Actor Centrality Correlates to Project Based Coordination

Liaquat Hossain
School of Information Technologies
University of Sydney
NSW 2006 Australia
+61 2 9351 4927
lhossain@it.usyd.edu.au

Andrè Wu
Global Business Consulting
IBM Australia
NSW Australia
+61 404 069 520
awu@ibm.com.au

Kenneth K S Chung
School of Information Technologies
University of Sydney
NSW 2006 Australia
+61 9351 5639
ken@it.usyd.edu.au

ABSTRACT
In this study, we draw on network centrality concepts and coordination theory to understand how project team members interact when working towards a common goal. A text-mining application based on the constructs of coordination theory was developed to measure the coordinative activity of each employee. Results show that high network centrality is correlated with the ability of an actor to coordinate actions of others in a project group. Furthermore, highly centralised actors coordinate better than others. In conclusion, we suggest implications of appropriate network structure for supporting organisational coordination more effectively and efficiently.

Categories and Subject Descriptors
H.1.1 Systems and Information Theory
H.4.3 Communications Applications

General Terms
Design, Human Factors, Theory

Keywords
Social network; coordination; actor centrality; network structure; communication; email content analysis

1. INTRODUCTION
Advances in technology have revolutionised how people share information and coordinate activities in organisations through various forms of organisational structures [9]. As a result, studies on how people coordinate through communication, information-seeking and knowledge transfer have grown rapidly in past years [8, 22]. The socio-technical significance of the relationship between technology, human action and recently, social structure is becoming important for exploring coordination process in organisations [2, 29, 32]. However, empirical research on determinants of coordination for organisational information environments is still not prominent. A key tenet in the study is that understanding the interplay between organisational structure and different types of communications network is important for the design of an effective information based organisation. In this study, we pose questions such as how do we assign coordination role among team members? How do we assign job responsibilities to individuals in an organisation? Do we have any empirical basis to suggest the varied requirement for emerging as a successful coordinator?

In light of the questions above, Pearce and David [30] have suggested that the macro-level characteristics of an organisational design affect group structural properties which in turn directly impacts performance. According to them, the two ends of organisational design continuum are mechanistic and organic. Mechanistic designs deploy the use of highly formalised and specific job descriptions leading to centralisation. Organic designs are less formal and flatter in the organisational hierarchy. According to Pearce and David [30], an organisation is "conceived of as a bounded social system in which there is a relatively stable network of interpersonal linkages through which messages flow and which affect the productivity and maintenance of the system". Drawing on this assumption, a way to study the relationships within the organisation is to explore the formal and informal communication ties, which bind the actors within groups in the organisation.

We utilised the Enron email corpus for various reasons [35]. Firstly, it is a large scale collection of emails from a real organisation. Secondly, the email corpus spans a period of three and a half years. Thirdly, the dataset allows for content, attribute and social network data analysis. Finally, the corpus appeal to researchers who have adopted social networks approach to explore organisational design-performance [1, 2]. After developing our contextual model, we developed a text-mining application in order to extract key phrases and text to calculate the coordination score for each actor involved in the project. We then extracted email addresses from the email headers to facilitate mapping and analysis of social network data for our study.

The novelty of the study lies in its approach to operationalise coordination constructs with a view to measuring coordination empirically. By comparing the contextual network with centrality measures, we explore the correlation between structural position and coordination of actors involved in a major project for Enron corporation. From the results, we argue that high network centrality is correlated with the ability of an actor to coordinate the actions of others in a project group. Furthermore, highly centralised actors coordinate better than others.

In the next section, we introduce concepts of network centrality and its different measures followed by a review of literature on centrality and coordination. A section on hypothesis development is also provided. Next, we discuss the context of the Enron case study for exploring the correlation between centrality and

CSCW'06, November 4–8, 2006, Banff, Alberta, Canada.

coordination. Subsequently, the methodological steps for data extraction and development of coordination scores are outlined, followed by analysis and discussion of the results. Finally, we outline the implications of appropriate network structure for supporting organisational coordination more effectively and efficiently.

2. NETWORK CENTRALITY

In a graph theoretical sense, centrality is referred to as the structural attribute of nodes in a network that dictates its structural position. Centrality has been defined by leading social network researchers as a measure of potential importance, influence, and prominence of an actor in a network [14]. For instance, the concept of centrality indicates the number and strength of ties (degree) that an actor has with other actors. In previous studies, centrality has been measured to capture the flow of information in a network and estimate potential levels of coordination [4, 12, 16]. Tushman [38] argued that people with a high degree centrality have significantly more communication than do those with a lower degree centrality, with professional, technical, and operational areas inside and outside of the organisation.

There are three primary measures of network centrality: (a) Freeman degree centrality, (b) closeness and (c) betweeness. Degree centrality is determined by the number of direct links connecting a node. It highlights the node with the most ties to other actors in the network. The higher an actor's centrality value, the more direct contrast and 'adjacency' that actor has with the other actors in the network [40], and thus, may be seen as an 'index' of potential communication activity [16, 27]. Freeman's degree centrality is considered to be the most basic and easy to calculate [12, 18]. Degree measurements are primarily concerned with local point centrality. Therefore, it is mainly relevant in studies of popularity and activity of actors. Degree centrality measures might be criticised because they only take into account the immediate ties that an actor has, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighbourhood [40]. It is for this reason that centrality is being studied as a precursor to coordination. In the study of coordination, it is useful for measuring local authority. However, we suggest that relying on degree centrality alone may be inappropriate because of its short-sightedness in terms of the overall network structure.

