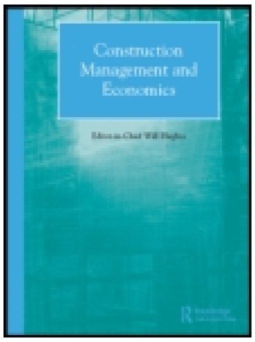


Routledge
Taylor & Francis Group

Construction Management and Economics

ISSN: 0144-6193 (Print) 1466-433X (Online) Journal homepage: https://www.tandfonline.com/loi/rcme20

Communications and coordination in construction projects

Liaquat Hossain

To cite this article: Liaquat Hossain (2009) Communications and coordination in construction projects, Construction Management and Economics, 27:1, 25-39, DOI: 10.1080/01446190802558923

To link to this article: https://doi.org/10.1080/01446190802558923

Published online: 25 Feb 2009.

Construction Management and Economics (January 2009) **27**, 25–39

Communications and coordination in construction projects

LIAQUAT HOSSAIN*

Project Management Graduate Programme, School of Civil Engineering, University of Sydney, Sydney, Australia

Received 27 April 2008; accepted 15 October 2008

Coordination can be seen as a process of managing resources in an organized manner so that a higher degree of operational efficiency can be achieved for a given project. Social network matrices are constructed using different centrality measures. These measurements are used to explore the association between network centrality and coordination for a construction project. Network 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 coordination? How is the actor's ability to coordinate projects related to his or her structural position in the communications network? Multi-layered test designs were developed to explore this relationship in a project-based coordination of Dabhol Power Company Construction company and Azurix Corporation. There are three major findings from this analysis. First, centrally positioned actors show more coordinative activity. Second, the betweenness index of centrality is the most potent predicate for coordination. Last, the influence of an actor is associated with coordination more than the actor's prominence.

Keywords: Communications, coordination, network centrality, social network analysis.

Coordination in construction

Construction can be viewed as a process of putting together all the materials in an orderly and timely manner by utilizing relevant resources to complete a structure as per designed specifications and quality standards. Building construction refers to the process of adding structure to real property such as residential, commercial, educational institution, hospital and sporting arena. Civil construction is adding infrastructure such as roads, railways, airports, harbours, dams, mines and golf courses to the environment, and industrial construction deals with large projects such as pharmaceutical, petroleum, chemical, power generation and manufacturing complexes. Previous studies on construction management highlight the need for understanding complexity factors for the construction industry which deals with uncertainties and interdependences, and inefficiency of operations (Dubois and Gadde, 2002).

The process of construction, depending on the complexity of the finished structure, requires a high level of coordination among all the professionals and trade persons from design office to the construction site until the project is completed. The persons involved at the higher management team consist, primarily, of the project manager, the construction manager, the commercial manager, the project control manager, the quality manager, the design manager, the environmental and heritage manager and the architect. Under them are the front line superintendent and supervisory staff which consist of the construction engineers, the project control engineers, the project architect, the quality engineer and the electrical and services engineer.

However, the most common types are design-bid-build, design and build and management contracting. In the design-bid-build arrangement, the architect or engineer acts as the project coordinator with a role of designing the works, preparing the specifications and producing the construction drawings, administering the contract, tendering out the works, and managing the works from inception to completion. In the design and build approach, the owner produces a list of requirements for a project, giving an overall view of the project's goals. The contractors would present different

*E-mail: lhossain@usyd.edu.au

Construction Management and Economics
ISSN 0144-6193 print/ISSN 1466-433X online
http://www.tandf.co.uk/journals
DOI: 10.1080/01446190802558923

ideas about how to accomplish these goals and the owner would select the preferred idea and hire the contractor or consortium. This is in contrast to a design-bid-build contract, where the project is completely designed by the owner, then bid on, then completed. In the management contracting arrangement, the client could have an active role in the procurement system, by entering into separate contracts with the designer, the construction manager, and individual trade contractors if so desired. Therefore the client could take on the contractual role, while the construction or project manager provides the active role of managing the separate trade contracts, and ensuring that they all work smoothly and effectively together or the client can have a manager to manage it all on their behalf.

We therefore focus our effort in determining the effects of network centrality on coordination. Building upon established coordination theories, we investigate differences in coordinative activity between individuals with high and low network centralization. The central research question may thus be phrased as: Are centrally 'well-connected' people able to exercise greater coordination within the network structure? Does an individual's 'potential for communicative activity' become reflected in its actual coordination score? Do the data reflect the theories of centrality and coordination? We investigate the following sub-problems for this study:

(1) whether high network centrality increases the ability to coordinate;
(2) whether low network centrality decreases the ability to coordinate.

