The Social Network within a Management Recruiting Firm: Network Structure and Output

NEIL GANDAL*

Harold Hartog School of Government and Policy, Tel Aviv University and CEPR

CHARLES KING III

Department of International Affairs and Government, Georgetown University

Greylock McKinnon Associates

MARSHALL VAN ALSTYNE

Department of Information Systems, Boston University and MIT

Abstract

To understand the relationship between information flows and white-collar output, we collected unique data on email communications to study the network connecting individuals in a management recruiting firm. We also gathered data on revenues and contracts at the individual level. Our empirical results suggest that the size of an individual's internal email network is more highly correlated with output than with the number of email messages, the time spent communicating, the external network size, and with all other measures of communication. This result suggests that a more favorable position in the network structure is associated with higher individual output.

1 Introduction

Social networks play important roles in the functioning of economic and social interaction. A key social network theory is that individuals who are positioned in higher volume information flows (Freeman 1979), or who bridge non-overlapping information pools (Granovetter 1973, Burt 1992) will be better positioned to arbitrage resources or broker opportunities. In other words, a favorable position in the communications network – at gatekeeper, entry, or exit points – should increase a person's output relative to that of people in less favorable positions.

* Contact author: Neil Gandal, Department of Public Policy, Tel Aviv University, Tel Aviv 69978, Israel. Email: gandal@post.tau.ac.il. We are grateful to the editor, Julian Wright, and to an anonymous referee for comments and suggestions that significantly improved the paper. We also thank Dan Ben David and seminar participants at Michigan State University and the Third Bi-annual IDEI Conference on the Economics of the Software and Internet Industries in Toulouse for helpful comments and suggestions. Van Alstyne gratefully acknowledges substantial funding via NSF Career Award 9876233 and Intel Corporation.

We have identified a setting that has measurable inputs and outputs, as well as a simple production process at the individual level, in order to empirically examine whether a favorable position in the network structure is indeed associated with higher output. Our data, from a management recruiting firm, include direct observations on individual use of information technology, as well as objective measures of individuals' dollar-denominated performance.

To undertake this research, we developed unique software instruments that allowed us to capture all computer mediated communication over the course of more than six months at the individual message level. Our primary measure of information flow is email communication.[1]

We used these data to precisely measure various attributes of email communication and social topology. For example, we can directly observe the size of an individual's internal contact network based on the number of unique individuals who sent email to him or her. Additionally, we can calculate the social network measure of "betweenness," which is a normalized count of the number of times an individual appears on the shortest path between all agent pairs. Individuals with higher values of betweenness are positioned at higher-value junctures in the information flow.[2]

Among many rich measures, we constructed measures of (i) the number of email messages sent and received, (ii) unique email contacts outside the firm, (iii) the topological structure of information flows, (iv) message sizes, (v) message initiation and response ratios, (vi) the presence of attachments, and (vii) proxies for time spent on email, phone, and face-to-face communication, respectively.

In addition to the wealth of data on use of information technology, we also have precise measures of individual performance in the form of (i) billing revenues generated on specific contracts, and (ii) the number of contract completions attributed to each person. Our data are sufficiently precise that although a majority of projects are completed by teams, the firm's accounting records include share weighted effort by person, so that we can attribute exact dollar revenues to particular individuals.

The goal of our project is to determine whether heavier information technology use and better network attributes correlate with individual workers' greater success. When a person uses IT, does the time he or she spends on email, his or her communications volume, the frequency with which he or she sends or receives communications, or his or her more frequent use of attachments better predict performance? Further, if heavier use of IT is associated with success, then is the size of the external network, the size of the internal network, or possessing a more favorable position within the network a stronger predictor of success? Management recruiting is an ideal setting for this analysis since, in this context, individual "production" is an extensive process of search and deliberation that matches "potential candidates" with "clients." In this industry, information is critical.

Our results suggest that an individual's internal network size is more highly correlated with revenues generated by that individual than are any of the standard demographic or human capital factors. Internal network size is also a better predictor of revenue than other communications measures such as message volume, external

[1] Secondary data on information technology come from our survey data on perceptions.

[2] For example, individual w26 in Figure 1 has a high value of "betweenness," while individual w98 has a low value.

message size, and declared time spent on email. Interestingly, inflows are better predictors than are outflows. The results are slightly less significant, but robust to using the number of completed contracts – rather than revenues generated – as a measure of success (or output).

