Value of Social Network -- A Large-Scale Analysis on Network Structure Impact to Financial Revenue of Information Technology Consultants[1]

Lynn Wu
MIT Sloan School of Management
IBM Research

Ching-Yung Lin
IBM Research

Sinan Aral
NYU Stern
MIT Sloan School of Management

Erik Brynjolfsson
MIT Sloan School of Management

Working Paper

Abstract

A large body of literature on social networks in organizations demonstrates that certain types of network topology are optimal. However, little research leverages the ample data created by people's electronic communications to refine and verify theories. This gap is problematic, because the literature on organizational networks suffers from the same deficits as much of the social network literature: both tend to be focused on small, static networks. In this study, we mitigate this gap by collecting and mining the largest organizational social network ever collected. We find that not only does the population level topology of social network correlate with performance, attributes of the nodes in a social network such as human capital and status that can be beneficial to work performance. In addition to an individual's own human capital and network position, the human capital and status in one's network can be instrumental to one's success.

[1] **This work was presented at Winter Information Systems Conference, Salt Lake City, UT, Feb. 2009. Presentation slides can be accessed at http://smallblue.research.ibm.com/**

Introduction

This study presents new empirical evidence on the relationship between information worker productivity and social capital generated from social networks. As the information content of work increases, studying how information workers generate value through both technological and social means is important. There is a large body of literature on social networks and organizations that describes the benefit of social networks on work performance in various settings. For example, a stream of literature shows that individuals and organizations with specific types of network topology, such as one that spans multiple structural holes and occupies a central position within a network is more likely to be successful (Burt 1992). However, little research leverages the ample data that are created by people's interactions, such as e-mail, call logs, text messaging, document repositories, web 2.0 tools, and so on. This gap is problematic, because the literature on organizational networks suffers from the same deficits as much of the social network literature does: both tend to be focused on small, static networks. As a result, important questions like 'does the optimal network structure for performance hold for a large network' and 'what is the appropriate timing of communication to actors of interest', and the like have been completely neglected.

Recent advances in information technology give researchers the opportunity to solicit real-time email communication data. Since email archives record detailed communication logs, such as who has emailed whom, the exact time of the interaction, and the content of the exchange, constructing social networks using email archives allows researchers to eliminate errors and bias that are often introduced in self-reports. Recent empirical work has started to capture real-time communication between people in various settings. Aral et. al (2006, 2007) and Aral and Van Alstyne (2007) record real-time electronic communication within an executive recruiting firm to provide some of the first empirical evidence demonstrating how information workers use information technology and their social networks to generate business value. Similarly, Wu et. al (2008) use "sociometric badges" to examine the impact of face-to-face communication networks on the productivity of hardware configuration specialists. While these works leverage new data collection techniques to advance our understanding of how information workers utilize social networks to improve their work performance, we have yet to see a study that explores a large-scale dataset involving more than a handful of people over a brief period of time. Lin et. al (2008) invented a privacy-preserving organizational social network analysis system that uses social sensors to gather, crawl and mine various types of data sources, including content of individual email and instant message communications, calendars, hierarchical structure of the organization as well as individual role assignments. The system is deployed in more than 70

countries and collected electronic communication content over 3 years. After anonymize the identity and the content of these communications, we are able to quantitatively infer the social networks of 400,000 employees within a large organization. To precisely estimate the value of social networks, we collect performance metrics of a subset of these employees from the corporate project and personal revenue database. To our knowledge, this is the largest social network ever constructed to study the impact of social networks on information worker productivity.

In particular, we focus on a class of information workers that have rarely been examined in the past: consultants that represent a large population of information workers who generate revenue by logging "billable hours." To explore the relationship between social network and productivity of these consultants, we collect detailed and objective performance measures of more than 1000 consultants including the number of billable hours, participated projects and the revenue generated. To understand how consultants generate economic value, we also conducted extensive interviews to 15 consultants at various stages of their career. Through these interviews, we find that efficient access to useful information is crucial, as timely and valuable information can facilitate fast and high-quality decision making. Satisfying customers is the cornerstone for consulting services, as generating repeat businesses from existing clients is one of the key performance indicators for the health of a consulting business. If expedient access to information improves productivity, understanding the mechanisms of how information workers access the information through both social and technological means is important. Continuing with the micro-analysis research on worker productivity pioneered by Ichniowski, Shaw and Prennushi (1997) and Aral Brynjolfsson and Van Alstyne (2006), we study a single industry in depth and examine how information workers obtain information through various communication channels and social networks. With the cooperation of the company and employees, we monitored email and instant messaging usage to analyze the flow of information and its relationship to productivity and performance.

The sheer volume of the data allows us to more precisely estimate how population-level topology in a network contributes to information worker productivity, after controlling for a host of factors. By capturing communication patterns and organizational roles of a large and diverse body of employees we identify a new mechanism of how information workers generate values for their projects. We find that not only does network topology have a strong relationship with work performance, human capital and status of one's social contacts instrumental as well. This demonstrates that in addition to one's own human capital and network position, it is

important to incorporate a third type of capital, the cumulative human capital inside one's social contacts. We demonstrate that this type of capital has a unique contribution to work performance in addition to the traditional human capital and network characteristics.

Specifically, we uncover four key results. First, we find that the structural diversity and centrality of social networks are positively correlated with performance for both individual consultant and project teams, corroborating previous work (Burt 2004, Aral el al. 2006, 2007, Aral and Van Alstyne 2007). Second, while having many weak ties to management is negatively correlated with work performance, strong ties to powerful individuals, such as access to executives is positively correlated with work performance. This demonstrates that it is important to examine the content of the network in addition to network topologies. The strength of connections and attributes of the nodes in the network, such as human capital, power and status, can have a unique relationship with work performance. Third, a team with strong ties to the management can be beneficial for work performance, having many managers working on the same project exhibits an inverted U-shape relationship with performance. Placing managers in a project is beneficial, but only up to a point. When a project has too many managers, the productivity actually decreases. Lastly, we study if a person' own network or the network of the overall team has a stronger association with a person's own work performance. We find that consultants derive the lion share of benefits from social networks of their project. This result highlights that composing a project team with desirable network characteristics is instrumental for both the success of the project and the work performance of individual team members. Participating in projects with the appropriate social capital can boost consultants' work performance in addition to their own social capital.