A measure of global centrality is based around what Freeman termed the closeness of the points [16]. Closeness centrality emphasises the distance of an actor to all others in the network by focusing on the geodesic distance from each actor to all others. Two points are connected if there are links between them. These links can span a sequence of lines (hops). The number of these lines to reach the other point is the distance. The sum of these geodesic distances for each actor is the "farness" of the actor from all others. We can convert this into a measure of closeness centrality by taking the reciprocal (one divided by the farness) and normalising it relative to the most central actor [33]. Using this form of measurement, a point is central if it lies within a short distance from many other points, both directly and indirectly connected.

Closeness refers to the extent to which a position is close to all other positions. For example, the position C in the centre of the wheel has a relatively high closeness, because the total distance (i.e., the number of hops) from C to all other positions in the network is relatively small. Positions A, B, D, and E on the periphery of the wheel have relatively low closeness, because the total distance from any one of these positions to all other positions in the network is relatively large. Therefore, closeness represents the potential for independence and efficiency, and signifies a group member who can avoid influence from others [28]. It is also argued that closeness indicates nodes that can spread a message to others in the group in a minimal amount of time. Closeness is the component of centrality which is considered in the original social psychological studies of communication network centrality [4, 24, 34].

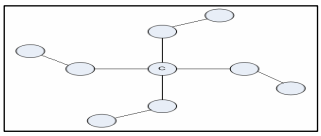

Figure 1. Simple social network showing closeness in node C

Betweeness refers to the frequency with which a node falls between pairs of other nodes in the network. The betweeness centrality of a node i is the number of shortest paths between pairs of other nodes which run through i. Nodes that occur on many shortest paths between other nodes have higher betweeness than those that do not [14]. Betweeness measures usually represent the potential for flow of information or resources between nodes [12]. The rationale behind this measure is that a point with a relatively low degree may play an important 'intermediary' role and hence, very central to the network. Betweeness of a point is measured by the extent to which the agent can play the part of a 'broker' or 'gatekeeper' [33].

In terms of coordination, betweeness may be the most appropriate measure of centrality because of its position in network structure, which allows for a balanced view of the influential control of each node. In figure 2, the actor B with a degree of two of direct connections is considered to be the bridge between two large sub-networks. Here, node B is considered highly central in terms of betweenness and least central in terms of degree.

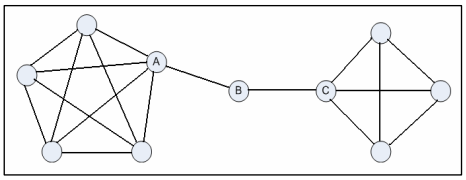

Figure 2. Betweeness highlights the structural bridges or gatekeepers

3. CENTRALITY AND COORDINATION

The formal properties of centrality were first documented in the 1950s by a team at the Massachusetts Institute of Technology (MIT) [4, 24]. The experiment demonstrated a link between network centrality and coordination of small groups. In the late 1970s; Freeman, Roeder and Mulholland replicated the MIT experiment [16] to five zero-history people positioned in different structural formations. By enforcing a strict pattern of permitted communication, the researchers were able to determine that network positions are most conducive to coordination. Post-

experimental interviews indicated that central nodes were most often identified as the leader of the group. In comparing various measures of centrality, it was determined that betweeness was the most useful predicate for coordination.

Coordination is abstract and difficult to quantify. It has been measured using a combination of other factors, such as centrality and the strength of social ties. Coordination is typically measured using qualitative methods, which identify those who have the potential to lead and are able to influence others. If coordination is defined as "managing dependencies between activities" [25], its connotations are broad in network management. Measuring coordination is to measure the effectiveness of key coordination processes as defined by Malone and Crowston [25, 39]. Measurements must be taken as the actor's ability to effectively coordinate scarce resources and also to maximise the impact of these resources. The second process of coordination is to manage the producer/consumer relationships. This means that the coordinator must be able to transfer information between other members of the network. The third process is to manage simultaneity constraints, meaning the coordinator must be able to synchronise tasks between actors. Lastly, effective coordination requires the management of tasks and subtasks to achieve a higher level goal.

Traditional approaches to coordination tended to focus on the delegation of more authority to a single actor. The idea was to secure coordination by control from the top, a 'coordination by command' approach [10]. This notion has been a contentious issue due to 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 [27, 37]. Larry Minear, a leading researcher in this field has argued that "coordination is multilayered, involving the orchestration of relationships not only at headquarters but also at the regional, national and field levels" [26]. Hierarchical models of resource allocation and coordination emphasise the power and dependencies, which develop during interorganisational transactions. Research within resource dependency has also argued that there is a positive association between organisations' network centrality and their supposed influence in community affairs [17]. The central and more visible actors in the network are more likely to be potential allies for other powerful actors, and thereby appearing even more powerful [5, 17]. To measure the effect of centrality on these coordination processes, data can be collected on the actors, which enact these coordination processes. We can then compare this data to its relative centralities to determine if a correlation exists. It has been argued that centrality measures are an explanatory variable in studies of actor influence and actor control [13, 16]. In our study, we seek to answer asking 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?

3.1 Hypothesis One

Individuals centrally positioned in a network are more likely to emerge as coordinator.

Freeman [16] defined three measures -- degree, betweeness and closeness -- of centrality and explained their structural implications for the study of centrality and information flow in organisations. Freeman suggests that the degree of a point seemed to be an index of that position's potential for activity in the network. Betweeness can be taken to be an index of potential for control of communication. Closeness measures the distance of a point to all others. According to Freeman [14] these kinds of centrality imply three competing theories of how centrality might affect group processes, as shown in table 1.