After accounting for the individual's number of unique contacts within the firm, the social network measure of betweenness is also highly correlated with an individual's ability to generate revenues. That is, revenue is higher for an individual when the social network measure of betweenness is higher. Individuals with higher values of betweenness appear to be positioned at critical junctures in the information flow. Hence, this result suggests that the individuals positioned in heavier information flows are also more successful. It should be pointed out that the result regarding betweenness is not robust to certain alternative functional forms. In contrast, the result regarding internal network size is extremely robust.

Once one controls for these two variables, it turns out that information volume has little explanatory power.[3] Our main results are that (i) measures of contact network size and betweenness are highly positively correlated with individuals' output, and that (ii) these are much better predictors of success than are several labor variables, including demographic and human capital factors.

We are careful not to attach a causal interpretation to these results because it is not possible to determine from the data whether, on one hand, a favorable network position or a larger contact network increases success, or whether, on the other hand, highly successful individuals are therefore able to obtain a more favorable location in the network and a larger contact network.

Our paper is closely related to Ahuja (2000), Ahuja, Galleta, and Carley (2003), Calvo-Armengol, Patacchini, and Zenou (2008), Ahuja (2000), and Fershtman and Gandal (2008), who also consider the relationship between network structure and performance. Calvo-Armengol et. al. use data on friendship networks and examine how the network structure affects pupils' school performance. Using patents as a proxy for innovation, Ahuja (2000) studies the relationship between the network of technical collaboration among firms in the chemical industry and innovation. Ahuja et. al. (2003) examine the relationship between communication patterns in an R&D setting. They find that network centrality predicts increased publication. Focusing on open source projects, Fershtman and Gandal (2009) examine how the structure of the open source software network affects project success.[4]

This paper enables us to shed light empirically on what a network benefit function might look like, in part by bridging two diverse literatures. Typical models in the IO literature assume that the network benefit function depends directly or indirectly on the number of members the network has. A "network effect," for example, is generally a demand economy of scale (Katz & Shaprio, 1984; Leibowitz & Margolis, 1994) rather than an implication of network topology. Our results provide some support for this model, since we find that network size (whether internal or external) is a better predictor of revenues than are other measures such as message volume,

[3] In addition to the (objective) email measures, we find that higher values of perceived skill in using internal database tools and a higher perceived benefit from face-to-face (FTF) contacts are also correlated with higher individual output, but these effects are not statistically significant.

[4] Our paper is also related to the literature on the role of social networks in the functioning of the economy. For general surveys, see Jackson (2006, 2008) and Goyal (2007).

demographic attributes, and declared time spent on email. However, we do find that, after accounting for network size, the network topology or structure also correlates highly with revenues, as predicted by theories in social network literature (Granovetter 1973, 1974; Freeman 1979; Burt 1992; Bulkley & Van Alstyne 2004). In particular, the "betweenness" index that measures frequency of an individual's appearing on the shortest information path between all pairs of other individuals correlates both with revenues and with project completions, and is significant in most, although not all, specifications. This suggests that the position/location of individuals within a network might affect the network benefit function as well. While this latter effect might not matter much in a passive network (such as, say, the network encompassing holders of a particular credit card), it can matter in networks where communication among members is important. This suggests that network topology matters for network economics.

2 The Setting

We negotiated access to a midsize executive recruiting firm. These firms offer a consulting service in which recruiting experts identify, vet, and recruit executives to fill employee vacancies at the client firms who hire them. The industry standard process involves (i) having a recruiter negotiate a contract with a client; (ii) identifying attributes of an ideal candidate; (iii) identifying potential candidates among databases, client competitors, and similar industries; (iv) conducting due diligence on candidate qualifications; (v) arranging client interviews with a winnowed pool of final candidates; and (vi) negotiating final terms with the client-preferred candidate and persuading candidates to accept. The firm allowed us to instrument their email system, yielding 10 months of useful data. The firm also provided more than a year of project accounting data to us. Additionally, we surveyed workers on their behaviors, perceptions, use of time, and human capital attributes. (See Table A4 for the survey questions.)