Theory and Literature

Information Worker Productivity

A large body of literature examines factors that can improve productivity of individuals and firms, but most of the works in the past largely focus on manufacturing sectors where output can be easily measured by counting the number of physical goods produced. However, as information work is growing to be a significant part of our economy, examining how information workers produce values in our economy becomes increasingly important.

Recently, a few papers try to advance our understanding of the production function of information workers. By analyzing the technology usage and email communication of executive recruiters, Aral et. al(2007, 2008) identify several key inputs that are theorized to impact the productive outputs of information workers. Similarly, Wu et. al (2008) studies verbal communication and examine the effect of face-to-face networks on worker productivity by recording all conversations of a group of hardware configuration. While these studies improve our understanding of how information workers generate values, more studies should be conducted to broaden our perspectives on information work. Since a large population of information workers, such as consultants, accountants and lawyers, generate revenue by logging "billable hours," studying consultants can advance our understanding of how information workers generate value. In this study, we delve into the production function of information technology (IT) consultants and understand the factors that contribute to their productivity. Specifically, we focus on intangible assets such as social capital and access to human capital that are generated from social networks and how these different types of capital are associated with worker productivity. In addition to individual's own human capital and network topology, we study how connections to power and status, and the strength of these connections are correlated with performance.

Network Topology and Information Advantage

The ability to access diverse and valuable information effectively promotes information worker productivity through two means. First, accessing information related to work at hand can directly improve the quality of work, as it can increase the chance of finding solutions to difficult problems. Second, accessing diverse information makes it more likely that new opportunities and resources will be discovered more quickly. To be first to learn about new opportunities allows a person to enter the queue faster and prompting her to suitably reposition her strategy, and therefore improving her chance of winning the opportunity. In IT consulting business, accessing information expediently is the key to improve productivity. Since consultants generate revenue through billable hours, it is crucial to avoid bench time as much as possible and increase utilizations by lining up projects ahead of time. At the same time searching for high-value projects that command higher wage is also important as these projects can generate higher revenue for each hour worked. Accessing a wider array of information about new project opportunities gives consultants the first mover advantage. Being the first to apply for high-value projects increases the likelihood of a consultant to be selected. Knowing where to obtain expertise to solve difficult problems

and produce high quality work also improves the likelihood of participating in high valued ventures. As word-of-mouth is extremely effective in the consulting business, reputation of a consultant can quickly disseminate throughout the organization. Reputable consultants are often in high demand as managers generally prefer them to handle high-value ventures, since they are more likely to satisfy customers and retain them for future services.

Social networks play an instrumental role in helping consultants to access diverse and valuable information through their contacts. Structural diversity and network centrality are two important topological characteristics of social networks that are theorized to improve work performance. Structural diversity with abundance of structural holes in a network is shown to have positive impact on innovation (Burt 2004), early promotion (Burt 1992), and R&D productivity (Reagan & Zuckerman 2001). Actors in a structural diverse network derive information benefits, since they are more likely to receive non-redundant information coming from various loosely connected networks neighborhood (Burt 1992). Economic value of information stems from the fact that information are distributed unevenly in a social network, and thus the ability of an consultant to tap into diverse and unique information sources through a structural diverse social network is more likely to solve hard problems as well as participating in new consulting opportunities. On the other hand, structural redundant networks tend to provide similar information to the consultant and a dense network of strong ties is likely to spread information quickly to many people and thus reducing the ability of anyone to act on the information obtained. A dense and redundant network also have high maintenance costs, since direct ties and a dense network of third-party ties require time and effort to culminate and maintain. From our qualitative interviews, the ability to tap into diverse pools of information is not only crucial to solve difficult problems but also important for finding new consulting opportunities in the future. Thus, we expect a structural diverse network with abundance of structural holes to be highly correlated with information worker productivity.

While structural holes provide diverse information through tapping into different network neighborhoods, proximity to the information source can also be conducive to productivity. An actor's direct and indirect contacts can affect her ability to search for relevant information (Hansen 2002). While direct contacts enable actors to immediately initiate information transfers, indirect contacts quicken the search process because actors can learn about information and opportunities in the network through word of mouth. Individuals can therefore leverage their web of third-party ties to obtain the desired information. However, indirect contacts can also distort information content. When information gets passed through long path lengths (Freeman 1979), the chance of distortion is

particularly high, as people tend to misunderstand or misinterpret information (Collins and Guetzkow 1964, Huber an Daft 1987, Gilovich 1991, Hansen 2002). Imprecise or inaccurate information can have a negative performance impact on the focal actor. Acting on vague information obtained indirectly, the focal actor may need to use her ties to connect to the original source of the information, only to find it was not what she sought. Eliminating misleading information is costly, as verifying each incorrect lead wastes valuable time and effort. When an actor has relatively short path lengths to other experts, not only is she exposed to less information distortion, she can also access knowledge experts more quickly. One way to gauge the ability to reach a broad range of actors in a network is betweenness centrality which measures how often an actor is positioned on the shortest path between other pairs of actors in the network (Freeman 1979). When an information worker is positioned in the network where she can access other actors quickly, she is more likely to be in the most effective position (Freeman 1979, Brass & Burkhardt 1992, Burt 1992, Hansen 2001) and is more likely to access more novel information more quickly. Thus, we expect consultants and project teams with high betweenness centrality to be strongly correlated with generating higher revenue.