Table 1. Centrality Measures and its social implications [15]

Measure	Social Implication
Betweeness	Control
Degree	Activity
Closeness	Independence

Using data from thirty years of major social psychological journals, Mullen, et. al., [28] performed a meta-analysis of the effects of centrality on groups. They studied three aspects of individual behaviour as part of a group task; leadership, satisfaction, and messages sent. They concluded that betweeness centrality was the most powerful independent predictor of the effects of centrality on leadership. Furthermore, individuals situated in most centralised position of a network (i.e. highly dense betweeness) were likely to emerge as the leader, to be more satisfied, and to participate more in the task solution. This further indicates that the potential for control of communication seems to be uniquely important in the development of leadership in communication networks. These behavioural characteristics support those of the MIT study [24], and of the study by Freeman et. al. [16].

3.2 Hypothesis Two

In undirected graphs, betweeness is the best predictor for coordination.

A social network can be directed or undirected. A network is directional if the ties are oriented from one actor to another [40]. Undirected graphs indicate the presence of a relation but do not specify the direction of the relationship. In a directed graph, the nodes can be adjacent to or adjacent from another node depending on the direction of the relation. In social network applications, out-centrality measures the expansiveness and in-centrality measures the receptivity or popularity [40]. In-centrality can be regarded as an indicator for the prominence of an actor, whereas out-centrality measures the influence of the actor.

3.3 Hypothesis Three

In directed graphs; out-centrality is a better predictor for coordination ability.

With an understanding of the implications of coordination and centrality, we describe how these hypotheses of the study are tested in the next section. Using the Enron corpus dataset, we performed text mining techniques based on established coordination theories to measure the coordinative activity of each employee within the project scope. By comparing the contextual network with centrality data, we explore the correlation between structural position and coordination.

4. CONTEXT OF THE STUDY

In this study, we explore the correlation between network centrality and coordination using the email correspondence between Enron employees working on the Dabhol Power project. Enron Corporation was an energy trading, natural gas, and electric utilities company based in Houston, Texas that 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 and released the email communication logs to the public.

In 1997, Enron created 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 involved two phases: (i) the construction of the power plant [740 megawatts] and (ii) the eventual expansion of its output capacity was the second phase [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.

In October 2003, over half a million emails sent by Enron employees were made public and posted on the Internet by the Federal Energy Regulatory Commission (FERC) during its investigation. There are at least three versions of Enron dataset available. In this study, we performed the analysis on the MySQL database prepared by Jitesh Shetty at the University of Southern California (see figure 3). The dataset was cleaned and all personal and empty messages were removed. The cleaned Enron corpus contains 252,759 messages belonging to 17,568 users. The dataset was used to extract evidence of coordination and to perform centrality measurements. In the next section, we describe the process of extraction and analyses of the Enron corpus.

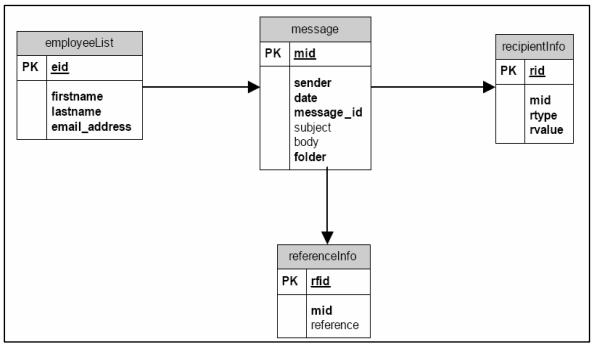

Figure 3: Enron Database Schema [35]

5. DESIGN OF THE STUDY

The study was performed on the Enron dataset as prepared by the team at USC (University of Southern California). This dataset is in MySQL format and is flexible for data extraction. The sampling frame used in the study was the emails about the Dabhol Power Company project. The motivation for studying a particular project was that the messages would be more likely to show coordination as the employees would be working towards a common goal. This motivation builds on the tradition set by communication-network researchers such as Ahuja, Galletta and Carley [3] and Adamic and Adar [1]. The methodology of the study involves four research phases:

1. Cataloguing of coordination key phrases
2. Calculation of coordination score based on project scope
3. Construction of social network matrices using the centrality measures, and,
4. Exploring the association between network centrality and coordination score

5.1 Cataloguing of Coordination Key Phrases

For this study, text-mining techniques were applied to measure coordination. Text mining is the application of analytical functions relying on sophisticated text analysis techniques that extract information from free-text documents [11]. This section begins with the existing literature on grammatical models and coordination theory.

In a study of grammatical models of organisational processes, Brian Pentland suggested a variety of coordination constraints based on the kinds of interdependencies between actions [25]. In organisational theory, it is difficult to construct a grammar that could sustain a rigorous analogy to the structure of human brain. Coordinative action is historically situated, culturally embedded, and generally stands in a recursive relation to action [19]. 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 [31]. Due to the lack of a universal grammar, this study used a context specific taxonomy. For construct validity, the taxonomy was compiled directly from the Enron email corpus.

The process of coordination was broken down into four key coordination processes as defined by Malone and Crowston [25] and operationalised for the study of the email corpus. The four processes along with the interpretations are shown below:

- 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(s). Allocating a time for a particular event(s). Passing information about the time of an event(s).
- Managing task / subtask dependencies: Planning tasks and strategy to achieve the overall goal.

Using text-mining techniques, these four processes were operationalised into key phrases for the extraction from email datasets. The process of text mining consists of three stages [23]: (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). These processes were identified within a project scope.

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 project. The intention was to improve data validity by using a different training set from the primary data for the study. The model building phase to compile the list of keywords was broken down into a few steps. The first of these steps was the extraction of sentences indicative of one of the processes of coordination. Each sentence was categorised into the specific coordination process and catalogued (see figure 4).

