at a higher level in other complex and sensitive settings to study coordination.

The context frame of this study is focused on three of Enron's prominent undertakings. These three projects are: (i) the Dabhol Power Company, (ii) the Azurix Water Company, and (iii) the JEDI Partnership. The motivation for studying coordination on a project based scope is to better capture the coordinative processes as the employees work towards a common goal. This definition of project scope goes beyond the pattern of messaging and takes into account the reason for messaging. The emails are more likely so support messages that were useful, meaningful and oriented toward the project goal. Building on the ideas of Carley et al. (2003) and Ibarra (1993), three distinct projects are used to minimise limitations of examining coordination and its determinants from a single point in time. It also serves as a basis for improving the confidence that any findings are not attributable to one particular project or environment.

The project scope is extracted from the email dataset using a keyword match on the project name, along with any replies to those emails. In addition to the project name, common names associated with the project are also used. In some cases, such as Maharashtra, the name of an Indian region is used because Enron's sole dealing with the region is through the Dabhol project. The list of project names and the alternate associations are provided in Table 1.

Table 1. Project Names and Alternate Associations	
Formal Project / Company name	Alternate associations
Dabhol Power Company	Dabhol, DPC, Maharashtra, MSEB [Maharashtra State Electricity Board]
Azurix Water Company	Azurix, Wessex Water, BOT Contract, WaterDesk.com, Water2Water.com, American Water Works
JEDI Partnership	Joint Energy Development Investment Limited Partnership, Chewco, Belco, California Public Employees Retirement System

The methodology of the study involves four research phases:

(i) Extraction and cataloguing of coordination key phrases;

(ii) Calculation of coordination score bounded by project scope;

(iii) Construction of social network matrices using the centrality measures; and,

(iv) Hypothesis testing of the association between network centrality and coordination score

Phase One – Coordination Key Phrases (Model Building)

The study of coordination requires a clear definition and a standard method of measurement. The challenge is to measure coordination in an objective manner so that comparisons can be made between different people or of the same person in different contexts. In this study, coordination is measured with the application of text-mining techniques. Crowston (1994) suggested that coordination processes depended on the mechanisms chosen to manage dependences among tasks and resources. These mechanisms primarily involve the creation, dissemination and processing of information. It was suggested that by systematically comparing the processes, common patterns became evident. If the organisations are performing essentially the same task, typically the same basic steps are required. Building on the core coordination mechanisms and the four key processes as defined by Malone and Crowston (1994), this paper proposes the process-action approach. This approach involves the extraction and weighting of coordination key phrases. It also provides a mechanism for calculating the score for each person. It is called the process-action approach because it combines the original process-oriented coordination approach with the study of action oriented key phrases.

Using the process-action approach, coordination is measured with the application of text-mining techniques. Text mining is the application of analytical functions relying on sophisticated text analysis techniques that extract information from free-text documents (Dörre et al., 1999). Prior to the application of text mining, it is important to investigate the existing literature on grammatical models and coordination theory. In a study of grammatical models of organisational processes, Pentland (1994) suggested a wide variety of coordination constraints based on the kinds of interdependencies between actions (Malone et al., 1993). In organisational theory, it is difficult to construct a grammar that could sustain a rigorous analogy to the structure of the human brain. Coordinative action is historically situated, culturally embedded, and generally stands in a recursive relation to action (Giddens, 1984). Pentland states that it is difficult to imagine an institutional, technological, cultural, or coordination constraint that does not vary with context and is not subject to revision with the passage of time. Universality is simply not a characteristic that applies to the social world. The lack of an organisational "language faculty" eliminates the possibility of a universal grammar for organisational processes: a single set of universal rules or principles that govern the syntactic structure of all organising processes (Pentland, 1994). Due to the lack of a universal grammar, this study uses a context specific taxonomy. For construct validity, the taxonomy is compiled directly from the Enron email corpus.

Sakurai and Suyama (2004) presented a simple yet intuitive method to discovering key concepts from textual data. The method decomposes textual data into word sets by using lexical analysis on training examples. The key phrases are extracted from the word sets given by the user. These key phrases are then used to map to specific concepts in

the dataset. The paper reports on the application of the method to e-mail analysis tasks for a customer support centre. This methodology builds on the approach outlined by Sakurai and Suyama (2004) in extracting the key phrases indicative of coordination.

This study builds on existing coordination theory and presents a new approach for exploring organisational processes. The underlying assumptions of coordination theory are utilised. These assumptions involve the creation, dissemination and processing of information. By identifying and mapping these coordination processes, we are able to identify the specific instances of coordination. The process of coordination was broken down into four key coordination processes as defined by Malone and Crowston (1994). These four processes were interpreted and operationalised for the study of the email corpus.

Managing shared resources
- Instructing or suggesting a person to perform a task

Managing producer / consumer relationships
- The creation or dissemination of information

Managing simultaneity constraints
- Synchronising tasks between actors
- Taking possible times for an event
- Allocating a time for a particular event
- Passing information about the time of an event

Managing task / subtask dependencies
- Planning tasks and strategy to achieve a higher-level overall goal

Using text-mining techniques, these four processes are operationalised into key phrases to be extracted from the email dataset. Building on a seminal text-mining study by Han and Kamber (2000), the process-action approach to text-mining consists of three stages: (1) the initial exploration, (2) model building or pattern identification with validation/verification, and (3) deployment (i.e., the application of the model to new data). This three stage process for text mining was replicated in a study by IBM Germany (Dörre et al., 1999).

As a training set for the text mining, the CalPERS dataset was used. In 1993, The California Public Employees' Retirement System (CalPERS) engaged in an investment partnership with Enron, each committing $250 million over three years. The CalPERS project dataset was used as the training data to build the list of key phrases. Those key phrases were then used to measure coordination in the Dabhol, Azurix and JEDI project datasets; this is further described in the phase two. The intention was to improve data validity by using a different training set from the primary data for the study. Initial exploration was conducted on the CalPERS dataset to ascertain the type of messages being sent. The model building phase to compile the list of key phrases was broken down into three steps. The first of these steps was the extraction of sentences indicative of one of the four processes of coordination. Each sentence was categorised into the specific coordination process and catalogued (see Figure 5).