Human Capital of Social Networks

While ample research has shown the importance of occupying strategic positions in the network is advantageous (Burt 1992, Freeman 1979). Traditional metrics for positional advantage such as network constraint and various centrality measures treat every node in the network to be same. This type of analysis is designed to demonstrate that all else equal, a person can derive premium by occupying a strategic position in the network. However, focusing on network topology alone misses important aspects of the network. Content of the network such as knowledge and expertise accumulated within one's social contacts can have a significant consequence on work performance (Rodan & Galunic 2004). To fully understand the relationship between social networks and productivity, it is also important to examine the attributes of nodes as well as the strength of ties within a network. Differentiating each person's social capital solely by network positions is narrow and simplistic. It is possible that someone endowed with positional advantages in a social network would still perform poorly if the contacts in his network cannot communicate the important information that is relevant to his work. However, the same person could still be productive without these positional advantages if he has access to a few brilliant advisors who counsel him with valuable guidance. It is therefore important to differentiate the human capital characteristics inside one's

social contacts to gain an understanding of how they can influence work performance on top of the topological characteristics of the network.

One important aspect of human capital is status and power endowed to a person. Strong ties to powerful individuals can be beneficial. Authorities such as partners and executives in a consulting organization often have the discretion to pick specific consultants to staff important projects that have high earning potential. Project managers prefer to choose consultants whom they trust to deliver excellent results and therefore, having a good relationship with project managers and gaining their trust is crucial to increase utilization and the likelihood to participate in high valued ventures. Consultants with strong ties to managers who are often experts are also likely to perform better as they can access useful knowledge and subject area expertise. Receiving targeted and useful information directly from the manager with minimum information distortion, consultants with strong ties to management are even more likely to complete a project. This forms a virtuous cycle where strong connections to managers increase the chance of accomplishing a project which then enhances a consultant's reputation and attracts even more connections to project managers. Similarly, we expect project teams with strong ties to managers outside of the team to be more successful as well.

While a project with external advisors can be beneficial to generate project revenue, having too many managers in the same project could have the opposite effect. A team with several project managers may have different visions on how the project should be carried out. When too many managers work in the same project without a central authority to make the necessary decision, the project is more likely to derail as multiple and often contradictory directions from different leaders hamper the team members from making any significant progress. From our qualitative interviews to low and mid tier consultants, we find that consultants often feel confused about the direction of the project when there are too many project leads that often have contradictory views. Without a clear vision and directions from the project leaders, consultants often received mixed if not contradicting requests that impede their progress to complete the project on time. Thus, we expect that while project managers should have a positive impact on a project since they tend to be subject area experts, having too many managers working on the same project could have the opposite effect of derailing the project.

Data and Network Construction

We analyze an electronic communication social network of more than 7000 employees over 3 years. The data contains email and instant messaging activities inside a global information technology firm with more than 30 product divisions. We collected detailed electronic communication records of 7043 volunteer employees in more than 70 countries. From the 7043 volunteers, we derive a social network of more than 400,000 people within the firm. In our analysis, we constrain the analysis to focus on the sub-network for the 7043 volunteers for which we have their complete electronic communication data. To eliminate any potential self-selection effects from using volunteer data, we compare the network characteristics and job roles of the volunteers with the rest of the firm. We find minimal differences between the two populations to warrant any concerns of using only the sub-network of the 7043 volunteers.

To understand how social network is related to work performance, we also collected detailed financial performance data of more than 10,000 consultants. Combining the financial data with social network data yields a total of more than 1000 consultants whom we have both network and financial performance data. To protect the privacy of the participants, their identities are replaced with hash identifiers. We also eliminated spam and mass email announcements in order to construct a view of the network that reflects the real communications between actors. We build a single cross-sectional network based on all electronic communications of our participants to capture the characteristics of the overall network. To capture changes in social networks over time, we also assemble a set of longitudinal social networks based on the quarterly communications from 2007 to 2008. Each network in the set is built using only the communications occurred in a specific quarter of the year. The summary statistics for the network is shown in Table 1 and 2.

Network Topology Measures

To gauge the topological characteristics of an individual's social network, we employ network in-degree, out-degree, betweenness centrality and structural diversity to approximate the range of a worker's network.

Direct contacts are measured using two measures: In-degree and out-degree. *In-degree* is measured as the number of email sent to the actor while *out-degree* is measured as the number of emails sent from the actor.

Betweenness centrality $B(n_i)$ measures the probability that an individual *i* will fall on the shortest path between any two other individuals in a network (Freeman 1979), where $g_{jk}(n)$ is the number of shortest geodesic paths from *i* to *j* that pass through a node *n*, while g_{jk} is the number of shortest geodesic paths from *i* to *j*:

$$B(n_i) = \sum_{j<k} g_{jk}(n_i) \Big/ g_{jk};$$

Network constraint C_i measures the degree to which an individual's contacts are connected to each other. P_{ij} is the proportion of *i*'s network time and energy invested in communicating with j. Network constraint can be used as proxy for measuring network cohesion (Burt 1992), and network diversity is simply computed as 1-C.

$$C_i = \sum_j \left(p_{ij} + \sum_q p_{iq} p_{qj} \right)^2, \quad q \neq i, j.$$

Strength of Ties

In addition to counting the number of communication occurrences between two actors, we also invented measurement for the strength of ties as defined as below. This measure of ego strength has been extensively tested and is shown to accurately reflect the strength of tie between actors (Lin et al., 2008). To make comparable companrision, the measure is normalizes to vary between 0 and 1. We define strong ties to be values above 0.8[2].

$$strength_{i,j} = \frac{\log(X'_{ij})}{\max_j \log(X'_j)}$$

$$X'_{i,j} = \begin{cases} 0 : if\{X_{i,j} \leq 3 + \log(X_{i.j})\} \\ X_{i,j} : otherwise \end{cases}$$

where X_{ij} is the total communications between actor *i* and *j*.