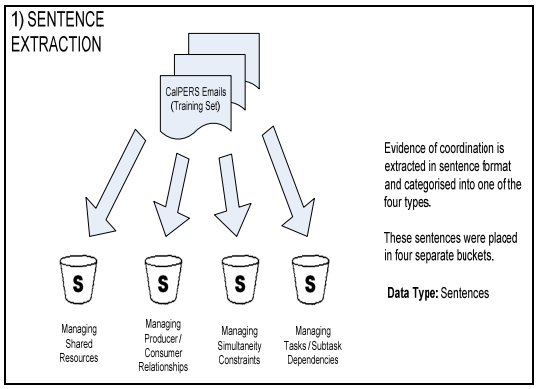

Figure 4. Sentence Extraction Process

In the second step, the list of sentences was sorted through and the key phrases which underlie the coordinative action were identified and marked (see figure 5). In the third step, 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 (see figure 6). 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 highest average weight was 4.1. The weight was measured using the Dabhol project. This was done to ensure accuracy in allocating the weights. The final list of coordination phrases and their respective weights are shown in table 2 below. For the deployment of the model, the key phrases were matched among the dataset.

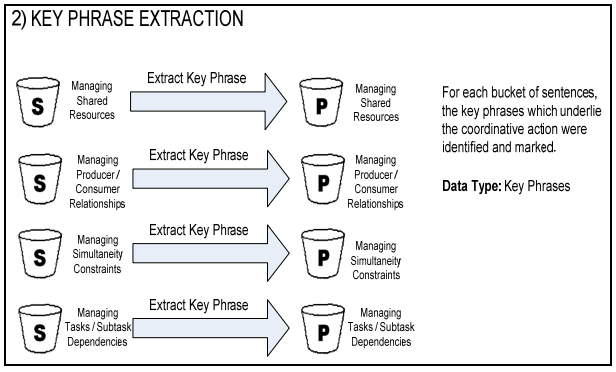

Figure 5. Process for Extracting Key Phrases

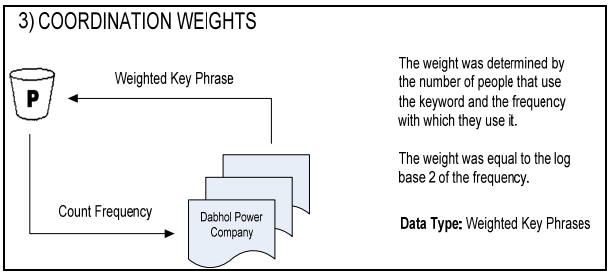

Figure 6. Process for Assigning Coordination Weights

Table 2. Coordination Key Phrases

Coordination key Phrase	***Weight***		
Resource Allocation		**Simultaneity Constraints**	
Help coordinate	3.28	As we move closer	2.63
Please allow	2.35	Please allow time	1.57
Please communicate	4.28	On track	7.15
Please coordinate	5.61	Sufficient time	2.26
Please do	5.97	Take the time	4.24
Please get	3.61	Agenda	7.79
Please make sure	2.80	On time	4.85
Do you want to	2.00	Make a timetable	6.00
I request	1.58		
I would appreciate	4.95	**Tasks/Subtasks**	
I would like to	6.49	I have considered	1.65
I would like your	1.87	I recommend	3.16
I would ask	2.31	I suggest	2.32
Look into	7.46	I wanted to	4.45
Make sure that	5.67	I would like to	6.49
Please see	4.58	I would suggest	2.32
Please speak	4.18	We can discuss	3.32
Please work	3.54	We can then	1.82
Put this together	2.65	We have seriously	1.42
ProducerRelationships			
Are as follows	5.12	Let me know if	6.93
Attached is a	4.64	Please let me know	6.89
Attached please find	2.32	We have had	4.16
The bottom line is	4.58	Would probably be	4.08
The purpose is	1.58	I believe you are	2.16
For your information	3.90	Which brings me to	1.65

5.2 Calculation of Coordination Score Based on Project Scope

The key phrases from the list above were extracted from the Dabhol project dataset. We built a text-mining application to iterate through the list of keywords and performed SQL queries based on the project scope (see figure 7). The coordination score of each person was collated by aggregating the weighted scores based on keyword matches from the list above. The text-mining program outputted a list of coordinators and their total coordination score (see figure 8).

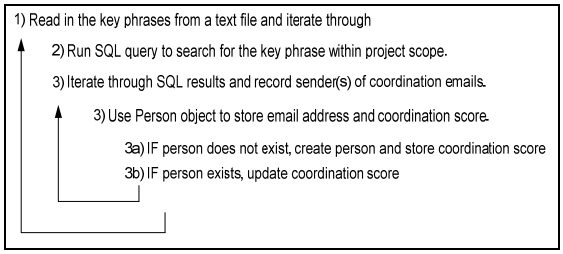

Figure 7. Pseudocode for the text-mining application to calculate coordination score

Name	Res. Alloc.	Prod/Cons Rel.	Simul. Con.	Tasks/Sub	Total Coord. Score
wade.cline@enron.com	25	24	16	11	76
steven.kean@enron.com	45	20	25	6	96
loretta.brelsford@enron.com	36	22	8	18	84
sandeep.kohli@enron.com	10	4	0	8	22
jonathan.whitehead@enron.com	4	0	0	4	8
...					

Figure 8. Sample output of Coordination Score by text mining program

5.3 Construction of Social Network Matrices using Centrality Measures

Centrality was measured by the number of emails sent and received by each person as part of the Dabhol project. Centrality is defined by 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 measured using Ucinet 6 [6]. This methodology studied centrality based on two axes; measurement, and directional analysis. Three centrality measures were used in this study; degree, closeness and betweeness. Flow betweeness 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. Figure 9 shows a sample of the undirected centrality index scores. The full data contained 712 employees.