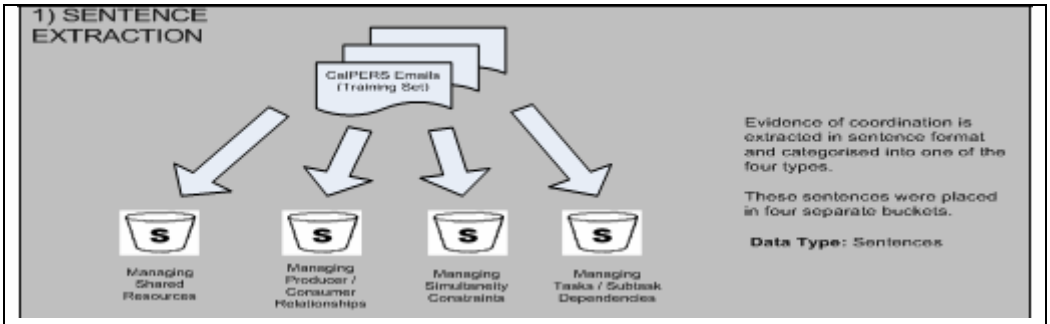


Figure 5. Sentence Extraction

In the second step, the sentences were sorted and the key phrases which underlie the coordinative action were identified and marked. These key phrases would then be put into a distinct 'bucket' for each type of coordination (see Figure 6).

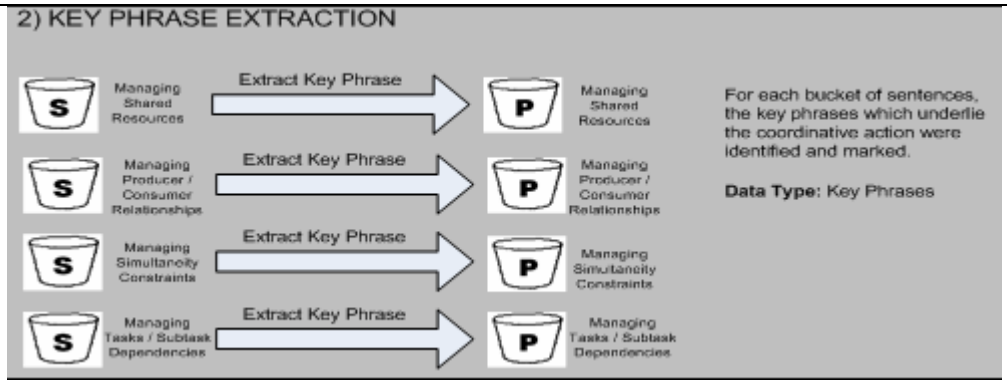


Figure 6. Key phrase Extraction

In the third step of the process-action approach, each of the coordination phrases was assigned a weight based on their level of significance. The weight was determined by the number of people that use the keyword and the frequency with which they use it. The weight of the words is equal to the base two log of the sum of the usage frequency of the words. A word used more commonly was assigned a greater weight. The reason for using the base two log of the frequency was to capture the effect of words with higher frequency without creating substantial outliers. This creates a normal distribution of the coordination weights and reduces the outliers. The weight was measured using the primary data, that is, the Dabhol, Azurix and JEDI Projects. This was done to ensure accuracy in allocating the weights. The weights of the words varied from 1.46 to 9.47, with an average of 4.1 (see Figure 7).

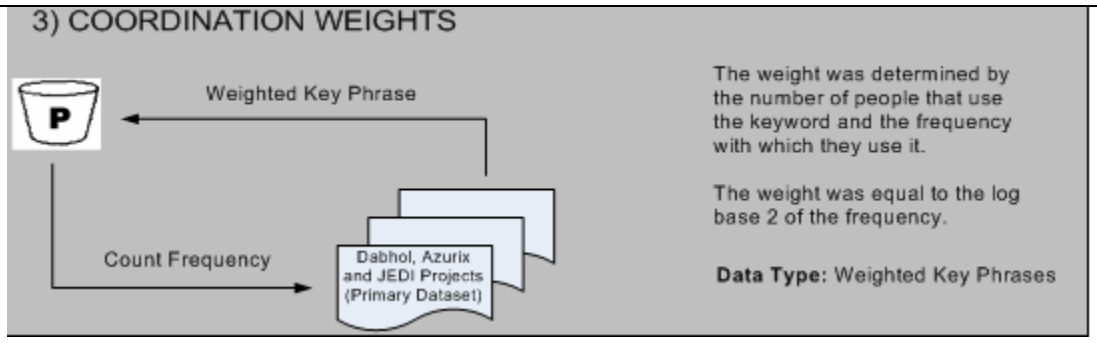


Figure 7. Assigning Coordination Weights

The final list of coordination phrases and their respective weights are shown in Table 2. For the deployment of the model, the key phrases were projected and extracted from the primary dataset. This process is further detailed in phase two.

Table 1. Weighted Coordination Key Phrases Compiled in the Model Building Phase			
Managing Shared Resource	Weight	Managing Task / Subtask Dependencies	Weight
Help coordinate	3.28	I have considered	1.65
Please allow	2.35	I recommend	3.16
Please communicate	4.28	I suggest	2.32
Please coordinate	5.61	I wanted to	4.45
Please do	5.97	I would like to	6.49
Please get	3.61	I would suggest	2.32
Please make arrangements	1.00	We can discuss	3.32
Please make sure	2.80	We can then	1.82
Please update	2.58	We have seriously	1.42
Do you want to	2.00	We need to	7.82
I Request	1.58	We Should	7.02
I would appreciate	4.95	To ensure that	5.74
I would like to	6.49	It will need	3.45
I would like your	1.87	I am changing	1.70
I would ask	2.31	I Believe	6.47
Look into	7.46	Let me know if	6.93
Make sure that	5.67	Please let me know	6.89
Please see	4.58	We have had	4.16
Please speak	4.18	Would probably be	4.08
Please work	3.54	I believe you are	2.16
Managing Producer / Consumer Relationships	Weight	Managing Simultaneity Constraints	Weight
Are as follows	5.12	As we move closer	2.63
Attached is a	4.64	Please allow time	1.57
Attached please find	2.32	On track	7.15
The bottom line is	4.58	Sufficient time	2.26
The purpose is	1.58	Take the time	4.24
For your information	3.90	Agenda	7.79
FYI	5.35	Follow up	5.16
		On time	4.85
		Make a schedule	9.47
		Make a timetable	6.00

Phase Two – Text Mining and Coordination Score (Data Collection)

In the second step of the process-action approach, the key phrases from the list above were extracted from the primary datasets--the Dabhol, Azurix and JEDI projects. This is the deployment phase of the text mining process. A text-mining application was constructed in Java to deploy the model to calculate coordination score from the Enron dataset. Each project scope was treated separately and so the operation was repeated three times. The process of measuring coordination and assigning a coordination score is relatively simple and intuitive once given keyword list. The application reads in the list of coordination keywords and their weights from a text file. The keywords are iterated through and processed individually. The dataset was queried to identify the employees that have used the key phrase in an email. For each match, the program recorded the employee's details added the weight of the keyword to his or her coordination score. This process is repeated for each keyword in the list until the end is reached. Figure 8 shows the pseudocode for the text mining application.