Person-level and Project-level Networks

To understand a broader view of how social network is associated with performance, we construct two types of social networks, at person level and project level. In the employee or person level, each node in the network represents an employee and each link between two nodes represents the total number of electronic communications

[2] We also tried other values to measure strong links. Our empirical results do not change qualitatively.

between the two people. Using employee-level networks, we explore the type of network characteristics are strongly correlated with work performance of individuals. However, the same network characteristics that are beneficial at the individual level may not directly translate to better performance at the team level. For example, if everyone in the project team strives to have the same optimal network characteristics, we would not necessarily expect the project to be successful, for the similar reason that having too many managers can be detrimental to project success. Thus, it is important to study network characteristics at the project level and explore the type of project network characteristics and team composition that are most beneficial to project performance. To construct project level networks, each node in the networks is a project instead of an individual and links between projects is the sum of all communications between members of the two projects, ignoring all communications happening within the same project team. If a person belongs to two project teams, we assign the contribution of this person to the overall communication between the two projects to be the maximum inter-team communication between any pair of individuals in these two teams. We hypothesize that the network characteristics that are most conducive for project networks may differ from those at the individual level. For example, the ability of a consultant to generate revenue depends on his ability to bill hours and participate in high valued projects, while project revenue depends not only on the ability of individual team member but also on how they work together to complete the project.

To link network characteristics to work performance, we obtain performance measures for all the projects a consultant has worked on, including both his own contribution to the project as well as the overall project performance. In order to protect the privacy of these consultants, both identities of projects and consultants are hashed with unique identifiers. Similarly, we also hashed the identity of the clients these consultants have served. We collect financial performance such as billable hours and revenue they generated on a monthly basis, as well as project characteristics such the difficulty of the tasks, the region where the project is performed, the industry of the client, line of business for the consultants and clients, and duration it takes for the project to be completed. To control for individual characteristics of consultants, we also collected demographic data, such as the technical expertise, geographical locations, managerial roles, and the division they belong within the organization. Unfortunately, we cannot obtain information on how long a consultant has worked in his current job role. This could skew our results as we may expect that employees who are around for a long time are more likely to have a larger network since they have more time and opportunities to build their social network. However, we can eliminate this bias using longitudinal networks. By looking at the changes between social network composition and changes in

their performance over time, we can eliminate any time invariant properties such a person's relative tenure or innate abilities by using fixed-effect specifications in our model.

Table 1: Summary Statistics for Person-Level Networks					
Variable	Obs	Mean	Std. Dev.	Min	Max
In-degree	7043	13.35	7.81	0	125
Out-degree	7043	13.41	8.72	0	132
Betweenness centrality	7043	.001	.001	0	.018
Network Constraint	7043	.564	.284	.074	1.810
Direct contacts with project managers	7043	.120	.466	0	5
Gender: male = 1	7043	.834	0.372	0	1

Table 2: Summary Statistics for Project-Level Networks					
Variable	Obs	Mean	Std. Dev.	Min	Max
Project indegree	2952	1.664	3.653	0	26
Project outdegree	2952	1.700	4.520	0	28
Project betweenness	2952	4.90e-06	.00003	0	.0008
Project constraint	2952	.826	.294	.064	1.021
Direct contacts with project managers	2952	.544	1.476	0	13
Number of managers in a project	2952	.235	.503	0	5
Number of divisions in a project	2952	2.92	10.105	1	324
Number of male in a project	2952	1.07	.957	0	16

To understand in detail how consultants work, we conducted extensive interviews with 15 consultants in various roles and geographic locations to understand the different factors inside the production function of a consultant. From these interviews, we find that a person's ability to generate revenue is the most appropriate productivity measures. In addition to resource utilization or billable hours, the hourly wage a consultant can command also affects the revenue. Although the hourly wage for different types of tasks can be fixed for a given project, reputable consultants are often more likely to be selected to execute projects that have high visibility or revenue potential.

Combining the financial data with our network data, we are left with 1029 consultants and 2952 projects these consultants have done in the last 13 months from June, 2007 to July, 2008. The scope of these projects involves 39 countries around the world with the majority of the projects being done in the US. Of the 1029 consultants, 864 are male, 66 are project managers. Low to mid level consultants represent 69.4% of the population while the rest are senior consultants or executives. While a large proportion of the projects consist of a few people, the number of consultants working on a project can be up to 20 people. Some project may have no managerial involvement while some project can have up to five managers.

Empirical Methods

We employ a linear regression model to examine the effect of network characteristics on revenue for both employee and project networks. In the person-level network, the dependent variable—revenue generated—is measured as the total US dollars a consultant generates in a month from all the projects he has participated. In the project-level network, project revenue are the monthly income generated by all members in the project team.

$$\text{Revenue}_{it} = \alpha + \beta_1 \text{hours}_{it} + \boldsymbol{\beta_2}\ \text{in-}\mathbf{degree}_i + \boldsymbol{\beta_3}\ \mathbf{out\text{-}degree}_i + \boldsymbol{\beta_4}\ \mathbf{betweenness}_i + \boldsymbol{\beta_5}\ \mathbf{constraint}_i + \boldsymbol{\beta_4}\ \mathbf{managers\_in\_network}_i + + \boldsymbol{\beta_4}\ \mathbf{StrongLinks} + \boldsymbol{\beta_4}\ \mathbf{StrongLinksToMgr} + \Sigma\ \text{personal\_characterisitcs} + \Sigma\ \text{job\_characteristics} + \Sigma \text{regional\_characteristics} + \varepsilon$$

$$\text{Revenue}_{jt} = \alpha + \beta_1 \text{hours}_{it} + \boldsymbol{\beta_2}\ \mathbf{in\text{-}degree}_j + \boldsymbol{\beta_3}\ \mathbf{out\text{-}degree}_j + \boldsymbol{\beta_4}\ \mathbf{betweenness}_j + \boldsymbol{\beta_5}\ \mathbf{constraint}_j + \boldsymbol{\beta_4}\ \mathbf{managers\_in\_network}_j + \boldsymbol{\beta_5}\ \mathbf{managers\_in\_project}_j + \boldsymbol{\beta_5}\ \mathbf{managers\_in\_preject}^2_j + \boldsymbol{\beta_5}\ \mathbf{genders} + \boldsymbol{\beta_5}\ \mathbf{gender}^2_j + \Sigma\ \text{personal\_characterisitcs} + \Sigma\ \text{job\_characteristics} + \Sigma\ \text{regional\_characteristics} + \varepsilon$$

We expect the coefficient of betweenness centrality to be positive while the coefficient for network constraint to be negative as predicted by our theory. We also expect the coefficients for the number of managers in a person's or a project's direct to be positive.