If (1) and (2) hold true; then we determine:

(a) which measure of centrality is the most potent predicate for coordination; degree, closeness or betweenness;
(b) whether in- or out-centrality correlates closer to 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 synthesized 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 a more important or more powerful predictor 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, an attribute not of actors themselves, but of their structural position in the network. Measures of centrality are based on attributes such 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 *et al.*, 2002). 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). Freeman (1979) defined three measures of centrality and explained their structural implications. The three measures identified were degree, betweenness and closeness. Freeman stated that the degree of a point 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.

Coordination is an abstract concept that is difficult to measure 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. 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 because of the difficulties in selecting a suitable governing body. Contemporary thinkers contend that it may be timely to consider whether an organization should be reconceived as constituting a social network (Moore *et al.*, 2003; Stephenson, 2004).

Minear (2002) argued that 'coordination is multi-layered, 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.

Measuring coordination through centrality

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 *et al.* (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 centralization 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 of participants identified the most central position as the leader. The study found evidence that centrality is associated with leadership. We take this finding further 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 centralization (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 centralized 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.

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 hypotheses were presented to outline the origins of the propositions. 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 organizational 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 are collected on the actors that enact these coordination processes and their relative centralities are further analysed to determine if a correlation exists.

Coordination dataset of Dabhol Power Corporation and Azurix 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. 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 caused violent protests owing to the environmental impact of the new plant. The project faced another problem in that the newly elected local 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 (1444 megawatts). By May 1999, phase one was completed and the DPC reported profits of $42 million during the first year of its operations. However, phase two of the project stagnated and in December 2001, Enron filed for Chapter 11 bankruptcy before the project was completed. 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). 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 2 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.

E-mail communications and coordination logs from the Enron Corporation between 1997 and 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 e-mail datasets that can be used for research purposes. It is more extensive than any other research-friendly e-mail corpus by several orders of magnitude (SGI, 2005). There are at least three versions of the 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 e-mails 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 organizational hierarchy. To accommodate this, the role of each employee was added to the MySQL database.

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 e-mails are more likely to 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 minimize 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 e-mail dataset using a keyword match on the project name, along with any replies to those e-mails. 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.

Coordination key phrases (model building)

The study of coordination requires a clear definition and a standard method of measurement. The challenge

Table 1 Project names and alternate association

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

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 organizations 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 organizational processes, Brian Pentland (1994) suggested a wide variety of coordination constraints based on the kinds of interdependencies between actions (Malone *et al.*, 1993). In organizational 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 organizational 'language faculty' eliminates the possibility of a universal grammar for organizational processes: a single set of universal rules or principles that govern the syntactic structure of all organizing processes (Pentland, 1994). Owing 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 e-mail 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 further builds on existing coordination theory and presents a new approach for exploring organizational processes. The underlying assumptions of coordination theory are utilized. 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. Using these discrete instances, it becomes possible to measure and compare levels of coordination. The process of coordination was broken down into four key coordination processes as defined by Malone and Crowston (1994). The four processes were then interpreted and operationalized for the study of the e-mail corpus. The four processes along with the interpretations are shown below:

(1) Managing shared resources
- Instructing or suggesting a person to perform a task.

(2) Managing producer–consumer relationships
- The creation or dissemination of information.

(3) Managing simultaneity constraints
- Synchronizing tasks between actors.
- Taking possible times for an event.
- Allocating a time for a particular event.
- Passing information about the time of an event.

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

Using text mining techniques, these four processes are operationalized into key phrases to be extracted from the e-mail 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).

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 and Azurix project datasets; this is further described in 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 categorized into the specific coordination process and catalogued. Figure 1 presents the model for sentence extraction. In the second step, the list of sentences was sorted and the key phrases that 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 2 for key phrase extraction).

In the third step of the process–action approach, each of the coordination phrases was assigned a weight based on its 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 and Azurix 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 Figures 3 and 4 for details related to assigning coordination weights and keyword weight frequency). 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.

Text mining and coordination score (data collection)

In the final step of the process–action approach, the key phrases from the list above were extracted from the primary datasets—the Dabhol and Azurix 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