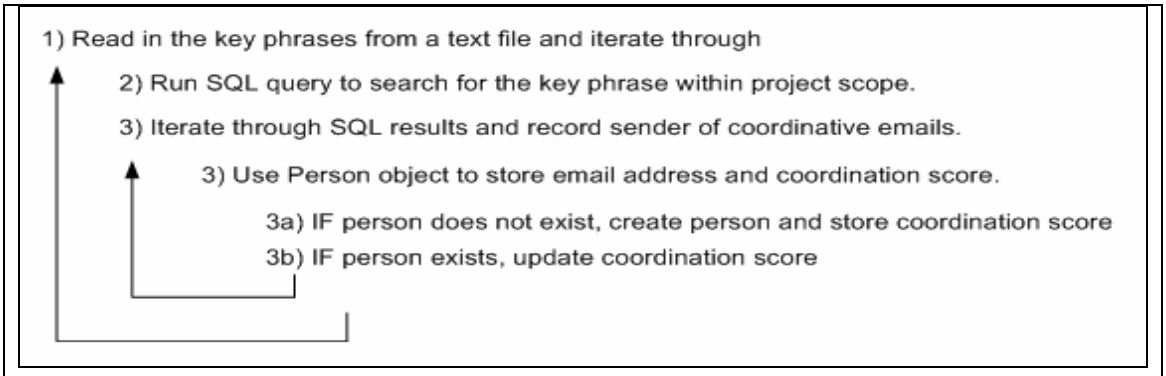


Figure 8. Pseudocode for the Text-mining Application to Calculate Coordination Score

The final coordination score of each person was collated by aggregating the weighted scores based on keyword matches. The text-mining program outputted a list of coordinators and their total coordination score. The program also counted the number of emails sent by each employee in order to gauge the level of activity from each user. The application only counts the emails that are within the project scope. The final coordination score was calculated by tallying the number of hits on each weighted keyword. The coordination score was measured for each employee within the project scope to conduct the hypothesis testing.

Phase Three – Social Network Matrices (Data Manipulation)

Network centrality was measured by the number of emails sent and received by each person as part of the project scope. The list of senders was matched to the recipients based on the recipient type TO, CC (BCC was ignored). The rationale for ignoring the emails received as BCC is further explained by Klimt and Yang (2004), in which BCC emails are usually intended as passive information propagation, rather than establishing a two-way relationship. The strength of the relationship depended on the frequency of emails exchanged. Centrality can be measured with a range of classifications and measurements. The intricacies and social implications of these measurements provide a solid foundation for the results and allow for meaningful conclusions. Centrality was calculated using UCINET 6 for Windows (Borgatti et al., 2002). This study measures centrality on two axes: (i) measurement, and (ii) directional analysis.

Three measures of centrality are used in this study: (i) betweenness, (ii) closeness, and (iii) degree. Flow betweenness measures the extent to which each node lies on the shortest path between two nodes. Closeness is measured by the reciprocal geodesic distances based only on the directed links. Degree measures the number of direct relations of each node. UCINET gives the option of whether to treat data as symmetric or asymmetric for degree centrality. Asymmetric data means the sending and receiving of emails are treated as distinct activities. For the study of directional relationships, the data was treated as asymmetric, following Ibarra and Andrews (1993) and Carley et al. (2003). A directional analysis was performed to investigate the differences between in and out-centrality and its effects on coordination. Directed degree is measured by counting the number of emails sent and the number received by an individual. In-centrality seems to indicate the prominence of an actor, whereas out-centrality measures the influence of the actor (Wasserman and Faust, 1994). In the last step, the text-mining application brings it all together by reading in the UCINET centrality results and combines the final coordination score to output a large spreadsheet with all the figures required for the study.

Hypothesis Testing

This phase ties it together by correlating the centrality measurements with the coordination score. The test designs look at this relationship in a macro and micro level. The four hypotheses tested were that; (H1) centrality is correlated to increases in coordination ability; (H2) betweenness is the best structural characteristic for predicting coordination; (H3) out-centrality is a better predictor than in-centrality and that (H4) network centrality correlates more closely to coordination ability than does organisational position. The hypotheses will be tested with four

different tests (i) project based coordination, (ii) cross-project coordination, (iii) directed centrality and, (iv) organisational position.

Test 1 – Project Based Coordination

This test was performed by dividing the sample into two groups, dichotomised by those 'high' and 'low' in centrality. The groups were divided by the median centrality measurement. The coordination scores from the group high in centrality were compared to those low in centrality. The hypothesis testing determined whether the two groups are different from each other. If the group with high centrality is found to have a higher coordination score and the difference between two groups is found to be statistically significant, it is evidence that centrality positioned individuals show more coordinative activity.

The statistical strength was measured using the Mann-Whitney U-test. The null hypothesis for H1 is that there is no statistical difference between the two groups. The test was repeated for each measure of centrality. All hypotheses are one-tailed, and thus the statistical significance level was set at 0.05 (5%). If the high and low groups were shown to be statistically different for all three measures of centrality, the null hypothesis will be rejected and H1 will be accepted. The Mann-Whitney U test was used to determine whether there is a statistical difference between the two groups of data. The U-test is non-parametric and does not assume a normal distribution. The strength of the U-test is that it factors for the variability and dispersion of the figures.

Seven hundred and twelve people were extracted as part of the Dabhol project scope. This list was loaded into NetDraw for a visual representation. The coordinative activity of each employee was measured using the text-mining application as described in phase one and two. From this set, 173 people were found to have demonstrated coordination in the Dabhol scope. To cleanse the data, the employees with an in or out-degree fewer than three were removed. Those addresses were not used very often and their relevance is questionable. Figure 9 below shows the distribution of the coordination score. A quick investigation of the score distribution shows a right-skew with a long right tail.