We then try to disaggregate the data further to examine revenue generated by each person for each project. This allows us to understand how personal network and project network characteristics simultaneously affect a person's ability to generate revenue.

$$\text{Revenue}_{ijt} = \alpha + \beta_1 \text{hours}_{it} + \beta_2 \text{ in-degree}_i + \beta_3 \text{ out-degree}_i + \mathbf{\beta_4\ betweenness_i} + \mathbf{\beta_5\ constraint_i} + \mathbf{\beta_4\ managers\_in\_network_i} + \beta_5 \text{ in-degree}_j + \beta_6 \text{ out-degree}_j + \mathbf{\beta_7\ betweenness_j} + \mathbf{\beta_8\ constraint_j} + \mathbf{\beta_9\ managers\_in\_network_j} + \mathbf{\beta_{10}\ managers\_in\_project_j} + \Sigma \text{ personal\_characterisitcs} + \Sigma \text{ job\_characteristics} + \Sigma \text{ regional\_characteristics} + \Sigma \text{ project\_characteristics} + \varepsilon$$

We pose this as an empirical question to see if project network characteristics have a stronger correlation with productivity than a person's own network characteristics.

Results

Employee-level networks

We first test the relationships between network structures and individual employee productivity (Table 3). Since our network data is longitudinal by nature, we calculate the network characteristics on a quarterly basis and use both random effect and fixed effect specifications as shown in column 1 and 2 in Table 3 respectively. The results from random effect (column 1) and fixed effect (column 2) models are largely similar. We will explain results of the fixed effect model in detail since it gives more credibility to our interpretations by eliminating time invariant factors that are not observable. As expected, the number of billable is highly correlated with revenue. Each additional hour billed is associated with $114 dollars in monthly revenue (Column 2). Network cohesion is negatively correlated with productivity, corroborating the results of previous email network studies (Aral et. al 2006, 2007, Aral and Van Alstyne 2007) demonstrating that structurally diverse networks are positively associated with productivity. One standard deviation increase in network constraint is associated with a decrease of 276% in monthly revenue.

Similarly, betwennness centrality is also positively correlated with performance, where one standard deviation in betweenness centrality is associated 513% increase in monthly revenue. Based on interviews with consultants at various stage of their career, we find that accurate and timely access to novel information through a structural diverse network is crucial to project completion by assisting consultants to solve complex problems and discover new opportunities. When an information worker is positioned at the center of the network where she can

access other actors quickly, she is in the most effective position to access novel information quickly (Freeman 1979, Brass & Burkhardt 1992, Burt 1992, Hansen 2001), and act on newly discovered opportunities. This is particularly important in consulting business, as lining up projects to avoid bench time is crucial for performance. Becoming aware of the newly available resources or projects allows the consultant get into the queue early, prompting the consultant to suitably reposition himself in order to take advantage of the newly discovered opportunities. Interestingly, we find that reaching many people inside an organization in 3 steps or less is weakly correlated with decreased performance. Each person a consultant can reach in 3 steps or less is associated with 2 cents decrease in revenue, demonstrating that the cheer number of people one can reach matters less than locating at the center of information highway.

In our model, network size, as measured by in-degrees and out-degrees, does not have a positive association with performance. In both random and fixed effects specifications, in-degrees and out-degrees have either negative or statistically insignificant correlations with performance. However, exploring the content of these conections, we find that the strength of the connection to authorities or experts inside one's network can moderate the one's work performance significantly. As shown in the fixed effect model (column 2 in Table 3), having strong connections to managers is shown to have a positive association with revenue. Knowing a manager well is associated $588 additional monthly revenue, suggesting that having strong connection to powerful individuals provide consultants consulting opportunities and avoids bench time. However, neither strong connection to colleagues in general or weak connections to management is shown to have a positive relationship to performance. In fact, too many weak links to managers actually negatively correlates with performance. Additional weak tie to a manager is associated with $98 decrease in monthly revenue.

From our interviews, we find that strong connections to managers, particularly project partners increase the chance of a consultant to find work. These managers are more likely to introduce consultants to high-value projects when they become available, providing consultants with first-mover advantage to win these opportunities. To minimize staffing risks, project managers also tend to choose consultants they know well to work on important projects. Having in-depth knowledge about the consultant through strong connections also enables managers to match the consultant to appropriate projects. On the other hand, managers would lack such insights about the consultant if the connection is not strongly linked. Simply knowing many managers is shown to negatively correlate

with performance. It could be that when consultants are linked to many managers, he may have conflicting demands from them on how the project should be executed. The consultants would be rendered less effective without a clear direction.

Table 3: Person-level Email Networks			
Dependent Var.	Monthly Personal Revenue	Monthly Personal Revenue	Monthly Personal Revenue
Controls: Project Complexity, Line of Business, Months, Regions, job level, month, job role			
	(1)	(2)	(3)
	Random Effect	Fixed Effect	Cross sectional network
hours	115.2***	113.7***	$119.84***
	(0.950)	(0.993)	(2.307)
Betweenness Centrality	321.7***	513.35 **	-283.5**
	(199.36)	(217.83)	(126.9)
Constraint	-328.49**	-276.64**	-313.5**
	(106.68)	(113.88)	(125.4)
In-degree	-13.51	-82.24*	-30.32
	(34.17)	(39.65)	(46.56)
Out-degree	8.358	9.833	99.31**
	(21.85)	(24.38)	(47.29)
Number of Managers in network	1.441	-98.48*	368.0
	(48.01)	(53.49)	(855.3)
Number of strong links	-52.14	-56.10	-2.427*
	(43.61)	(48.43)	(1.406)
Number of strong links to managers	808.2**	588.2*	1320**
	(365.2)	(389.4)	(615.8)
Total communication to managers	-7.418	-6.777	-10.19
	(4.588)	(4.880)	(7.554)
Reach in 3 steps	-0.00948**	-0.0191***	1.66e7***
	(0.00433)	(0.00489)	(4.02e8)
Divisions in network	-283.2**	-196.6	-254.6***
	(114.7)	(128.7)	(95.78)
Is Mananger	2685***	--	566.6
	(817.5)	--	(1376)
gender_code (1=male)	84.96	--	236.9
	(495.6)	--	(214.5)
Observations	5527	5527	5499
R-squared	0.81	0.81	0.78
*p<.1, **p<.05, ***p<.001. Huber-white robust standard errors are shown in parentheses			