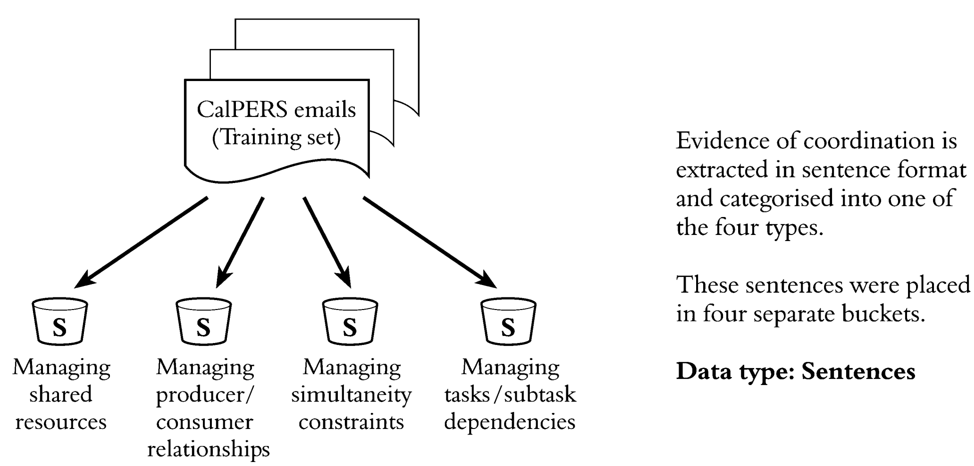


Figure 1 Sentence extraction

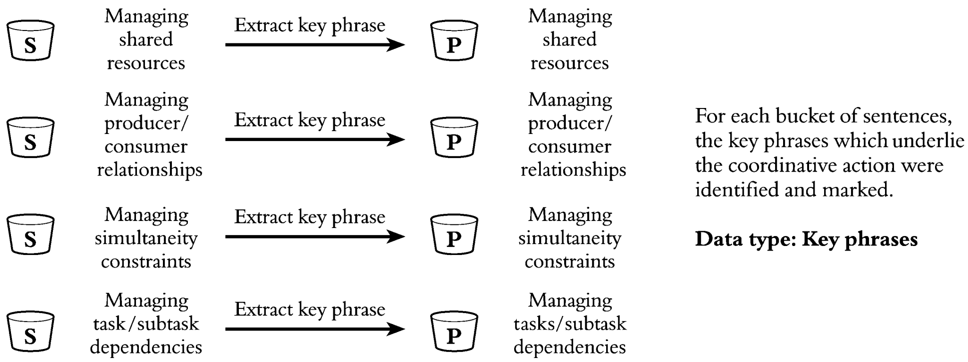


Figure 2 Key phrase extraction

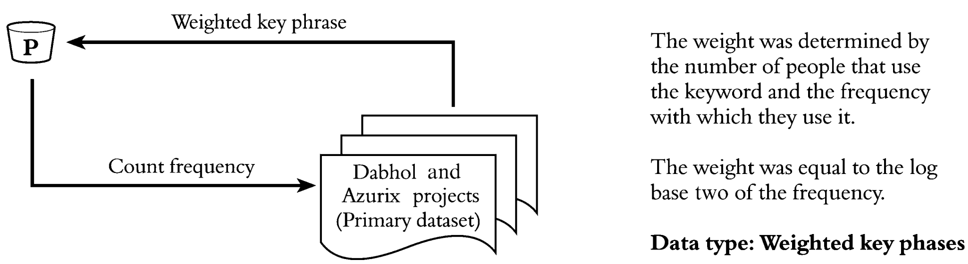


Figure 3 Assigning coordination weights

Weight	Frequency
1.0	1
1.5	2
2.0	9
2.5	8
3.0	4
3.5	4
4.0	3
4.5	7
5.0	6
5.5	3
6.0	5
6.5	3
7.0	2
7.5	4
8.0	2
8.5	0
9.0	0
9.5	1

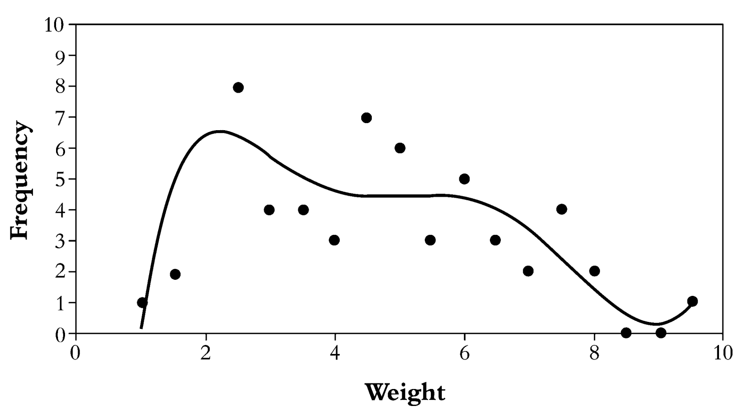


Figure 4 Keyword weight frequency

treated separately and so the operation was repeated three times.

In testing the accuracy of using Support Vector Machines (SVM) to classify the Enron e-mails into folders, Klimt and Yang (2004) determined that all the fields in an e-mail are vital, with none substantially more accurate than another. Rather, the combination of the fields provides greater precision. As such, the text mining application will query all fields of the e-mail. 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 who have used the key phrase in an e-mail. For each match, the program recorded the employee's details and 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 5 shows the pseudocode for the text mining application.