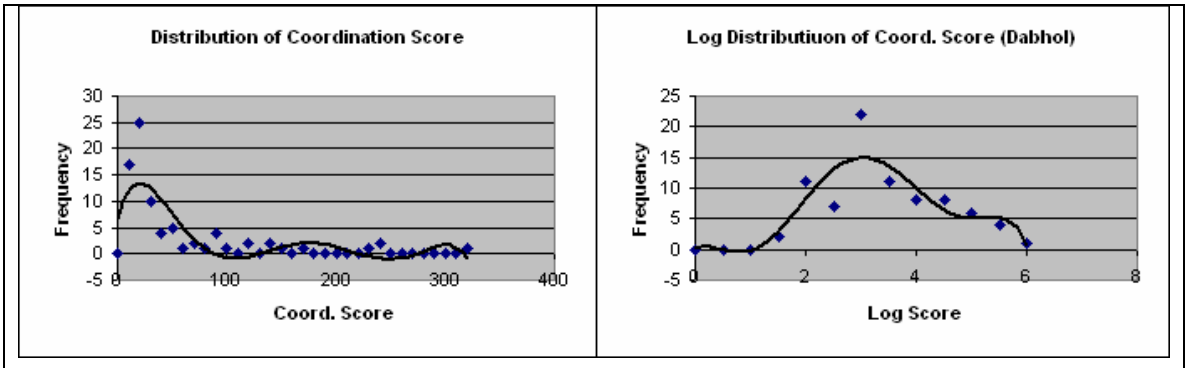


Figure 9. Distribution and Log Distribution of Coordination Score (Dabhol)

If the coordination levels of the two groups are found to be statistically different, it is evidence that centrality is correlated to coordination. The process was repeated for the three projects scopes: Dabhol, Azurix and JEDI. Table 3 shows the results for undirected graphs using Mann-Whitney U test.

Table 3. Results for Undirected Graphs Using the Mann-Whitney U Test

	Dabhol	Azurix	JEDI
UNDIRECTED GRAPHS - Mann-Whitney U Test (5% significance)			
Betweenness	P = 0.0029	P = 0.011	P = 0.0260
Degree	P = 0.0043	P = 0.029	P = 0.0421
Closeness	P = 0.0073	P = 0.0366	P = 0.0273

The three main measurements of centrality (i.e., betweenness, degree and closeness) shows a one-sided P of less than 0.05; the alpha level used for this study. Moreover, the results were replicated across all three projects. The difference in coordination score between the groups high and low in centrality was found to be statistically different. This allows us to reject the null hypothesis as evidence shows a significant correlation between coordination and centrality (H1).

Figure 10 shows an overview of coordinative activity on the Dabhol Social Network. The warmer colours (red, orange, and yellow) seem to be centrally clustered in the diagram. The cooler colours (green, blue, and purple) are more peripheral in the social network. The nodes with no evidence of coordination, or those removed by the data cleansing process, were shown as white.

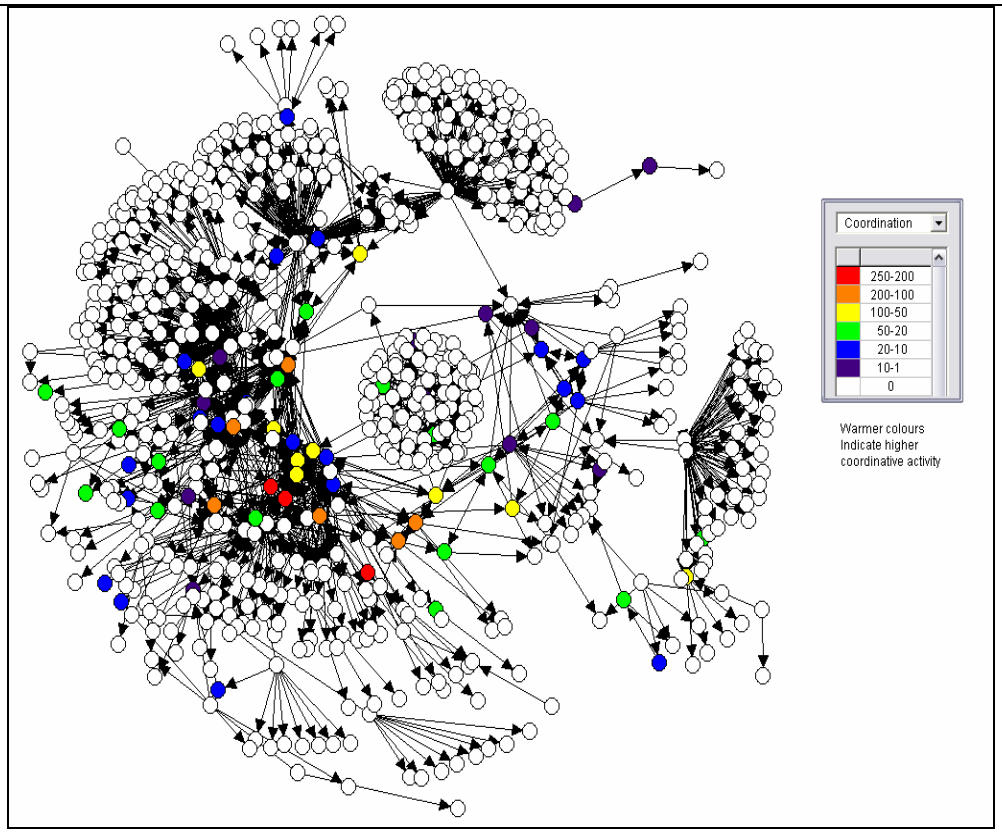


Figure 10. Social Network of Dabhol Project Showing Coordinative Activity

The second hypothesis (H2) is that betweenness is the most potent predicate for coordination in undirected graphs. This test investigated H2 by again using the statistical difference in coordination scores of the high and low groups of centrality. This time, the statistical significance for each centrality measurement was compared against each other. The measure of centrality showing the strongest statistical significance was taken to be the best predictor for coordination. The measure with the lowest p-value is taken to have the strongest significance because it has the least chance of obtaining those numbers by chance alone. In all three projects, it was found that the betweenness measure had the lowest p-value. The p-value for the other measurements (although statistically significant) was not as strong as the betweenness index. This indicates the strongest statistical significance in the relationship between betweenness and coordination. Figure 11 shows the substantial difference in coordination scores between the two groups.