In addition to node attributes such as managers, we also explore other type of network content by examining if knowing colleagues in different divisions has any effect on performance. It is possible that information diversity come from the ability of a consultant to reach into different divisions within the organization instead of structural diversity at the topological level. Our results do not support this hypothesis. Direct contacts with people

in different divisions do not have a statistically significant relationship with the performance. Reagan Zuckerman and McEvily (2004) also demonstrate that demographic characteristics of team members have ambiguous performance implications as oppose to social network characteristics.

Project-Level Network

We explore the relationship between topological characteristics of project networks and outcome. In project-level networks, each node is a project instead of a person and a link between two nodes measures all the communications exchanged between two projects. As predicted, beweenness centrality at the project level is also positively correlated with performance, similar to what is found in the employee-level networks. One standard deviation of betweenness centrality is associated $115.2 in monthly revenue, suggesting that when project teams can obtain diverse and novel information through efficiently reaching various actors inside the network, they are more likely to solve difficult problems and complete the project. Contrasting to employee-level network, network constraint at the project level ceases to have any association with performance. The coefficient is positive but not statistically different from zero, suggesting that the topological advantage at the project level primarily comes from occupying at the center of the information highway (high betweenness centrality) within a social network.

As shown in Table 3 and 4, betweenness centrality has strong correlations with performance at both personal and project level. However, it is still ambiguous to where the project centrality comes from. Centrality could come from the composition of the team where everyone has high centrality individually. It is also possible that a few consultants whose extraordinarily high personal centrality drives the overall centrality measure of the team. This scenario implies that high centrality individuals are perhaps the project gatekeepers who are responsible for collecting and disseminating information that their teammates need to complete the project. They may also act as advisors and leaders who help and direct other team members to solve difficult problems. To distinguish the two scenarios, we measured the variance of betweenness centrality of individuals in a team. If the variance of individual betweenness centrality is large, this may indicate the presence of a gate keeper. However, if the variance is minimal, the first scenario is more likely where the team as a whole is driving the centrality instead of any particular individual. As shown in Table 4, in addition to betweenness centrality of the project, the range of centralities of individual team members is also positively correlated with project revenue. One standard deviation in the range of betweenness centrality is correlated with 77.24% increase in monthly project revenue, lending evidence of the

presence of a gatekeeper in a project team. A team that consists of a few members with high betweenness centrality is positively correlated with team performance. This demonstrates that network characteristics that are conducive at the personal level do not directly translate to performance at the team level. Using person-level networks, we show that a person with high betweenness centrality is positively correlated with performance. However, a team that consists of solely members who occupies the central position of the networks is not necessarily beneficial for productivity. In fact, we show evidence that this composition is actually negatively associated with performance. Only a team consisting of only a few members with high betweenness centrality is positively correlated with performance.

Table 4:Project-level Email Networks	
Dependent Var.	Monthly Project Revenue
Controls: project Complexity, Line of Business, Months, Regions, max job level, number of people in the project	
Cross-sectional network, linear regression	
Hours	111.5***
	(2.226)
In Degree	102.83
	(106.24)
Out Degree	115.2***
	(23.54)
Betweenness	304.36***
	(100.3)
Betweenness range	77.42***
	(20.13)
Constraint	146.81
	(158.85)
Direct links to mgrs outside of the project	6395***
	(2267)
# Managers in project	2733.9***
	(537.5)
# Managers in project ^2	-682.02***
	(215.3)
Average Gender	4533***
	(1441)
Average Gender2	-4374***
	(1416)
Number of divisions project can access	-513.41***
	(128.55)
Observations	8018
R-squared	0.913
*p<.1, **p<.05, ***p<.001. Huber-white robust standard errors are shown in parentheses	

While the effect of network size as measured by in-degrees and out-degrees tend to be negative if not statistically insignificant for the employee-level networks, at the project level, out-degree is shown to have a positive association with performance (Table 4) where communicating to an additional project is correlated with generating 115.2% increase in monthly revenue. Next, we delve deeper and explore the characteristics of direct contacts to understand what type of projects are especially beneficial for performance. As shown in Table 4, knowing executives who are external to the project is associated with higher productivity where each external executive in a project's contact is associated with $6395 in revenue, controlling for a team's internal human capital and topological characteristics. This reinforces the notion that the human capital inside a project network is correlated with performance, on top of traditional human capital and network topologies. We also explore whether knowing people at various divisions within the organization is correlated with project performance. We find that when a project team is in contact with many divisions, it is actually negatively correlated with performance, suggesting that after controlling for structural diversity, knowing division diversity does not contribute to performance. Overall, it demonstrates the importance of examining the content within a social network including the attributes of nodes and the strength of ties. While having strong relationship to executives can be conducive to performance, knowing people in different division in the organization do not contribute to performance beyond network topological characteristics.

We then explore the effect of having direct managerial involvement in a project. Contrary to our previous finding that connections to external managers outside of the project is beneficial to performance, we find that managers directly involved a project to exhibits an inverted U-shape relationship with project performance. While having managers in a project can be beneficial to project execution, having too many in the same project is negatively correlated with project revenue. Having an additional manager in the project initially shows a positive relationship where each manager is associated with $1175 in revenue. As shown in Table 4 and Figure 1, the coefficient on quadratic of managers is negative, implying a concave relationship, such that more managers in a project team are associated with greater revenue to a point, after which there are diminishing marginal returns, then negative returns to increased number of managers. Perhaps, lacking a clear leadership role intensifies internal debates among managers and derails the consultant from making progress. Our interview with consultants further confirms our hypotheses.