Table 2 Weighted coordination key phrases compiled in the model building phase

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 arrangements	1.00	Follow up	5.16
Please make sure	2.80	On time	4.85
Please update	2.58	Make a schedule	9.47
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
You will be	4.58	We need to	7.82
You work with	1.46	We should	7.02
Ensure that	7.36	To ensure that	5.74
We can go	4.32	It will need	3.45
		I am changing	1.70
		I believe	6.47
Producer/consumer relationships			
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
FYI	5.35	We have begun	2.32

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. Figure 6 shows the text mining application running on the Dabhol project scope. The first few lines load the driver and connect to the MySQL database. The coordination keywords are then read in and searched iteratively. The keyword shown in Figure 6 is 'help coordinate' with a weight of 3.28. Each line represents a match to the keyword. The unique message ID is shown in the square brackets followed by the sender of the e-mail. The program also counted the number of e-mails sent by each employee in order to gauge the level of activity from each user. The application only counts the e-mails that are within the project scope. This becomes important in the data-cleansing phase described in phase three. Figure 7 shows the application running. The third step is to print out the final coordination score for each employee within the project scope. This coordination score was used to conduct the hypothesis testing for the correlation between coordination and centrality (see Figure 8).

Social network matrices (data manipulation)

Network centrality was measured by the number of e-mails sent and received by each person as part of the project scope. This study proposes a new technique for constructing social network matrices from an e-mail dataset. The list of senders was matched to the recipients based on the recipient type TO, CC (BCC was ignored). The rationale for ignoring the e-mails received as BCC is further explained by Klimt and Yang (2004), in which BCC e-mails are usually intended as passive information propagation, rather than establishing a two-way relationship. The strength of the relationship depended on the frequency of e-mails exchanged.

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 calculated using UCINET 6 for Windows (Borgatti *et al.*, 2002). This study measures centrality on two axes: (1) measurement; and (2) directional analysis. Three measures of centrality are used in this study: (1) betweenness; (2) closeness; and (3) 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 while computing degree centrality. Asymmetric data means the sending and receiving of e-mails are treated as distinct activities. For the study of directional analysis, the data were

(1) Read in the key phrases from a text file and iterate through

(2) Run SQL query to search for the key phrase within project scope
(3) Iterate through SQL results and record sender of coordinative emails

(3) Use person object to store email address and coordination score

(3a) IF person does not exist, create person and store coordination score
(3b) IF person exists, update coordination score

Figure 5 Pseudocode for the text mining application to calculate coordination score

treated as asymmetric, following Ibarra and Andrews (1993) and Carley *et al.* (2003). Figure 9 shows a sample of the undirected centrality index scores from the Dabhol project. The full data contained 712 employees.

A directional analysis is further performed to investigate the differences between in- and out-centrality and the effects on coordination ability. Directed degree is measured by counting the number of e-mails 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). Figure 10 shows a sample of the directed centrality index from the Dabhol project. 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 file with all the figures required for the study.

```
C:\j2sdk1.4.2_09\bin>java TextMiner
=> loading driver:
Driver loaded
=> connecting:
Connection Opened
---------------------------
Keyword: Help coordinate
Weight: 3.28

[33205]   richard.sanders@enron.com
[77555]   jeffrey.shankman@enron.com
[83229]   billy.dorsey@enron.com
[95610]   john.arnold@enron.com
[118048]  ricardo.charvel@enron.com
[118370]  j..kean@enron.com
[119998]  steven.kean@enron.com
[126273]  steven.kean@enron.com
[126521]  steven.kean@enron.com
[203380]  cathy.phillips@enron.com
[210549]  jeff.dasovich@enron.com
[233314]  james.derrick@enron.com
[241803]  drew.fossum@enron.com
[305707]  vince.kaminski@enron.com
[306439]  vince.kaminski@enron.com
[324693]  vince.kaminski@enron.com
[335820]  dkenne@houston.rr.com
[368645]  sally.beck@enron.com
[376478]  brenda.herod@enron.com
[376517]  brenda.herod@enron.com
---------------------------
Keyword: Please allow
Weight: 2.35
```

Figure 6 The text mining application running on the Dabhol project

```
--- Counting number of emails sent ---
tracy.geaccone@enron.com[33]
j..parrish@enron.com[19]
jeffrey.shankman@enron.com[16]
john.arnold@enron.com[47]
kay.mann@enron.com[6]
lance.schuler-legal @enron.com[3]
gerald.nemec@enron.com[7]
tim.belden@enron.com[4]
sean.riordan@enron.com[2]
tom.fischbach@enron.com[15]
jr..legal@enron.com[9]
milagros.velasquez@enron.com[2]
vince.kaminski@enron.com[18]
rosalee.fleming@enron.com[19]
steven.kean@enron.com[46]
stanley.horton@enron.com[14]
```