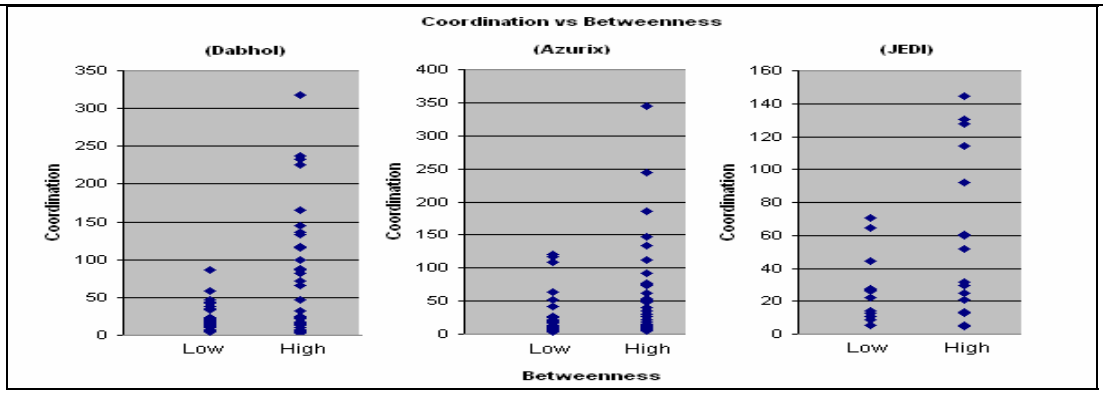


Figure 11. Coordination vs. Betweenness (Groups Based on Median Betweenness)

Test 2 – Cross-Project Coordination

The purpose of this test is to find evidence that employees show greater coordinative activity in projects in which they are centrally positioned. This test builds on the first hypothesis (H1) but instead of comparing levels of coordination between different people; this micro-level test makes comparisons with the same person across multiple projects. The tests were performed by extracting the employees who have demonstrated coordination in at least two projects. The bounds for the project scope are the same as in the macro-level tests (Test 1). Figure 12 shows the relative sizes of the three projects. The rows are divided into one-mode, two-mode and three modes. This test involves performing two and three mode analyses. The columns show the total number of people in each project scope as well as the number that have shown coordination. Test 1 covers the three-mode analysis and test 2 covers the two-mode analyses.

	Total	Show Coordination	
One Mode			
AZURIX	1444	159	11%
DABHOL	712	173	24%
JEDI	393	81	21%
Two Mode			
AZURIX and DABHOL	408	64	16%
DABHOL and JEDI	83	23	28%
JEDI and AZURIX	188	26	14%
Three Mode			
AZURIX, DABHOL and JEDI	70	16	23%

Figure 12. Multi-mode Data of Centrality and Coordination

A multi-mode analysis application was written in Java to perform the name-generation and figure comparisons on the Enron dataset. The application is flexible enough to extract names that are common to any two projects, or common to all three projects. The multi-mode analysis application compares the names, coordination score and normalised degree index for two or more distinct projects, and outputs the common employees along with their centrality and coordination scores.

For each employee spanning across at least two projects, the coordination score and normalised degree index of centrality was extracted from each project. The normalised degree allows for comparisons between networks of different sizes and densities. A normalised degree is achieved by expressing the degree as a percentage the number of actors in the network, less one (the ego). That is, nodes with a higher proportion of direct connections will have a high normalised degree. This figure is comparable between projects because it uses a percentage of the number of total actors in the network rather than the raw number of connections. Another important reason for using the normalised degree as the centrality measure is that every node in the network will have a non-zero positive

normalised degree. This is because the email dataset consists of connected senders and recipients, so the minimum degree an employee can have is one. The betweenness and closeness indexes are often zero, making it impossible to compare.

The goal is to find evidence that projects in which an employee has higher centrality is more likely to also have higher a coordination score. For each individual, the coordination score in the Dabhol and Azurix projects were ranked. The normalised degree in the Dabhol and Azurix projects were also ranked. If the two projects rank in the same order for both centrality and coordination, it is evidence of a correlation. The number of total instances in which the two projects rank in the same order is counted and divided by the total number of opportunities. A high percentage of matches indicate a strong agreement with the hypothesis. This test was then repeated across the three two-mode analyses based on the project permutations.

Three-Mode Analysis

The micro level tests of cross-project coordination begin with the three-mode analysis before moving on to the two-mode tests. The three-mode analysis involved employees across all three project scopes. Sixty-two people were found to have been part of all the three projects. Of these, only eight people demonstrated coordination in all three projects. These eight employees are listed with their details in Figure 13.

Name	Dabhol Coord.	Dabhol Centrality	Azurix Coord.	Azurix Centrality	JEDI Coord.	JEDI Centrality
courtney.votaw@enron.com	538	0.019	736	0.038	9	0.009
james.derrick@enron.com	54	0.118	36	0.036	25	0.007
m..schmidt@enron.com	2241	0.142	1162	0.131	70	0.023
mary.cook@enron.com	12	0.001	12	0.002	48	0.187
richard.sanders@enron.com	12	0.001	49	0.05	24	0.006
rick.buy@enron.com	66	0.044	5	0.005	4	0.004
steven.kean@enron.com	84	0.019	302	0.055	4	0.006
vince.kaminski@enron.com	47	0.031	89	0.015	35	0.004

Figure 13. Coordination across Three Projects

It was found that six of the eight employees had all three project rank in the precise correlation of centrality with coordination. These findings support the correlation between coordination and centrality (H1). However, to compensate for a relatively small sample size in the three-mode analysis, we perform two-mode analyses on larger samples to achieve stronger statistical significance.

Two-Mode Analysis

The two-mode tests of cross project coordination covered the Azurix, Dabhol and JEDI projects two at a time. In this first two-mode test of the Dabhol and Azurix projects, it is revealed that forty eight of a possible sixty four people had the coordination and centrality ranked in the same order. Seventy five percent of the people showed higher coordination in the project in which they were more central. We repeated this method for the other two-mode project scopes; Dabhol and Jedi, and Jedi and Azurix. A summary of the results are shown in Figure 14.

	Show Coordination	Equal Ranked	
Two Mode			
AZURIX and DABHOL	64	48	75%
DABHOL and JEDI	23	17	74%
JEDI and AZURIX	26	21	81%
Three Mode			
AZURIX, DABHOL and JEDI	8	6	75%

Figure 14. Results of Multi-mode Analysis

The three two-mode tests all indicated that people showed more coordinative activity in projects in which the person is more structurally central. An interesting note is that all four multi-mode tests showed that coordination and centrality were ranked equally approximately 75% of the time. These four multi-mode tests all found a strong correlation between coordination and centrality (H1). This supports the results found in the test of project-based coordination (Test 1).

Test 3 – Directed Centrality and Coordination

The third hypothesis (H3) is that out-centrality measures are the most potent predicate for coordination in directed graphs. This test investigated H3 by comparing the correlation between coordination and each of the four types of directed centrality. In-centrality was measured by the number of emails received from distinct individuals. Out-centrality was measured by the number of emails sent to distinct individuals. The use of distinct individuals aims to find people with a wider range and reach, rather than a person who frequently communicates within the same clique. The four types of directed centrality used in this study were the in-degree, out-degree, in-closeness and out-closeness measures. In-centrality was measured using both the in-degree and the in-closeness measurements. The same applies for out-centrality. The strength of the in-centrality measurements will be compared with that of the out-centrality.