Name	Betweenness	Rank	Closeness	Rank	Degree	Rank
wade.cline@enron.com	1519.539	1	129.5	3	70	10
steven.kean@enron.com	1475.576	2	120.456	6	92	6
loretta.brelsford@enron.com	1458.313	3	130.683	2	89	7
sandeep.kohli@enron.com	1243.107	4	104.543	12	43	13
jonathan.whitehead@enron.com	1058.002	5	83.595	31	17	36
...						

Figure 9. Undirected centralisation index scores and ranks for individuals

A directional analysis was performed to determine the correlation of in and out-centrality on coordination. Directed degree was measured by counting only the number of emails sent by and received by a person. In-centrality seems to indicate the prominence of an actor, whereas out-centrality measures the influence of the actor. The plus/minus was calculated to determine whether there was a relationship between coordination and skew directed centrality. Figure 10 shows a sample of the directed centrality index.

Name	outDegree	Rank	inDegree	Rank	outCloseness	Rank	inCloseness	Rank
wade.cline@enron.com	32	12	52	3	81.967	4	47.533	1
steven.kean@enron.com	31	14	73	1	74.756	8	45.7	2
loretta.brelsford@enron.com	88	4	12	25	99.083	3	31.6	18
sandeep.kohli@enron.com	36	11	9	31	77.9	6	26.643	30
jonathan.whitehead@enron.com	15	23	5	40	60.75	23	22.845	42
...								

Figure 10. Directed Centralisation index scores and ranks for individuals

5.3.1 Data Cleansing

The resulting centrality data was cleansed to separate the noise from the general pattern of data. Certain steps were taken to factor for the increased probability that a person who sent a high volume of emails would have an increased coordination score. The ratio between in and out-degree was measured for each person. This was titled the noise ratio because those with a low ratio would distort the data. The employees with a noise ratio less than 0.05 were removed from the dataset. This was done to reduce the effect of mass-mailers sending company announcements. Within the Enron dataset, it was common for a few email addresses to send out news updates and stock quotes. As an additional method of data cleansing, a threshold was introduced and those people with fewer than three in or out-degree were removed. Those addresses were not used very often and their relevance is questionable.

5.4 Exploring Association Between Network Centrality and Coordination Score

In this phase, we explore the association between coordination score and centrality measurements. We performed two types of tests—(i) parametric and (ii) non-parametric. Parametric tests assume the normal distribution (bell-shape) of scores. Non-parametric tests do not assume a normal distribution. Due to some high outliers, it was determined using the Shapiro-Wilks W test that coordination scores did not follow a normal distribution. Because the results were not normally distributed, non-parametric tests were introduced. However, the central limit theorem states that when a dataset is large (eg. n>80) the mean will follow the normal distribution even if the respective variable is not normally distributed in the population. As a result, both parametric and non-parametric tests were used for the hypothesis testing.

The three hypotheses tested were that; (H1) higher centrality is correlated to an increase in coordination ability; (H2) betweeness is the best measurement for predicting coordination and that (H3) out-centrality is a better predictor than in-centrality. The hypothesis testing determined whether the means of two groups are statistically different from each other. For this study, the two groups were dichotomised as those 'high' and low' in centrality

based on the median. The coordination scores from the group high in centrality were compared to those low in centrality.

The Mann-Whitney U test was then used to determine if there was a substantial difference between two groups of data. The Mann-Whitney U test is non-parametric and does not assume a normal distribution. H1 is a one-tailed hypothesis, with a statistical significance level set at 0.05 (5%). The t-test was used to determine whether the means of two groups are statistically different from each other. The strength of the t-test is that it factors for the variability and dispersion of the scores. The alpha level was set at 0.05 for all measurements. The t-test is parametric, meaning the normal distribution is assumed. Figure 14 depicts the process for calculating statistical significance.

6. RESULTS

This test studies coordination using all three project scopes, that is, the Dabhol project. The emails were extracted from the dataset based on a keyword match of the company name as well as the common names and companies associated exclusively with the project. Emails containing at least one of these keywords and all reply correspondences in that thread were included in the project scope. Using these project names and associations, the employees found to have either sent or received an email matching one of these was extracted. Seven hundred and twelve people were extracted as part of the Dabhol project scope. This list was loaded into UCINet's NetDraw module for a visual representation. Figure 11 shows the social network for the Dabhol project. The figure is colour coded based on the betweeness index. It can be seen that the majority of the nodes show a betweeness score of zero. The coordination for each employee was calculated using the text-mining application as defined in phase one and two. From this set, 173 people were found to have demonstrated coordination in the Dabhol scope. The coordination scores ranged from 3 to 244 with an average of 44.

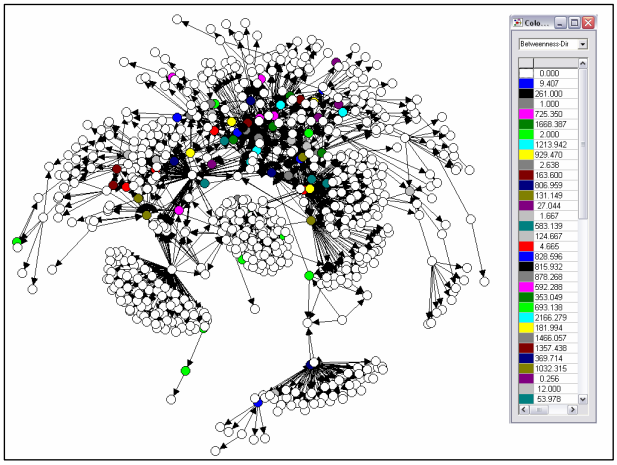

Figure 11. Social Network of Dabhol Project showing betweeness index

As part of the data cleansing as described in the previous section, 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. This process removed 85 nodes. A further attempt to separate the noise from the general pattern was performed by calculating the noise ratio. Using the noise ratio, people with a low in-degree in comparison to out-degree were removed from the dataset. This was done to reduce the effect of mass-mailers sending company announcements. In the Dabhol dataset, the noise ratio eliminated a further eight nodes. From the original 173 nodes, the cleansing process eliminated 92 nodes, leaving 80 employees with genuine data. The distribution of the coordination score is shown below (Figure 12). A quick investigation of the score distribution shows a right-skew with a long right tail.