Figure 7 Counting the number of e-mails sent by each employee

--- Final Coordination Score ---	
tracy.geaccone@enron.com	244.38
j..parrish@enron.com	13.79
jeffrey.shankman@enron.com	133.48
john.arnold@enron.com	24.37
kay.mann@enron.com	21.55
gerald.nemec@enron.com	16.43
tim.belden@enron.com	108.96
sean.riordan@enron.com	26.66
tom.fischbach@enron.com	13.33
jr..legal@enron.com	171.63
milagros.velasquez@enron.com	26.66
vince.kaminski@enron.com	117.83
rosalee.fleming@enron.com	121.02
steven.kean@enron.com	344.59
stanley.horton@enron.com	49.78
rod.hayslett@enron.com	111.57

Figure 8 Coordination score as outputted by the text mining application

Results and discussions

This phase ties it together by correlating the centrality measurements with the coordination score. The two test designs look at this relationship at macro and micro levels. The three hypotheses tested were that: (Hypothesis 1) centrality is correlated to increases in coordination ability; (Hypothesis 2) betweenness is the best structural characteristic for predicting coordination; and (Hypothesis 3) out-centrality is a better predictor than in-centrality.

The hypotheses will be tested with three different tests (1) project-based coordination; (2) cross-project coordination; and (3) directed centrality. The project-based coordination test investigates the first two hypotheses (Hypothesis 1 and Hypothesis 2). The cross-based coordination test focuses on Hypothesis 1. The first test (project-based coordination) studies the relationship between coordination and centrality on a macro level. It involves all employees within the project scope and treats each project separately. The data do not track specific employees, with the coordination and centrality figures treated as nameless identities. The second test (cross-project coordination) studies the relationship on a micro level. The coordinative activity of specific employees is measured within each project and studied to determine how the coordination varies as the centrality varies between projects. The third test investigates (Hypothesis 3) the effect of in- and out-centrality on coordination.

Name	Betweenness	Rank	Closeness	Rank	Degree	Rank
wade.cline@enron.com	1519.539	1	129.500	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 centralization index scores and ranks for individuals

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

Figure 10 Directed centralization index scores and ranks for individuals

The tests on project-based coordination were conducted once for each project scope. The tests were performed by dividing the sample into two groups, dichotomized by those ‘high’ and ‘low’ in centrality. They 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 difference between two groups is found to be statistically significant, it is evidence that centrality positioned individuals show more coordinative activity.

Hypothesis 1 is that centrally positioned individuals show more coordination. The project-based coordination test operationalizes Hypothesis 1 by measuring the statistical difference in coordination scores between the low and high groups of centrality. The statistical strength was measured using the Mann–Whitney U-test. The null hypothesis for Hypothesis 1 is that there is no statistical difference between the two groups. The test was repeated for each measure of centrality. 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 Hypothesis 1 will be accepted. Hypothesis 2 is that betweenness is the most potent predicate for coordination in undirected graphs. This test investigated Hypothesis 2 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 the others. The measure of centrality showing the strongest statistical significance was taken to be the best predictor for coordination. The Mann–Whitney U-test is used to determine whether there is a statistical difference between the two groups of data. The Mann–Whitney 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. The Mann–Whitney

U-test was used in this study to compare the coordination scores between those high and low in centrality. The null-hypothesis is that there is no significant difference between the two groups. All hypotheses are one-tailed, and thus the statistical significance level was set at 0.05 (5%). If the U-test shows a significant difference between the two groups, the null hypothesis is rejected.

This test studies coordination using two project scopes, that is, the Dabhol and Azurix projects. The e-mails 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. E-mails containing at least one of these keywords and all reply correspondence 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 e-mail 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 NetDraw for a visual representation.

The coordination for each employee was calculated using the text mining application as defined in phases 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. As part of the data cleansing, 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 in Figure 11. A quick investigation of the score distribution shows a right-skew with a long right tail.

This process was repeated for the Azurix project. The Azurix Water Company dataset contains 1444 people within the project scope (see Figure 12). Of this set, 159 people were found to have demonstrated coordination. After data cleansing there were 97 people. The distribution of the coordination score is shown in Figure 12. Like the Dabhol dataset in the first test, the Azurix distribution shows a right-skew with a long right tail.

Using these coordination scores, the process was to determine if there is a substantial difference in coordination between people with high and low centrality. This is 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 betweenness measure of the Dabhol project. This is shown in Figure 13.

The three main measurements of centrality (i.e. betweenness, degree and closeness) show a one-sided p of less than 0.05, the alpha level used for this study. Moreover, the results were replicated across two projects (Dabhol, Azurix). 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 (Hypothesis 1).