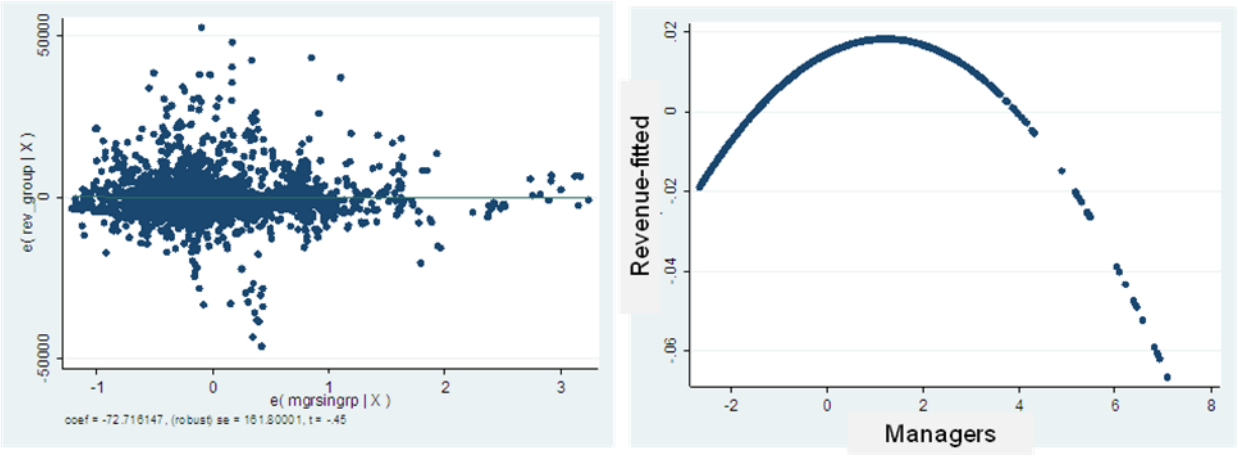

Figure 1: a) Inverted-U relationship between managers in a project and project revenue; b) Fitted value

Interestingly, there is an inverted-U shaped relationship between gender distribution and project revenue as well. While on average a project generates more revenue when there are more male consultants in the team, having too many male consultants is actually negatively correlated with revenue. However, when a quadratic relationship is tested, we find strong evidence of a non-linear relationship (Figure 2). This shows that a team composition with a diverse gender distribution is positively correlated with the success of the project. For our future work, we intend to create a diversity index in the team where we can take into account of other characteristics, such as job level distribution, job function distribution and other tangible team member characteristics to see if diversity has an impact on team performance.

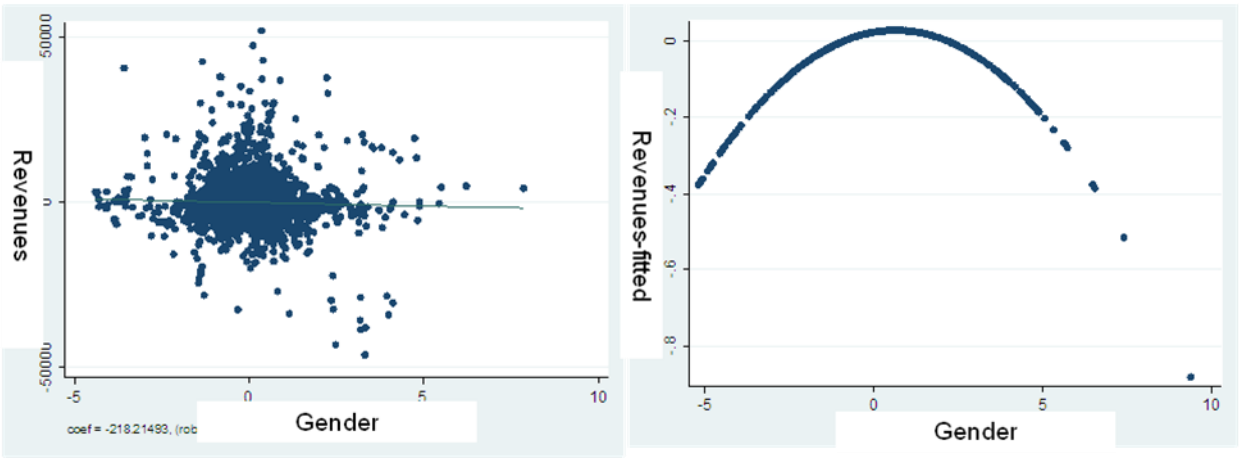

Figure 2: a) Inverted-U relationship between gender and project revenue; b) Fitted value

Project Network vs. Employee Network

In Table 5, we explore the effect of personal network and project network on project revenue. Instead of aggregating revenue at a person-month or a project-month level, each observation is revenue generated per project per person in a given month. When both person-level network characteristics and project-level network characteristics are in a single model, the project-level network characteristics are shown to have a stronger association with revenue than person-level network. As seen in Table 5, most of the coefficients on employee network variables are not significant from zero, while most of the project network variables remain to be statistically significant. T-tests further confirm that the coefficients at the project network level are significantly greater than the coefficients of individual network characteristics.

We find that project betweenness centrality continues to have a positive correlation with revenue. Similarly, we find that network constraint on the project level is also negatively correlated with project revenue. Together these results suggest that projects that can tap into diverse groups to access novel and unique information is more likely to generate more revenue. While the coefficient on the number of managers a project is associated outside of the project is no longer having a statistically significant, having too many managers within the same project continue to have a negative correlation with project revenue, suggesting internal debate within the project team continue to be negatively associate with performance.