Professionals at management recruiting firms generate revenues via both "booking" and "billing," where booking revenues are fees earned from bringing work into the firm, and billing revenues are earned by fulfilling contracts. Typically, more senior staff (e.g., partners) handle client interaction and bookings, while more junior staff (e.g., consultants) handle candidate interaction and billings. Persons at any level, however, may work on tasks at any level, and this is common throughout our data. Since we employ billing revenue as a measure, an important question is how people end up working on particular projects. Our discussions with executives in the host research site suggested that work typically comes into the firm in two ways: In the first case, the firm gets a request for a competitive proposal. Contact with the prospective client is then handled by a group leader for an area defined by industry (e.g., education, health care, etc.), who assembles a team to create and present the proposal. In the second case, an individual brings in the business directly through his/her contacts.[5]

[5] Norms for these cases differ by industry. In health care, for example, repeat business accounts for roughly 90% of all business, so case two applies here. In education, by contrast, repeat business accounts for only roughly 60% of the work; hence, case one would apply relatively more strongly here.

In the first case, that of competitive proposals, the team assembled to generate the business typically continues to work together to execute the contract once it is won. Here, assignment to projects occurs via an "internal market system." In the second case, where an individual brings in business, the "rainmaker" typically makes the decision regarding who is on the team. Since the rainmaker hopes to generate repeat business and has a reputation at stake, it is likely that he/she will assemble the best team possible; in other words, assignments are also likely made via the internal market system. In both cases, revenues are fixed at one-third the salary of a placed executive, representing the external market's valuation of the person found. Income does not depend on the time spent to fill a vacancy, so efficiency requires recruiters to actively manage the flow rate of placements. Hence, in both cases, billing revenues are performance based and are likely to be an appropriate measure of success.

We employed two distinct measures for individual output: The first is individual contract completions. A contract "begins" at the point in time when a client firm signs a contract with the recruiting firm to conduct a search. A contract "ends" when the client firm signs a contract with a job candidate. Thus, start and stop dates are precise. Since teams of recruiters (rather than individuals) conduct most searches, the accounting system also tracks the share-weighted effort of each person on a contract. A person's compensation is then proportional to his or her effort on each contract.

The second measure of output is the dollar-denominated revenue attributed to each person. Each recruiting project brings in an amount specified by contract, and that amount is proportional to the value of the position to be filled. Thus, looking to individuals' share-weighted effort on each contract allows us to allocate the revenues generated by individuals who have worked in teams. "Effort shares" are determined primarily by a combination of internal market forces and observation. The firm, at the time the data was gathered, was more than 25 years old. Accordingly, it had developed standardized project weights for tasks such as project management, client interaction, candidate generation, interviews, etc. These form a baseline. At the start of the project, the person who lands the contract can adjust these percentages slightly, in order to attract other consultants to the project. Adjustments are based on the duties that these other consultants will assume, and their opportunity costs on other projects to which they are currently assigned. (A great teammate overloaded with projects will cost more.) Upon project conclusion, baseline effort shares are again adjusted if effort was disproportionally large or small relative to the norm for similar projects.

In summary, ex ante, the firm uses an established formula based on expected effort and opportunity costs. Ex post, this formula is adjusted based on observed effort.[6] Given this discussion, it is not surprising that the correlation between these output measures (revenues and contract completions) is very high, 0.89.

3 Model and Data

We employ a simple model of the form $Q_i = \alpha + \beta H_i + \gamma X_i + \delta Y_i + \varepsilon_i$, where Q_i is Output ($, Completed Contracts); H_i includes job rank variables (i.e., Partner, Consultant, etc.); X_i includes human capital variables (i.e., Education, Experience,

[6] This suggests that the result that internal network size is highly associated with output is not driven by an internal bargaining process.

etc.); Y_i includes information technology variables (Network Size, Betweenness, Email Volume, etc.); and ε_i is white noise. Although we employ a linear model, as we discuss in section 4.3., our main results are robust to alternative functional forms.

Y_i is composed of IT use variables like internal network size, betweenness, email volume, etc. As we discussed in the introduction, IT use variables are possibly endogenous in the model, since highly successful workers might use IT more intensively, or intensive users might be more successful.[7] Since we do not have exogenous instruments for measuring the IT use variables, we are only able to examine the raw correlations in the data, rather than examining causality. Since the main goal of our project is to determine which IT use variables are most highly correlated with success, we believe that the OLS regressions are sufficient for our analysis. Again, we are careful not to attach a causal interpretation to the results.