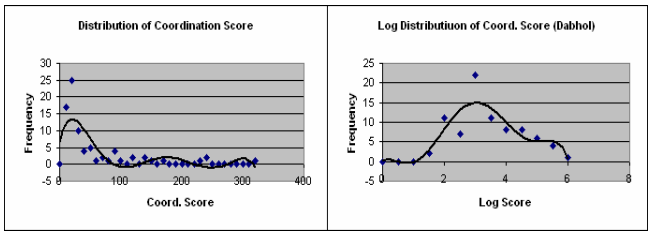

Figure 12. Distribution and Log Distribution of Coordination Score (Dabhol)

Using these coordination scores, the process was to determine if there was a substantial difference in coordination between people with high and low centrality. This was done by splitting the data into two groups ordered by centrality and testing the statistical difference in coordination between the high and low groups. To demonstrate the process, the operation is presented in full for the betweeness measure of the Dabhol project. This is shown in figure 13. Data from the operation in figure 13 revealed a U test p-value of 0.0029. This data was entered into table 3 and the operation was repeated for the degree and closeness centrality indexes.

The three main measurements of centrality (i.e., betweeness, degree and closeness) shows a one-sided P of less than 0.05; the alpha level used for this study. 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).

Table 3. Results for undirected graphs using the Mann-Whitney U Test (5% significance)

Network Centrality Measure: Dabhol Project	
Betweeness	**P = 0.0029**
Degree	**P = 0.0043**
Closeness	**P = 0.0073**

Figure 13 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. The second hypothesis is tested by comparing the p-values between the three measures of centrality. The measure with the lowest p-value is taken to have the strongest statistical significance because it has the least chance of obtaining those numbers by chance alone.

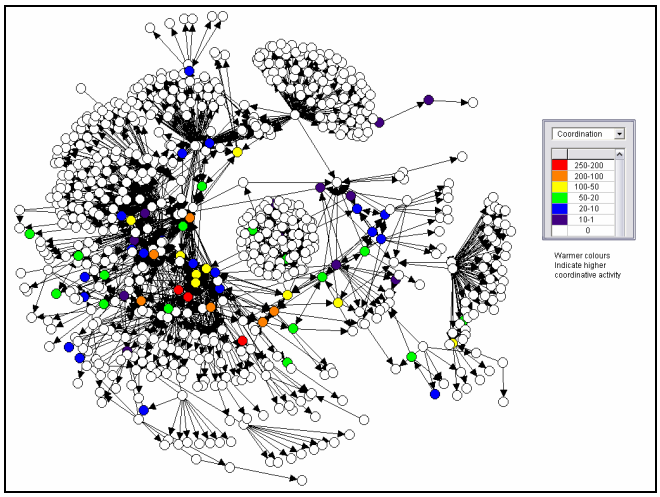

Figure 13. Sociogram showing Coordination in Dabhol Project

The test on directed centrality and coordination investigates the correlation between in and out-centrality to coordination. The hypothesis to test is that out-centrality measurements have a stronger correlation to coordination. Due to the non-normal distribution of the data, non-parametric tests must be used. The Spearman Rank Test was used to measure correlation. The r-estimate is the correlation found using the Spearman Rank test. The p-value indicates the probability of receiving those r-estimates by chance alone. In line with the other experiments, the significance level used a p-value of 0.05 (see table 4). The Dabhol project show clear results highlighting that out-centrality measurements correlated with coordination much stronger than did the in-centrality counterparts (i.e. the p-value is found to be significant, and thus all the measures are accepted as evidence).

Table 4. Results for directed graphs in the Dabhol Project using the Spearman Rank Correlation

Out-Degree	**r= 0.364; p= 0.000882**
In-Degree	**r= 0.268; p= 0.015915**
Out-Closeness	**r= 0.345; p= 0.001748**
In-Closeness	**r= 0.262; p= 0.000959**

7. DISCUSSION

All tests of undirected 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 another example of the power of social networks in affecting our day-to-day interactions. These results are in line with the findings of Freeman, Roeder and Mulholland that network position and centrality has an effect on coordination [16]. This is interesting because the study by Freeman et al. was experimental using groups of five. This study found similar results with data-driven approach and with a group size of over seven hundred.