Hypothesis 2 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. In all three projects, the U-test showed the lowest p-value for the betweenness measure. The p-value for the other measurements (although statistically significant) is not as strong as the betweenness index. This indicates the strongest statistical significance in the relationship between betweenness and coordination. 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 tested the people that were involved in all three project scopes. Sixty-two people are found to have been part of all three projects.

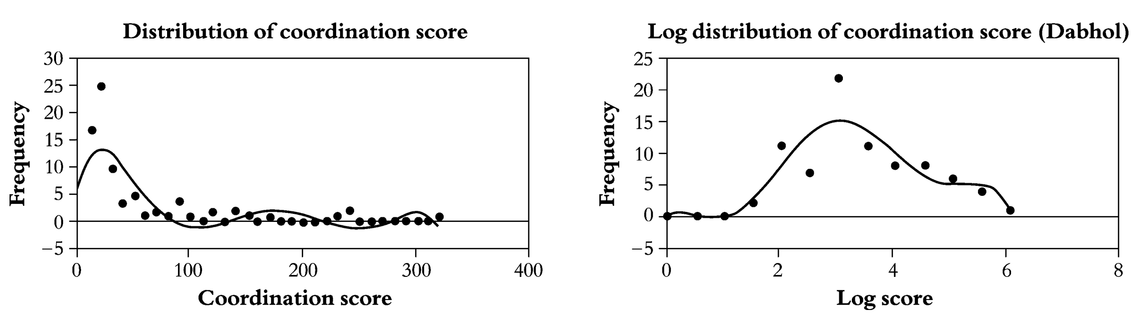


Figure 11 Distribution and log distribution of coordination score (Dabhol)

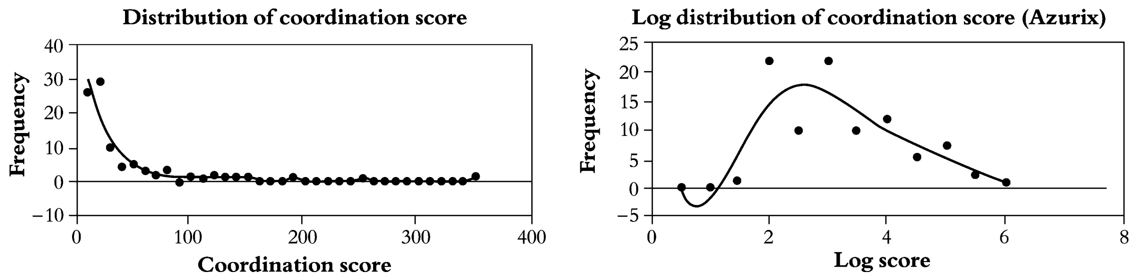


Figure 12 Distribution and log distribution of coordination score (Azurix)

Of these, only eight people demonstrated coordination in all three projects. The others are discarded because this study requires both a coordination score and a normalized degree for the rankings to work. The next step is to rank the projects by coordination and degree centrality for each person (see Figure 14). If the projects' rank is in the same order for both centrality and coordination, it is evidence that they are correlated and the person is marked as a match. The number of instances in which they 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 Hypothesis 1.

Of the eight people in Figure 14, it is found that six of them matched precisely and two did not match. Mary Cook and Vince Kaminski are the two that did not match. In the case of Mary Cook, a close inspection reveals that the Dabhol and Azurix coordination scores are equal and the centrality score is only slightly higher. The two-mode tests of cross-project coordination covered the Azurix and Dabhol projects. The first two-mode test involved the Dabhol and Azurix projects. This was performed by ranking the Dabhol and Azurix coordination score, and also ranking the Dabhol degree with the Azurix degree. If the projects rank is in the same order for both centrality and coordination, it is evidence that they are correlated. In this first two-mode test of the Dabhol and Azurix projects, it is revealed that 48 of a possible 64 people had the coordination and centrality ranked in the same order. Seventy-five per cent of the people showed higher coordination in the project in which they were more central. 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 (Hypothesis 1). This supports the results found in the test of project-based coordination (Test 1). The hypothesis to test (Hypothesis 3) is that out-centrality measurements have a stronger correlation to coordination. Owing 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. 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 (see Table 4), the p-value is found to be significant, and thus all the measures are accepted as evidence.

Conclusions

Using the process–action approach for measuring coordination, we investigated the relationship between social structures on coordination. It was found that centrality had a profound effect on coordination. This provides organizations with a new approach to establishing coordination mechanisms. 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 organizations 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.

The hypotheses tested here reveal some common patterns with recurring characteristics. Dhabol 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 conclusion, the project-based tests find that Hypothesis 1 holds true. Individuals centrally positioned in a network show more coordination. The betweenness

Measuring Statistical Significance

1) The list was sorted by the centrality measurement, in this case, the betweenness index. The median was found.