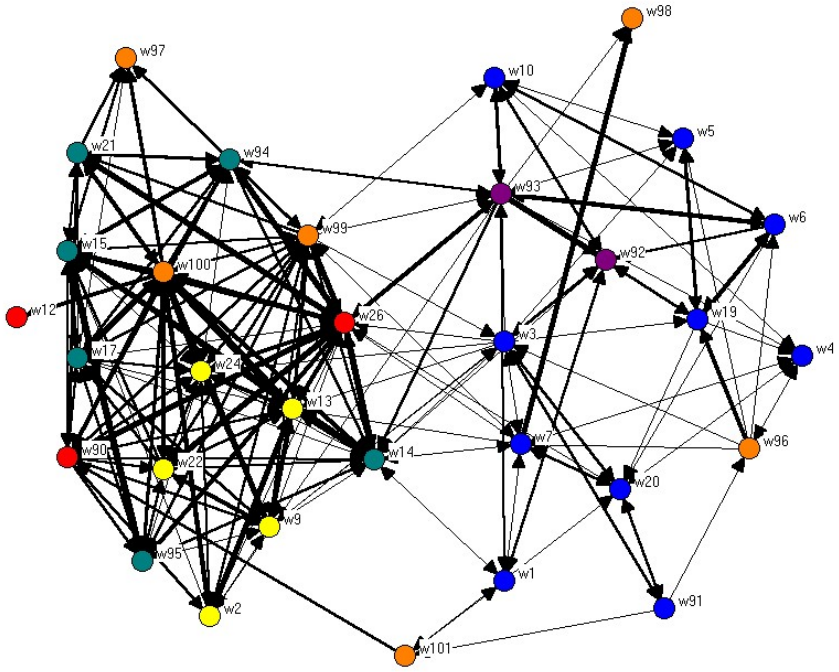

Figure 1 – The observed contact network over a six week period[8]

Data for this study come from three sources: Communications data represent direct observation of all email traffic on a corporate mail server from August 2002 to June 2003. Raw data were encrypted prior to disclosure to ensure privacy, and each voluntary participant was paid $100 if he or she chose not to opt out of the study.[9]

[7] Likewise, people who bring in more revenues might be more popular, or people who are more popular might bring in more revenues.

[8] In order to preserve firm confidentiality, we cannot reveal the firm or its industry sectors, but sectors include some of the following industries: consumer products, real estate, health care, computer industry, and education. Estimated coefficients on dummy variables for sectors are insignificant in the regressions below; hence, given the limited number of observations, we do not include them in the analysis. Email illustration (Figure 1) is from outside the sample period.

[9] See Van Alstyne, M. & Zhang, J. (2003) for a description of the development of tools used to gather these data. See Van Alstyne & Zhang US Patent 7,503,070 for a mechanism to permit analysis of

Figure 1 shows the observed communications network from one six week interval. Nodes represent individual people; links represent messages between individuals, while thicker lines designate more frequent communication. Individuals such as w26 have both a relatively high number of incoming contacts and also a relatively high index of "betweenness."

Perception data were gathered using a 52 question online survey. Participants were contacted by the chief technology officer, received $25 for completed surveys, and were offered a chance to view their own responses, ranked against those of average respondents, upon conclusion of the study. Voluntary participation in both activities (the email activity and the survey) exceeded 85%.[10]

We have data on individual billing revenues and completed contracts from 2002 (the full year) and 2003 (the first half of the year). The regressions with 2002 revenues as the dependent variable fit the data much better than the regressions using 2003 revenues as the dependent variable.[11,12]

3.1 Variables employed in the study

A description of the variables we employ appears in Table 1 below. For clarity in notation, we use the following conventions in defining our variables: "internal" and "external" refer, respectively, to contacts located inside or outside the firm; "in" and "out" refer, respectively, to the direction of the communication, where "in" means receiving and "out" means sending; "net" refers to the number of unique network contacts; and "vol" refers to message volume.

Summary Statistics for these variables are shown in Table 2 below. The Table shows that the standard deviation and the range of BETWEENNESS are both quite high, suggesting that there some people who are very central to information flows, and others who are not essential at all (See figure 1.) The size of an individual's internal contact network also ranges quite a bit. The demographic variables show that this is highly educated group of individuals; each has completed at least four years of study beyond high school.

Of all of the IT variables, INTERNAL IN-NET and BETWEENNESS are the ones most highly correlated with REVENUES (0.43). Interestingly, these variables are not that highly correlated with each other (0.46). We examine the correlations between INTERNAL IN-NET and the alternative measures of email in section 4.1.