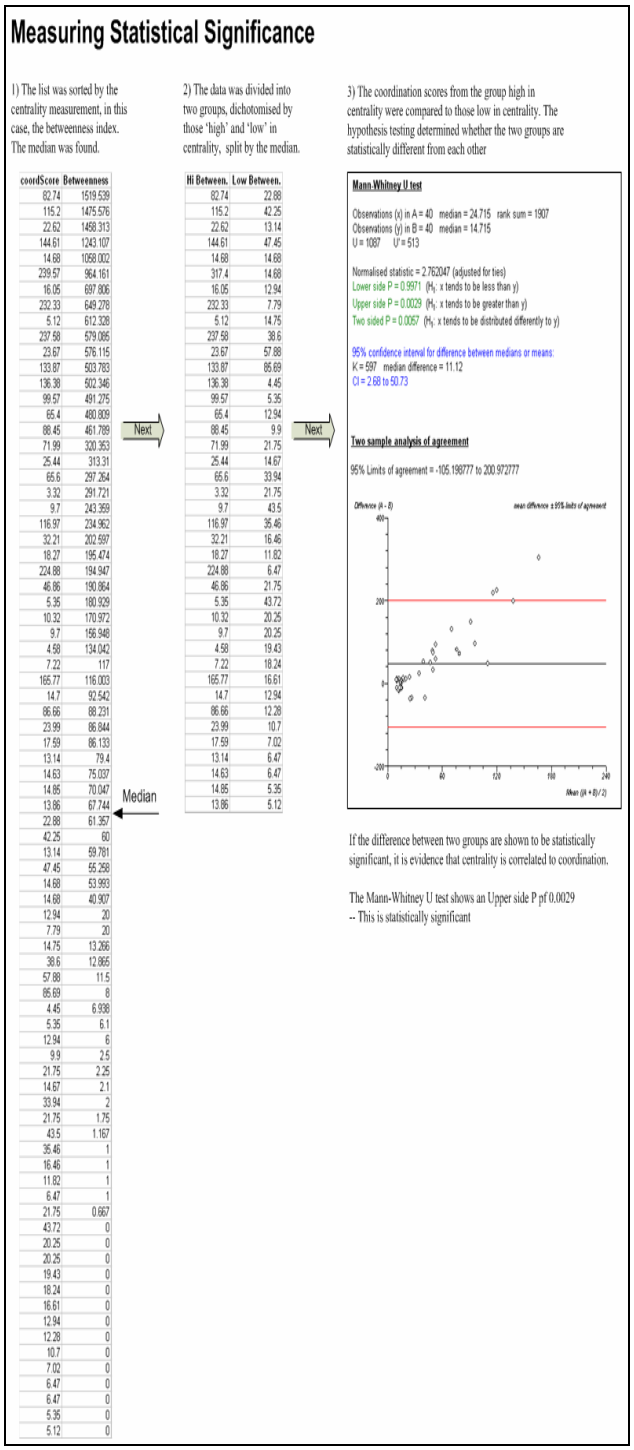

Figure 14. Process for Calculating Statistical Significance

In this study, betweeness was shown to be the best independent predictor for coordination. It makes sense that a person with a high betweeness index would be able to coordinate because the position dictates a role of leadership. The person would lie between many other people and thus have greater influence through the weak links. According to Granovetter, it is through these weak links that the diffusion of information and influence

are able to spread [21]. These findings support the work of Mullen, Johnson and Salas; who stated that the betweeness index is the best predictor for the effects of centrality [28]. This indicates that the effect of centrality is an increase in coordination ability. It was also shown that closeness was the weakest in terms of statistical strength, this supports the findings of Freeman et al. [16].

A potential concern about this study is the comprehensiveness of the keywords in measuring the ability to coordinate. Pentland [31] argued that the lack of a universal organisational grammar meant that processes such as coordination are best studied using a context specific taxonomy. To cater for this, the context specific taxonomy was built directly from the Enron corpus. A limitation is that the dataset had some integrity errors with the email addresses. This study extracted the 'From' and 'To' fields from the email headers to measure centrality and the coordination score. In some instances, a single employee could have several email addresses. For example, vince.j.kaminski@enron.com, vince.kaminski@enron.com and vincent.j.kaminski@enron.com all belong to the same person. For this study, they were treated as discrete nodes.

A limitation of the study is that non-email communications are not reflected in the data. However, it has been shown that the email network is fairly representative of the underlying communication network within people in an organisation. A recent study in the UK found that email was used to communicate with 80% of one's social network for the 25–35 age group [60% for the 50–60 age group] [36]. This is particularly true for Enron, a company striving to set itself amongst the leaders in technology utilisation.

Enron had a culture of ultra-competitiveness with other organisations and between project groups. This may fetter effective coordination as ambitious employees seek to serve their own gain. However, considering the competition was generally between project groups, the data from this study should be relatively unharmed from this because the tests look at a single project. It is for this reason that projects are an appropriate context to study coordination. The methodology of this study was conducted in a generalised way so as to assist the extensibility of the results.

8. CONCLUSION

Although the collapse of Enron may be attributed to a plethora of deficit-based reasons that relate to mismanagement, deceit, manipulation, and the like, we utilise social network approach to learn about the role of informal organisational network structure and the successful coordination of one of Enron's multinational projects. Building on existing coordination and organisational process theory, we designed and implemented a text mining tool to measure coordination from a large dataset of emails. Using this method, we were able to demonstrate with statistical significance a relationship between network centrality and coordination. Not only has this study extended the work of Pearce and David's [30] with empirical evidence to show that centrality as a group structure property is directly correlated with coordination, but we also demonstrate that the relationship was further defined in that betweeness centrality was shown to be the best predictor for centrality. In directed graphs, it was shown that out-centrality measures are better predictor for centrality than in-centrality measures. We can conclude that centrally 'well-connected' people are able to exercise greater coordination within the network structure. The implication of these inferences mean that for organisations to design an information environment where coordination and knowledge sharing is facilitated, it is important to account for the social structure and position in designing and mapping coordinated groups. In particular, those who broker information between other people indicate a higher ability to coordinate because their position dictates a role of leadership in the social structure. We conclude that centrality is an important social network property that moderates the relationship between organisational design and building an information environment.

9. REFERENCES

[1] Adamic, L. and Adar, E. How to Search a Social Network. *Social Networks*, *27* (2005), 185-203.

[2] Ahuja, M.K., Carley, K. and Galletta, D.F., Individual Performance in Distributed Design Groups: An Empirical Study. in *Proceedings of the 1997 ACM SIGCPR conference on Computer personnel research* (San Francisco, California, United States, 1997), ACM Press, NY, 160-170.