Due to the non-normal distribution of the data, non-parametric tests must be used. The Spearman Rank Correlation was used for this test. The Spearman Rank Correlation is a nonparametric (distribution-free) rank statistic used to measure of the strength of the associations between two variables. The Spearman rank correlation gives the correlation as an R-estimate which falls between -1 and +1, and a p-value for the statistical strength. In line with the other experiments, a p-value of 0.05 was the significance level. The results are shown in Table 4.

Table 4. Results for Directed Graphs using the Spearman Rank Correlation

	Dabhol	Azurix	JEDI
DIRECTED GRAPHS – Spearman Rank Correlation			
Out-Degree	r= 0.364 p= 0.000882	r= 0.557 p= 0.000069355	r= 0.373 p= 0.032412
In-Degree	r= 0.268 p= 0.015915	r= 0.306 p= 0.00284	r= 0.032 p= 0.85839
Out-Closeness	r= 0.345 p= 0.001748	r= 0.370 p= 0.00026071	r= 0.109 p= 0.04602
In-Closeness	r= 0.262 p= 0.000959	r= 0.271 p= 0.0024487	r= -0.131 p= 0.46636

The Dabhol and Azurix projects show clear results. In all cases, the out-centrality measurements correlated with coordination much stronger than did the in-centrality counterparts. In all the tests for the Dabhol and Azurix projects, the p-value was found to be significant, and thus all the measures were accepted as evidence. The results for the JEDI project were not so obvious. It must be noted that the in-degree showed an insignificant p-value of 0.858. The in-closeness also showed an insignificant p-value. This means the in-centrality measurements for the JEDI project showed no significant correlation. In contrast, the out-centrality measurements for JEDI did show positive (albeit weak) correlations with significant p-values. This again indicates that the out-centrality measures have a stronger correlation to coordination than in-centrality. A probable reason for the high p-values for the JEDI project was the small size of the dataset. With only 32 nodes, a strong statistical significance is not often found. In summation, the out-centrality measurements were consistently found to be more strongly correlated to coordination than the in-centrality measurements. This is evidence supporting (H3), that out-centrality measurements are the most potent predicate for coordination in directed graphs

Test 4 – Organisational Position and Coordination

The purpose of this test is to determine whether the organisational position of an individual has an effect on levels of coordination. This effect is then is compared to that of centrality and coordination. This is to test the fourth hypothesis (H4), that network centrality correlates more closely to coordination ability than does organisational position. The organisational position (or title) of each employee was categorised into one of the eight ranked types; CEO, President, Senior Vice President, Vice President, Director, Managing Director, Manager and (non-managerial) Employee. These eight titles were adapted from the study of the Enron dataset by Shetty and Adibi (2004). The basic premise is that positions higher in the organisational hierarchy show more coordination.

The Spearman correlation analysis was performed using the ranked title and coordination score. This measures the relationship between title and coordination score. The strength of this relationship was compared with the correlation of coordination to centrality. Like in the cross-project coordination tests, the Freeman degree of centrality was used because it provides the most non-zero indexes. The goal of this test is to determine whether an actor's organisational title or network centrality correlates more closely to coordination. This project used the Dabhol Power Company as the sample scope because it contains the most complete dataset in terms of title and role. In all, there were fifty six coordinators with the title and role information available. These coordinators also passed the filtering mechanisms such as the noise ratio.

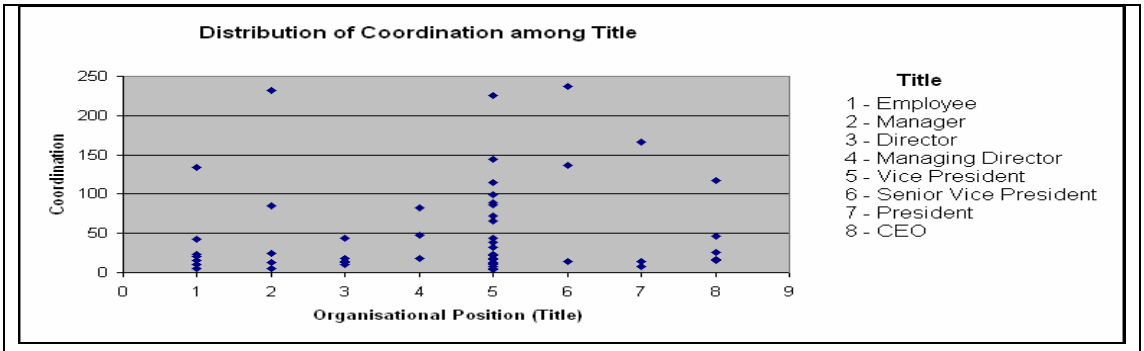
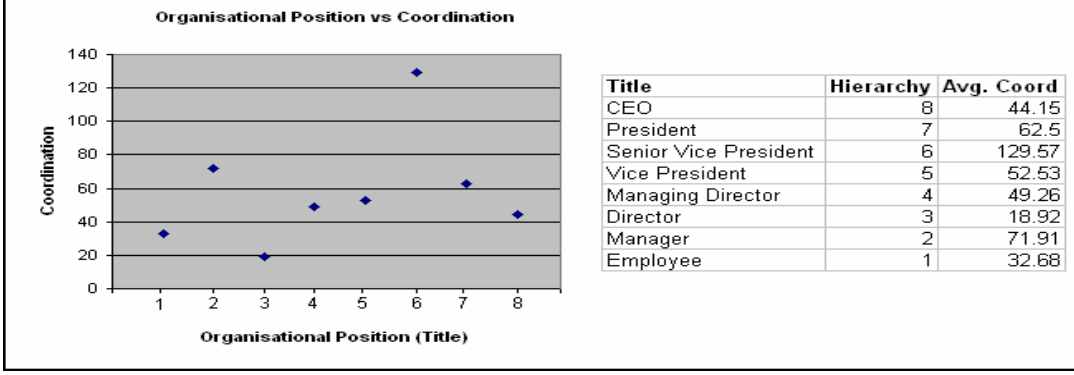


Figure 15. Distribution of Coordination by Organisational Title

The distribution shown in Figure 15 shows a fairly even distribution of coordination across the different organisational titles. Using the organisational hierarchy as a guide, the titles were ranked in the order shown in Figure 16; that is, 8 = CEO, 7 = President and so on.