---

communication content that preserves privacy. See Bulkley & Van Alstyne (2004) for development of competing alternative hypotheses.

[10] The differences between participants and non-participants in terms of billing revenues and completed contracts are not statistically significant. Hence, there is no selection issue.

[11] The correlation between revenues from 2002 and completed contracts completed is very high (0.89), while the correlation between completed contracts and revenues from 2003 is relatively low (0.25).

[12] Pursuant to our contract on the use of human subjects, no data on individual identities are available as a result of this study.

REVENUES - Revenues in dollars billed by the individual for completed contracts in 2002.	**COMPLETED CONTRACTS** – Number of contracts completed during the August 2002-June 2003 period. Full-time equivalents are based on the share of revenues attributed to an individual.[13]
SALARY – 2002 salary in dollars.	**EDUCATION** - Years of education.
AGE – Age of the individual.	**EXPERIENCE** – Years of experience in the industry.
GENDER - A dummy variable that takes on the value 1 if the individual is male, and zero if she is female.	**PROJECTS** – The total number of projects on which an individual is working, averaged over two-week intervals and weighted by the fraction of reported effort on that project.
PARTNER - A dummy variable that takes on the value 1 if the individual is a partner, and zero otherwise.	**CONSULTANT** - A dummy variable that takes on the value 1 if the individual is a consultant and zero otherwise.[14]
INTERNAL (EXTERNAL) IN-NET - Size of the individual's internal (external) contact network, as measured by the number of unique individuals within the firm including support staff (from the outside) who *sent email to* the relevant individual. This includes cc'd messages.	**INTERNAL (EXTERNAL) OUT-NET** - Size of the individual's internal (external) contact network, as measured by the number of unique individuals within the firm (from the outside) who *received email from* the relevant individual. This includes cc'd messages.
INTERNAL (EXTERNAL) IN-VOL – Daily average volume of incoming mail from contacts inside (outside) the firm.	**INTERNAL (EXTERNAL) OUT-VOL** – Daily average volume of outgoing mail to contacts inside (outside) the firm.
INTERNAL (EXTERNAL) IN-MSG-SIZE – Daily average size (in bytes) of messages from contacts inside (outside) the firm.	**INTERNAL (EXTERNAL) EMAILS WITH ATTACHMENT** – Daily average number of messages with an attachment received by an individual.
BETWEENNESS – A normalized count of the number of times an individual appears on the shortest path between all agent pairs, including staff. A link between two agents exists if they exchanged 30 or more messages over six months. (There is very little difference in results if we use 10, 20, 30, 40, or 50 messages as a link.)	**SEARCH TOOLS** – A variable that that takes on a value from 0 to 500, where the individual is asked to agree or disagree with the following statement: "I am highly effective at using our in-house proprietary search tools. This means I know what information they contain and I can easily find, add and modify the records I need." A higher value of the variable means stronger agreement with the statement.
PFTF VALUE – The perceived (percent) value from face-to-face contacts. This value can range between 0 and 100, but perceived value on all measures of communication (face-to-face, telephone, email, instant messenger, computer display and hardcopy) was scaled to total 100.[15]	**PFTF TIME** – The declared percent of time spent on face-to-face contacts. This value can range between 0 and 100. Further, the percent of time spent on all measures of communication (face-to-face, telephone, email, instant messenger, computer display and hardcopy) was required to total 100.[16]
PTEL VALUE – The perceived (percent) value from using the telephone.	**PTEL TIME** – The declared percent of time spent on the telephone.
PEMAIL VALUE - The perceived (percent) value from using email.	**PEMAIL TIME** – The declared percent of time spent on email.

Table 1: Variables and their Definitions

[13] A contract is complete when a candidate employee signs a contract with the client firm. When that occurs, the management recruiting firm receives its fee.

[14] The other individuals in the study are "researchers." While partners are higher up in the firm than consultants, our empirical results would not qualitatively change if we lumped those two categories together. See footnote 19 below.

[15] The average perceived *value* from the sum of telephone, face-to-face, and email communication exceeds 90%, leaving less than 10% for hardcopy, instant messaging, and "other" communications. Hence, we focus on these three, rather than looking at all six measures.

[16] The average perceived percent of *time* spent on telephone, face-to-face, and email communication exceeds 85%, leaving less than 15% for hardcopy, instant messaging, and "other" communications. Hence, we focus on the first three, rather than looking at all six measures.