Coord. score	Betweenness
82.74	1519.599
115.20	1475.576
22.62	1458.313
144.61	1243.107
14.68	1058.002
239.57	964.161
16.05	697.806
232.33	649.278
5.12	612.328
237.58	579.085
23.67	576.115
133.87	503.783
136.38	502.346
99.57	491.275
65.40	480.809
88.45	461.789
71.99	320.353
25.44	313.31
65.60	297.264
3.32	291.721
9.70	243.359
116.97	234.962
32.21	202.597
18.27	195.474
224.88	194.947
46.86	190.864
5.35	180.929
10.32	170.972
9.70	156.948
4.58	134.042
7.22	117
165.77	116.003
14.70	92.542
86.66	88.231
23.99	86.844
17.59	86.133
13.14	79.40
14.63	75.037
14.85	70.047
13.86	67.744
22.88	61.357
42.25	60
13.14	59.781
47.45	55.258
14.68	53.993
14.68	40.907
12.94	20
7.79	20
14.75	13.266
38.60	12.865
57.88	11.5
85.69	8
4.45	6.938
5.35	6.10
12.94	6
9.90	2.5
21.75	2.25
14.67	2.10
33.94	2
21.75	1.75
43.5	1.167
35.46	1
16.46	1
11.82	1
6.47	1
21.75	0.667
43.72	0
20.25	0
20.25	0
19.43	0
18.24	0
16.61	0
12.94	0
12.28	0
10.70	0
7.02	0
6.47	0
6.47	0
5.35	0
5.12	

Median

Next

2) The data was divided into two groups, dichotomised by those 'high' and 'low' in centrality, split by the median.

Hi Between	Low Between
82.74	22.88
115.20	42.25
22.62	13.14
144.61	47.45
14.68	14.68
317.40	14.68
16.05	12.94
232.33	7.79
5.12	14.75
237.58	38.60
23.67	57.88
133.87	85.69
136.38	4.45
99.57	5.35
65.40	12.94
88.45	9.90
71.99	21.75
25.44	14.67
65.60	33.94
3.32	21.75
9.70	43.50
116.97	35.46
32.21	16.46
18.27	11.82
224.88	6.47
46.86	21.75
5.35	43.72
10.32	20.25
9.70	20.25
4.58	19.43
7.22	18.24
165.77	16.61
14.70	12.94
86.66	12.28
23.99	10.70
17.59	7.02
13.14	6.47
14.63	6.47
14.85	5.35
13.86	5.12

Next

3) 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 statistically different from each other.

Mann-Whitney U test

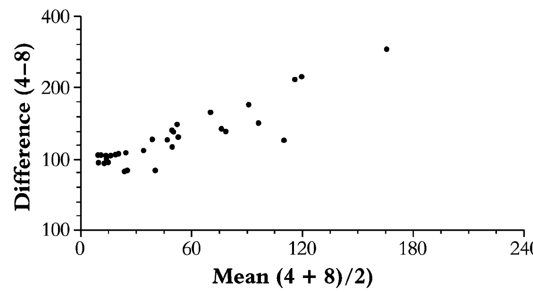
Observations (x) in A = 40 median = 24.715 rank sum = 1907
Observations (y) in B = 40 median = 14.715
U = 1087 U′ = 513

Normalised statistics = 2.762047 (adjusted for ties)
Lower side P = 0.9971 (H_1: × tends to be less than y)
Upper side P = 0.0029 (H_1: × tends to be greater than y)
Two sided P = 0.0057 (H_1: × tends to be distributed differently to y)

95% confidence interval for difference between medians or means:
K = 597 median difference = 11.12
CI = 2.68 to 50.73

Two sample analysis of agreement

95% limits of agreement = –105.198777 to 200.972777



If the difference between two groups are shown to be statistically significant, it is evidence that centrality is correlated to coordination.

The Mann-Whitney U-test shows an Upper side P pf. 0.0029 – This is statistically significant

Figure 13 Process for calculating statistical significance

courtney.votaw@enron.com		
	Coordination	Centrality
Dabhol	2	2
Azurix	1	1
JEDI	3	3

MATCH

james.derrick@enron.com		
	Coordination	Centrality
Dabhol	1	1
Azurix	2	2
JEDI	3	3

MATCH

m..schmidt@enron.com		
	Coordination	Centrality
Dabhol	1	1
Azurix	2	2
JEDI	3	3

MATCH

mary.cook@enron.com		
	Coordination	Centrality
Dabhol	2	3
Azurix	2	2
JEDI	1	1

NO MATCH

richard.sanders@enron.com		
	Coordination	Centrality
Dabhol	3	3
Azurix	1	1
JEDI	2	2

MATCH

rick.buy@enron.com		
	Coordination	Centrality
Dabhol	1	1
Azurix	2	2
JEDI	3	3

MATCH

steven.kean@enron.com		
	Coordination	Centrality
Dabhol	2	2
Azurix	1	1
JEDI	3	3

MATCH

vince.kaminski@enron.com		
	Coordination	Centrality
Dabhol	2	1
Azurix	1	2
JEDI	3	3

NO MATCH

Figure 14 Ranking coordination and centrality between three projects

measure was consistently found to have the strongest statistical significance of a relationship to coordination. The strength of the other centrality measurements varied. Tests showed a stronger relationship for degree and one test showed a stronger relationship for closeness.