Table 5: Person-Project-Level Email Networks	
Dependent Var.	Personal Revenue
Controls: Average Project Complexity, Line of Business, Months, Regions, Job Level , the number of people working in the project	
Hours	129.1*** (3.182)
In Degree	-59.98 (49.45)
Out Degree	135.0*** (49.10)
Betweenness	159.29 (157.12)
Constraint	151.66 (114.9)
gender	-651.2 (414.6)
isManager	9145** (4397)
Managers in Direct Contacts	-2967 (2142)
Project_gender	1957 (1352)
Project gender^2	-1432 (1187)
Number of managers in project	1315.47*** (400.01)
Number of mangers in Project ^2	-236.20** (106.28)
# managers the project members know	-2967 (2142)
Project indegree	-256.2*** (43.63)
Project outdegree	19.47 (21.66)
Project betweenness	331.17** (196.97)
Project constraint	-276.95* (178.93)
Betweeness Range of employees in a project	512.68** (163.75)
Observations	7458
R-squared	0.741

Conclusion and Future Work

By analyzing one of the largest organizational electronic communication networks ever collected and combining it with detailed performance data, we show initial evidence that not only are network topology

characteristics associated with productivity, human capital of inside one's social network and tie strength to the appropriate human capital are also positively correlated with information worker productivity. We find that having strong links to management is associated with higher revenue generation, but simply knowing many managers and have minimal correlations with performance. This demonstrates the importance of distinguishing one's social network not only by its network topologies but also the content of the network such as the cumulative human capital inside the network and the strength of ties connecting to this type of capital.

By leveraging our unique data source, we are able to analyze productivity at the project team level as well. We find that both network diversity and betweenness centrality are strongly correlated with performance, corroborating earlier results. We also analyzed the team composition to work performance and find that the number of executives participating in a project exhibits an inverted-U shaped relationship. While having managers in a project can be beneficial to project execution, having too many mangers in the same project is negatively correlated with project revenue. Perhaps, having too many mangers in a project team signals the lack of leadership role which can intensify internal debates and inhibit project progress.

By constructing social networks at both the individual and project levels, we contrast the properties of individual and project network and find that project network characteristic are more dominant in predicting revenue than individual network characteristics, demonstrating the importance of forming well balanced project teams. In the future, we intend to perform robustness checks, and identify other intermediate mechanisms such as the actual information that is disseminated inside the network and where the expertise is located inside a network. Understanding these mechanisms can increase our understanding of how precisely social networks are related to performance.

References

Aral, S., Brynjolfsson, E., & Van Alstyne, M. 2006. “Information, Technology and Information Worker Productivity: Task Level Evidence.” Proceedings of the 27th Annual International Conference on Information Systems, Milwaukee, Wisconsin.

Aral, S., Brynjolfsson, E., & Van Alstyne, M. 2007. “Productivity Effects of Information Diffusion in Networks.” Proceedings of the 28th Annual International Conference on Information Systems, Montreal, CA.

Aral, S., & Van Alstyne, M. 2007. "Network Structure & Information Advantage" International Conference on Network Science 2007

Burt, R. 1987. “Social Contagion & Innovation: Cohesion versus Structural Equivalence.” American Journal of Sociology, 92: 1287-1335.

Burt, R. 1992. “Structural Holes: The Social Structure of Competition.” Harvard University Press, Cambridge, MA.

Burt, R. 1997. “The Contingent Value of Social Capital”, Administrative Science Quarterly, Vol 42. No. 2

Burt, R. 2004. “Structural Holes & Good Ideas” American Journal of Sociology, (110): 349-99.

Lin,C., Ehrlich, K., Griffiths-Fisher, V., and Desforges, C., SmallBlue: People Mining for Expertise Search, IEEE Multimedia Magazine, Jan.-Mar. 2008.

Coleman, J.S. 1988. “Social Capital in the Creation of Human Capital” American Journal of Sociology, (94): S95-S120.

Freeman, L. 1979. Centrality in social networks: Conceptual clarification. Social Networks 1(3) 215–234. Garguilo, M., and A. Rus 2002 "Access and mobilization: Social capital and top management response to market shocks." Working paper, INSEAD.

Granovetter, M., 1973. "The strength of weak ties." American Journal of Sociology, 6: 1360-1380.

Granovetter, M., 1982. "The strength of weak ties: A network theory revisited." In P. V. Marsden and N. Lin(eds.), Social Structure and Network Analysis: 105-1 30.

Granovetter, M. 1985. “Economic Action & Social Structure: The Problem of Embeddedness.” American Journal of Sociology (91):1420-1443.

Granovetter, M. 1992. “Problems of Explanation in Economic Sociology.” In N. Nohria & R.G. Eccles (eds.), Networks & Organizations: 25-56. Harvard Business School Press, Boston.

Hansen, M. 1999. "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits." Administrative Science Quarterly (44:1): 82-111.

Hansen, M. 2002. "Knowledge networks: Explaining effective knowledge sharing in multiunit companies." Organization Science (13:3): 232-248.

Pentland., A. 2006. “Automatic mapping and modeling of human networks” Physica A: Statistical Mechanics and its Applications.

Podolny, J., and Baron, J. 1997. "Resources and relationships: Social networks and mobility in the work-place." American Sociological Review (62:5): 673-693.

Polanyi, M. 1966. “The Tacit Dimension.” New York: Anchor Day Books.

Reagans, R. and McEvily, B. 2003. “Network Structure & Knowledge Transfer: The Effects of Cohesion & Range.” Administrative Science Quarterly, (48): 240-67.

Reagans, R. and Zuckerman, E. 2001. "Networks, diversity, and productivity: The social capital of corporate R&D teams." Organization Science (12:4): 502-517

Waber, B.N., Olguín Olguín, D., Kim, T., Mohan, A., Ara, K., and Pentland, A. 2007. “Organizational Engineering using Sociometric Badges” International Conference on Network Science, New York, NY.

Wu, Waber, Aral, Brynjolfsson & Pentland “"Mining Face-to-Face Interaction Networks Using Sociometric Badges: Predicting Productivity in an IT Configuration Task”, International Conference on Information Systems, Paris, France, December 14 – 17, 2008.

Uzzi, B. 1996. “The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect.” American Sociological Review, (61):674-98.

Uzzi, B. 1997. “Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness.” Administrative Science Quarterly, 42: 35-67.