[3] Ahuja, M.K., Galletta, D.F. and Carley, K.M. Individual Centrality and Performance in R&D Groups: An Empirical Study. *Management Science*, *49*,1 (2003), 21-38.

[4] Bavelas, A. Communication Patterns in Task-Oriented Groups. *Journal of Acoustical Society of America*, *22* (1950), 725-730.

[5] Bonacich, P. and Lloyd, P. Eigenvector-like Measures of Centrality for Asymmetric Relations. *Social Networks*, *23*,3 (2001), 191-201.

[6] Borgatti, S.P., Everett, M.G. and Freeman, L.C. Ucinet for Windows: Software for Social Network Analysis, Analytic Technologies, Harvard, MA, 2002.

[7] Chatterjee, P. Enron Deal Blows a Fuse. *Multinational Monitor*, *16* (1995), 7-8.

[8] Cummings, J. Work Groups, Structural Diversity and Knowledge Sharing in a Global Organization. *Management Science*, *50* (2004), 352-364.

[9] DeSanctis, G. and Jackson, B.M. Coordination of Information Technology Management: Team-Based Structures and Computer-Based Communication Systems. *Journal of Management Information Systems*, *10*,4 (1994), 85-110.

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

[11] Dörre, J. and Gerstl, P., Text Mining: Finding Nuggets in Mountains of Textual Data. in *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (1999), 398-401.

[12] Faust, K. Centrality in Affiliation Networks. *Social Networks*, *19*,2 (1997), 157-191.

[13] Frank, O. Using Centrality Modeling in Network Surveys. *Social Networks*, *24*,4 (2002), 385-394.

[14] Freeman, L.C. Centrality in Social Networks: Conceptual Clarification. *Social Networks*, *1*,3 (1978), 215-239.

[15] Freeman, L.C., Roeder, D. and Mulholland, R.R. Centrality in Social Networks: ii. Experimental Results. *Social Networks*, *2*,2 (1980), 119-141.

[16] Freeman, L.C., Roeder, D. and Mulholland, R.R. Centrality in Social Networks: ii. Experimental Results. *Social Networks*, *2*,2 (1979), 119-141.

[17] Galaskiewicz, J. Interorganizational Relations. *Annual Review of Sociology*, *11*,1 (1985), 281-304.

[18] Garton, L., Haythornthwaite, C.A. and Wellman, B. Studying Online Social Networks. *Journal of Computer Mediated Communication*, *3*,1 (1997).

[19] Giddens, A. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press, Berkeley, 1984.

[20] Granovetter, M. *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press, Chicago, 1995.

[21] Granovetter, M.S. The Strength of Weak Ties. *The American Journal of Sociology*, *78*,6 (1973), 1360-1380.

[22] Guetzkow, H. and Simon, H. The Impact of Certain Communication Nets upon Organization and Performance in Task-Oriented Groups. *Management Science*, *1*,233-250 (1955).

[23] Han, J. and Kamber, M. *Data Mining: Concepts and Techniques*. Morgan-Kaufman, New York, 2000.

[24] Leavitt, H. Some Effects of Certain Communication Patterns on Group Performance. *Journal of Abnormal and Social Psychology*, *46* (1951), 38-50.

[25] Malone, T.W. and Crowston, K., What is Coordination Theory and how can it help design Cooperative Work Systems? in *CSCW 90 Proceedings*, (Cambridge, MA, 1990), Massachusetts Institute of Technology.

[26] Minear, L. *The Humanitarian Enterprise: Dilemmas and Discoveries*. Kumarian Press, Bloomfield, 2002.

[27] Moore, S., Eng, E. and Daniel, M. 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 (2003), 305-318.

[28] Mullen, B., Johnson, C. and Salas, E. Effects of Communication Network Structure: Components of Positional Centrality. *Social Networks*, *13*,2 (1991), 169-185.

[29] Orlikowski, W. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, *3*,3 (1992), 398-427.

[30] Pearce, J.A.I. and David, F.R. A Social Network Approach to Organizational Design-Performance. *Academy of Management Review*, *8*,3 (1983), 436-444.

[31] Pentland, B.T. Grammatical Models of Organizational Processes. Sloan WP 3720-94, CCS Working Paper number 176, 1994.

[32] Rice, R.E., Grant, A., Schmitz, J. and Torobin, J. Individual and Network Influences on the Adoption and Perceived Outcomes of Electronic Messaging. *Social Networks*, *12*,1 (1990), 27-55.

[33] Scott, J. *Social Network Analysis: A Handbook*. SAGE Publications, London, 2000.

[34] Shaw, M. Communication Networks. in Berkowitz, L. ed. *Advances in Experimental Social Psychology*, Academic Press, New York, 1964, 111-147.

[35] Shetty, J. and Adibi, J. The Enron Email Dataset Database Schema and Brief Statistical Report, Los Angeles, CA, n. d.

[36] Smith, H. and Rogers, Y. *Managing One's Social Network: Does Age Make a Difference?* Kluwer Academic, 2003.

[37] Stephenson-Jr., M. 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* (2004).

[38] Tushman, M.K. Special Boundary Roles of the Innovation Process. *Administrative Science Quarterly*, *22* (1977), 587-605.

[39] Vick, R.M. and Crosby, M.E. Assessment of Resource Coordination Effectiveness Through Analysis of Distributed Cognitive Traces in Team Decision Making. *25th Annual Meeting of the Cognitive Science Society, CogSci 2003 (Poster Paper)* (2003), 1182-1187.

[40] Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, 1994.

[41] Weick, K. and Sutcliffe, K. *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. Jossey-Bass Publishers, San Francisco, 2001.

[42] Zemljic, B. and Hlebec, V. Reliability of Measures of Centrality and Prominence. *Social Networks*, *27*,1 (2005), 73-88.