Title	Hierarchy	Avg. Coord
CEO	8	44.15
President	7	62.5
Senior Vice President	6	129.57
Vice President	5	52.53
Managing Director	4	49.26
Director	3	18.92
Manager	2	71.91
Employee	1	32.68

Figure 16. Average Coordination by Organisational Title

Due to the non-normal distribution of the coordination data we used non-parametric tests to measure correlation. The Spearman Rank test showed a weak correlation of 0.28 and a high p-value of 0.49. These results indicated no substantial relationship between the organisational hierarchy and coordination. This can be seen in the figure 13 which shows a relatively flat distribution with a high outlier. This indicates that the distribution of coordination is evenly spread across the organisational positions and no substantial correlation exists.

In the second step, the association between organisational title and coordination was compared to that of centrality. The Freeman degree of centrality was used for the correlation analysis. To follow on with the non-parametric measures, the Spearman rank correlation was used again. The Spearman test requires the data to be grouped into sets. This was achieved by dividing the dataset based on the centrality. The coordination scores are grouped into those with centrality between 0-10, 10-20, 20-30 and so on. This is shown in Figure 17.

Centrality	Coordination
0-10	26.29
10-20	33.74
20-30	35.35
30-40	31.26
40-50	107.79
70-80	76.21
80-90	85.60
90-100	133.87
100-110	107.38
120-130	237.58

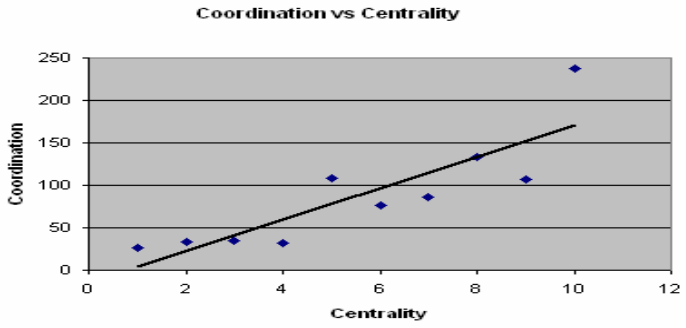


Figure 17. Grouped Centrality and Coordination

The grouped centrality showed a Spearman Rank Correlation score of 0.867 with a p-value of 0.0012. This indicates a strong correlation between centrality and coordination. This test found that the centrality of the individual is strongly associated with coordination while the organisational position showed little correlation. These results provide evidence supporting the forth hypothesis (H4), that network centrality correlates more closely to coordination ability than does organisational position.

Conclusion

The four tests reveal some common patterns with recurring characteristics. For hypothesis one, all project-based tests show a significant relationship between centrality and coordination. The tests consistently found a strong difference in coordinative activity between the groups high and low in centrality. In the tests of project-based coordination, H1 is accepted. The tests of cross-project coordination were performed once using all three projects (three-mode), and three more times using a rotation of two projects (two-mode). The cross-project coordination tests consistently found that 75 percent of employees showed more coordination in projects which they were centrally positioned. This is further evidence that network centrality correlates to increased coordination ability. In conclusion, both the project based and cross-project tests find that hypothesis one holds true. Individuals centrally positioned in a network show more coordination.

For hypothesis two, the betweenness measure was consistently found to have the strongest statistical significance of a relationship to coordination. The strength of the other centrality measurements varied. Two tests showed a stronger relationship for degree and one test showed a stronger relationship for closeness. In conclusion, it was found in all three tests that the second hypothesis holds true, betweenness is the best predictor for coordination ability in undirected graphs. In testing hypothesis three, the experiments of directed centrality consistently found that out-centrality was a greater predicate for coordination than in-centrality. The Dabhol and Azurix projects had clear data indicating this. The JEDI project showed no reliable correlation for the in-centrality measurements and a small but acceptable correlation for the out-centrality measurements. In conclusion, it is found in all three projects that hypothesis three holds true. Out-centrality measures are the most potent predicate for coordination in directed

graphs. In testing hypothesis four; we studied the effect of an employee's organisational position (title) and structural position. The test indicated that the distribution of coordination is evenly spread across the organisational titles; whereas the test on centrality showed a Spearman rank correlation of 0.867 with a p-value of 0.0012. This means the centrality of the individual had a strong correlation to coordination while the organisational position showed no correlation. Of the three project scopes, it was found that the strongest cases were the Dabhol and Azurix projects. It must be noted that these two projects had the largest sample size. The JEDI project scope used the smallest sample size and had the least statistically significant data. This implies that the larger datasets demonstrate a stronger association between centrality and coordination. This is a vindicating result for this study. All hypotheses were supported by the data, with the larger datasets providing stronger agreement. Figure 18 below provides a summary of the results

Test	Hypothesis Tested	Findings of this study
T1	H1	1) The three main measurements of centrality (betweenness, degree and closeness) all show a significant difference in coordination levels between actors high and those low in centrality. These results were replicated across all three projects (Dabhol, Azurix and JEDI).
T1	H2	2) In all three projects, the betweenness index of centrality was found to be the most significant predicate for coordinative activity.
T2	H1	3) The four multi-mode tests all indicated that people showed higher coordinative activity in projects in which the person is more central.
T3	H3	4) In all tests of directed degree and closeness, the out-centrality measurements were consistently found to be more strongly correlated to coordination than the in-centrality measurements.
T4	H4	5) This test found that the centrality of the individual is strongly associated with coordination while the organisational position showed little correlation.

Figure 18. Summary of Results with Hypothesis

In this study, we investigated the relationship between social structures on coordination. Using the process-action approach for measuring coordination, it was found that centrality had a profound effect on coordination. This study further provides a new approach to establishing coordination mechanisms for organisations. All measures of centrality used in this study were shown to be statistically different between the high and low groups. The implications of these results mean that organisations should consider structural position in a network in designing and mapping coordinated groups. These findings are a strong testament to the power of social networks in affecting our day-to-day interactions.

References

Adamic, L. and E. Adar (2005). "How to Search a Social Network." Social Networks In Press, Corrected Proof.

Alavi, M. (1993) "Making CASE an Organizational Reality: Strategies and New Capabilities Needed," Information Systems Management, Spring 1993, 15-20.

Bavelas, A. (1950). "Communication Patterns in Task-Oriented Groups." Journal of the Acoustical Society of America 22.

Bonacich, P. B. (1972) "Factoring and weighting approaches to status scores and clique identification." Journal of Mathematical Sociology. 2: 113-120.

Borgatti, S. P., M. G. Everett, L. C. Freeman (2002). "UCINET for Windows: Software for Social Network Analysis." Harvard, MA: Analytic Technologies.