The question that generated PTF VALUE, PTEL VALUE, and PEMAIL VALUE was as follows: "The relative amounts of value I get from different modes of information gathering and communication are (best guess):"[17] Similarly, the question that generated PTF TIME, PTEL TIME, and PEMAIL TIME was as follows: "The relative amounts of time I spend on different modes of information gathering and communication are (best guess):" If we define a new variable to be the ratio of "perceived percent value" to "declared percent of time," we (unsurprisingly) find that the mean of this variable is highest for face-to-face contacts (2.5). Interestingly, the mean of the "email" ratio variable (1.1) is higher than the mean of the telephone ratio variable (0.9). We discuss the effect of perceptions in section 4.3.

Variable	Mean	Standard Dev.	Minimum	Maximum
REVENUES	435,695	140,119	211,353	773,280
SALARY	249,028	117,489	70,783	510,027
GENDER	0.58	0.50	0	1
AGE	47.09	9.06	28	64
EDUCATION	17.78	1.36	16	21
EXPERIENCE	15.91	9.14	3	39
PARTNER	0.45	0.51	0	1
CONSULTANT	0.48	0.51	0	1
INTERNAL IN-NET	69.15	10.04	43	87
INTERNAL OUT-NET	47.67	15.26	13	83
EXTERNAL IN-NET	879.03	709.71	131	2483
EXTERNAL OUT-NET	297.61	295.07	35	1439
INTERNAL IN-VOL	7.08	2.68	3.21	12.03
INTERNAL OUT-VOL	4.51	2.52	0.69	10.82
EXTERNAL IN-VOL	15.54	10.17	3.61	47.3
EXTERNAL OUT-VOL	4.91	3.50	0.4	15.07
INTERNAL IN SIZE	37.02	13.36	11.5	75.89
EXTERNAL IN SIZE	33.95	12.93	15.58	62.08
INTERNAL IN ATTACH	4.29	1.57	1.4	8
BETWEENNESS	378.32	364.43	0	1625.72
SEARCH TOOLS	318.58	98.66	86	467
PROJECTS	4.70	2.13	1.5	10.22
COM CONTRACTS	6.04	2.29	1.15	10.38
PFTF VALUE	33.96	19.48	0	80
PTEL VALUE	35.37	15.08	10	70
PEMAIL VALUE	20.81	11.51	0	50
PFTF TIME	19.61	15.80	0	75
PTEL TIME	43.61	16.81	10	70
PEMAIL TIME	22.88	11.97	5	50

Table 2: Descriptive Statistics

[17] The questionnaire used to generate all perceptual measures is provided in Table A4 in the Appendix.

4 Empirical Results

The first regression of Table 3 contains our "first" preferred model. This "preferred" empirical model is not preferred for any a priori reason, but rather because it yields the most statistically significant results. We discuss alternative functional forms and show in section 4.3 that these functional forms yield qualitatively (very) similar regression results. Hence, the results are quite robust to functional form.

Two IT variables, INTERNAL IN-NET and BETWEENNESS, are included in this "preferred" regression, in addition to PARTNER and CONSULTANT, as well as the number of projects on which an individual is working. The adjusted R-squared is 0.68, while the R-squared itself is 0.73. (The correlations among the variables in this regression appear in Table A2 in the Appendix.)

The estimated coefficient on INTERNAL IN-NET is positive and statistically significant (6,024.4, t=3.39 in the first regression in Table 3.) This suggests that, controlling for the number of projects and rank, success is greater for individuals who have email contact with a large number of other workers in the firm. The coefficient on INTERNAL IN-NET suggests that the addition of one person in the internal incoming network is associated with a $6,024 increase in revenues to the firm.[18]

Even after controlling for an individual's number of unique contacts (INTERNAL IN-NET), the BETWEENNESS measure is also statistically significant (t=2.40 in the first regression in Table 3.) This suggests that individuals are more successful when their (social network measure of) betweenness is higher, that is, when more information flows through that person. Being better positioned in the network also matters because it affects information flow among people.[19,20] Adding age, education, years of experience, and gender to the first regression in Table 3 does not qualitatively change the results, but the adjusted R-squared falls from 0.68 to 0.65 when these four variables are added. For this reason, and because we have a limited number of observations, we do not include these variables in further analysis.[21]

The second regression in Table 3 shows an alternative preferred model that includes variables measuring the perceived benefit of other means of communication and skills (PFTFVALUE, PTELVALUE, and SEARCH TOOLS). We include these variables because we have no objective measures of the benefits from the use of the telephone, face-to-face contacts, or the ability to use the internal database. Our main results -- that an individual is more successful when he/she has a larger number of

[18] It is important to point out that this is not necessarily a causal relationship.