Table 3 Results for undirected graphs using the Mann–Whitney U-test

	Dabhol	Azurix
Undirected graphs: Mann–Whitney U-test (5% significance)		
Betweenness	P=0.0029	P=0.011
Degree	P=0.0043	P=0.029
Closeness	P=0.0073	P=0.0366

Table 4 Results for directed graphs using the Spearman rank correlation

	Dabhol	Azurix
Out-degree	r=0.364	r=0.557
	p=0.000882	p=0.000069355
Directed graphs: Spearman rank correlation		
In-degree	r=0.268	r=0.306
	p=0.015915	p=0.00284
Out-closeness	r=0.345	r=0.370
	p=0.001748	p=0.00026071
In-closeness	r=0.262	r=0.271
	p=0.000959	p=0.0024487

In conclusion, it was found in all three tests that the Hypothesis 2 holds true, betweenness is the best predictor for coordination ability in undirected graphs. In testing Hypothesis 3, the experiments of directed centrality consistently found that out-centrality was a greater predicate for coordination than in-centrality. The Dabhol project had clear data indicating this.

References

Bavelas, A. (1950) Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 22.

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

Carley, K.M., Ahuja, M.K. and Galletta, D.F. (2003) Individual centrality and performance in virtual R&D groups: an empirical study. *Management Science*, **49**(1), 21–38.

Chatterjee, P. (1995) Enron deal blows a fuse. *Multinational Monitor*, **16**(7–8).

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

Dörre, J. and Gerstl, P. (1999) Text mining: finding nuggets in mountains of textual data, in *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 398–401.

Dubois, A. and Gadde, L.-E. (2002) The construction industry as a loosely coupled system: implications for productivity and innovation. *Construction Management and Economics*, **20**(7), 621–31.

Faust, K. (1997) Centrality in affiliation networks. *Social Networks*, **19**(2), 157–91.

Freeman, L.C. (1979) Centrality in social networks: conceptual clarification. *Social Networks*, **1**(3), 215–39.

Freeman, L.C. and Roeder, D. (1980) Centrality in social networks: ii. Experimental results. *Social Networks*, **2**(2), 119–41.

Giddens, A. (1984) *The Constitution of Society: Outline of the Theory of Structuration*, University of California Press, Berkeley.

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

Ibarra, H. (1993) Network centrality, power, and innovation involvement: determinants of technical and administrative roles. *Academy of Management Journal*, **36**(3), 471–501.

Klimt, B. and Yang, Y. (2004) The Enron corpus: a new dataset for email classification research, in *Machine Learning: ECML 2004: Proceedings 15th European Conference on Machine Learning*, Pisa, Italy, 20–24 September, pp. 217–28.

Knoke, D. and Burt, R.S. (1983) Prominence, in Burt, R.S. and Minor, M.J. (eds) *Applied Network Analysis: A Methodological Introduction*, Sage, Beverly Hills, CA, pp. 95–122.

Leavitt, H.J. (1951) Some effects of communication patterns on group performance. *Journal of Abnormal and Social Psychology*, **46**, 38–50.

Malone, T.W. and Crowston, K. (1993) Tools for inventing organizations: towards a handbook of organizational processes, in *Proceedings of the 2nd IEEE Workshop on Enabling Technologies Infrastructure for Collaborative Enterprises*, Morgantown, WV.

Malone, T.W. and Crowston, K. (1994) The interdisciplinary study of coordination. *ACM Computing Surveys*, **26**(1), 87–119.

McLean, B. and Elkind, P. (2003) *Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*, Penguin Books Ltd., London.

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

Moore, S. and Eng, E. (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), 305–18.

Mullen, B. and Johnson, C. (1991) Effects of communication network structure: components of positional centrality. *Social Networks*, **13**(2), 169–86.

Pentland, B.T. (1994) Grammatical models of organizational processes, CCS Working Paper No. 176, Sloan WP 3720-94.

Stephenson Jr, M. (2004) Making humanitarian relief networks more effective: exploring the relationships among coordination, trust and sense making, Paper prepared for presentation at the Annual Conference of the Association for Research on Non-Profit and Voluntary Associations, Los Angeles, California, 18–20 November.

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

Zemljic, B. and Hlebec, V. (2005) Reliability of measures of centrality and prominence. *Social Networks*, **27**(1), 73–88.