Burt, R. S. (1982). Toward a Structural Theory of Action, Academic Press, NY.

Carley, Kathleen M., Ahuja Manju K., Galletta Dennis F. (2003) "Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study" Management Science Vol. 49, No. 1, January 2003

Chatterjee, P. (1995). "Enron Deal Blows A Fuse." Multinational Monitor 16(7-8).

Chwe, M. S.-Y. (2000). "Communication and Coordination in Social Networks." The Review of Economic Studies 6(1): 1-16.

Crowston, K. (1994) "Electronic communication and new organizational forms: A coordination theory approach" Working Paper Series 175, MIT Center for Coordination Science.

Donini, A. and N. Niland (1994). "Rwanda: Lessons Learned, A Report on the Coordination of Humanitarian Activities." United Nations Department of Humanitarian Affairs, New York.

Dörre, J., P. Gerstl, R. Seiffert (1999). "Text mining: finding nuggets in mountains of textual data." Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining: pp. 398-401.

Faust, K. (1997). "Centrality in Affiliation Networks." Social Networks 19(2): 157-191.

Fox, L. (2003). Enron: The Rise and Fall. Hoboken, New Jersey, John Wiley and Sons, Inc.

Freeman, L. C. (1977). "A Set of Measures of Centrality based on Betweenness." Sociometry 40: pp 35-41.

Freeman, L. C. (1979). "Centrality in social networks conceptual clarification." Social Networks 1(3): pp 215-239.

Freeman, L. C., D. Roeder, R. R. Mulholland. (1980). "Centrality in Social Networks: ii. Experimental Results." Social Networks 2(2) pp: 119-141.

Galaskiewicz, J. (1985). "Interorganizational Relations." Annual Review of Sociology 11(1): 281-304.

Granovetter, M. S. (1973). "The Strength of Weak Ties." American Journal of Sociology 78(6): pp. 1360-1380.

Han, J. and M. Kamber (2000). Data mining: Concepts and Techniques. New York, Morgan-Kaufman.

Hanneman, R. A. (2001). Introduction to Social Network Analysis. Department of Sociology, University of California, Riverside.

Ibarra, H. (1993). "Network Centrality, Power, and Innovation Involvement: Determinants of Technical and Administrative Roles," Academy of Management Journal, 36(3): pp. 471-501.

Kapucu, N. (2005) "Interorganisational Coordination in Dynamic Context: Networks in Emergency Response Management" Connections 26(2): pp. 33-48

Klimt, B. and Y. Yang (2004). "The Enron Corpus: A New Dataset for Email Classification Research." Machine Learning: ECML 2004: Proceedings 15th European Conference on Machine Learning, Pisa, Italy, September 20-24, 2004. Proceedings: 217-228.

Knoke, D. and R. S. Burt (1983). Prominence. Applied network analysis: A methodological introduction. Beverly Hills, CA, Sage: 95-122.

Krackhardt, D. & Hanson, J. (1993). "Informal Networks: The Company Behind the Chart," Harvard Business Review, 71, 104.

Leavitt, H. J. (1951). "Some Effects of Communication Patterns on Group Performance." Journal of Abnormal and Social Psychology 46: 38-50.

Malone, T. W. (1998). "What is Coordination theory?" National Science Foundation Coordination Theory Workshop, Massachusetts Institute of Technology, Massachusetts.

Malone, T. W. and K. Crowston (1990) "What is coordination theory and how can it help design cooperative work systems?" Proceedings of the 1990 ACM conference on Computer-Supported Cooperative Work, Los Angeles, California, United States, ACM Press.

Malone, T. W. and K. Crowston (1994). "The Interdisciplinary Study of Coordination." ACM Computing Surveys 26(1): 87-119.

Malone, T. W., K. Crowston, J. Lee, B. T. Pentland. (1993). "Tools for Inventing Organizations: Towards a Handbook of Organizational Processes." Proceedings of the 2nd IEEE Workshop on Enabling Technologies Infrastructure for Collaborative Enterprises. Morgantown, WV.

McLean, B. and P. Elkind (2003). Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron, Portfolio.

Minear, L. (2002). The Humanitarian Enterprise: Dilemmas and Discoveries., Kumarian Press, Bloomfield.

Monge, P. R. & E. M. Eisenberg, (1987). "Emergent Communication Networks," in F. M. Jablin et al. (eds.), Handbook of Organizational Communication, Newbury Park, CA, Sage, pp 305-342.

Moore, S., E. Eng, M. Daniel. (2003). "International NGOs and the Role of Network Centrality in Humanitarian Aid Operations: A Case Study of Coordination During the 2000 Mozambique Floods." Disasters 27(4): 305-318.

Mullen, B., C. Johnson, E. Salas (1991). "Effects of Communication Network Structure: Components of positional centrality." Social Networks 13(2): pp. 169-186.

Ove, F. (2002). "Using centrality modeling in network surveys." Social Networks 24(4): 385.

Pentland, B. T. (1994). "Grammatical Models of Organizational Processes." CCS Working Paper (176) Sloan WP 3720-94.

Powell, W. W. (1990). "Neither market nor hierarchy: Network form of organization." In B. M. Staw and L. L. Cummings (eds.), Research in Organizational Behavior, Vol. 12. Greenwich, CT: JAI Press, 295-336

Sakurai, S and A. Suyama (2004). "Rule Discovery from Textual Data Based on Key Phrase Patterns" Proceedings of the 2004 ACM symposium on Applied Computing: pp 606-612

SGI (2005). SGI: A site about Language Technology, Email Research and Java http://www.sgi.nu

Shetty, J. and J. Adibi (2004). "The Enron Email Dataset: Database Schema and Brief Statistical Report." Technical report, Information Sciences Institute.

Stephenson-Jr., M. (2004). "Making Humanitarian Relief Networks More Effective: Exploring the Relationships Among Coordination, Trust and Sense Making." Paper prepared for Delivery at the Annual Conference of the Association for Research on Non-Profit and Voluntary Associations.

Wasserman, S. and K. Faust (1994). Social Network Analysis: Methods and Applications. Cambridge, UK, Cambridge University Press.

Weick, K. and K. Sutcliffe (2001). Managing the Unexpected: Assuring High Performance In An Age of Complexity, Jossey-Bass Publishers, San Francisco.

Wikipedia (2005). "Enron Corporation - http://en.wikipedia.org/wiki/Enron."

Zemljic, B. and V. Hlebec (2005). "Reliability of Measures of Centrality and Prominence." Social Networks 27(1): 73-88.