[19] Not surprisingly, the coefficient on PROJECTS is positive and significant as well. That is, controlling for IT use, those working on more projects have higher revenues.

[20] Since 48% of the observations are consultants and 43% of the observations are partners, the base category (that of researchers) is very small. Hence, we re-ran the first regression in Table 1 without the variable "PARTNER." The key results are qualitatively unchanged. In the first regression in Table 1, the coefficients on both INTERNAL IN-NET (t=2.67) and BETWEENNESS (t=2.36) are statistically significant. Since the adjusted R-squared of this regression (without PARTNER) is lower, we prefer to include both PARTNER and CONSULTANT in the regression.

[21] See Appendix Table A1 for a preferred regression model including demographic variables.

unique internal contacts, and when his/her BETWEENNESS measure is higher -- are robust to whether the perception variables are included or excluded.[22]

	Regression 1: Preferred Model		Regression 2: Preferred Model With Perception Variables	
Independent Variables	Coefficient	T-stat	Coefficient	T-stat
CONSTANT	-355,896.3	-2.49	-527,067.2	-3.18
INTERNAL IN-NET	6,024.4	3.39	7,925.1	4.45
BETWEENNESS	104.8	2.40	77.5	1.79
PROJECTS	28,316.7	4.12	23,636.1	3.51
PARTNER	148,431.1	2.40	183,064.8	2.79
CONSULTANT	277,978.3	4.31	333,460.6	4.83
SEARCH TOOLS			168.79	1.13
PFTF VALUE			890.31	0.93
PTEL VALUE			1542.91	0.125
N of observations	33		33	
R-squared	0.73		0.80	
Adj. R-squared	0.68		0.73	

Table 3: Regression Results: Dependent Variable: REVENUES

4.1 Alternate measures of email use: Importance of internal networks

A fairly striking but perhaps intuitive result is that the strength of the correlation between success and email use depends on how email is used. Regressions that replace INTERNAL IN-NET with an alternate information technology measure as an explanatory variable yield weaker statistical significance. Results from each of these replacements are summarized in Table 4 and discussed below. Alternative information technology measures include contact location (inside versus outside the firm), volume (messages versus network size), and directionality (sending versus receiving). These alternatives show weaker (if any) statistical significance, and the adjusted R-squared of the regression falls. [23]

For these alternative email measures, it is interesting to briefly examine their correlations with INTERNAL IN-NET.[24] Some of the measures are (unsurprisingly) quite strongly correlated with this variable. The highest correlation is between INTERNAL IN-NET and INTERNAL OUT-NET (0.75,) while the correlation between INTERNAL IN-NET and INTERNAL IN-VOL is 0.69, and the correlation between INTERNAL IN-NET and INTERNAL OUT-VOL is 0.54. Interestingly, the correlation between INTERNAL IN-NET and PEMAIL TIME (the declared time

[22] The estimated coefficient on BETWEENNESS is statistically significant at the 90% level of confidence in the second regression in Table 2. If we include relative time spent, rather than perceived FTF or Telephone value, these two variables are completely insignificant and the other estimates remain virtually unchanged. See Section 4.3 for further discussion of the second regression in Table 2.

[23] We obtain qualitatively similar results if we use the second regression in Table 2.

[24] The correlations among the alternative measures of email from Table 3 are shown in Table A3 in the Appendix.

spent on email) is relatively small (0.32), suggesting that the size of the internal contact network does not necessarily create a burden to "keep in touch."

Variable included in first preferred regression	*T-Statistic*	*Adj R^2*
INTERNAL IN-NET	4.45	0.68
Internal In-Net excluded and replaced by	*T-Statistic*	*Adj R^2*
INTERNAL OUT-NET	2.34	0.62
EXTERNAL IN-NET	-1.35	0.57
EXTERNAL OUT-NET	-0.20	0.55
INTERNAL IN-VOL	2.46	0.62
INTERNAL OUT-VOL	0.91	0.55
EXTERNAL IN-VOL	-0.80	0.55
EXTERNAL OUT-VOL	0.45	0.54
INTERNAL IN-MSG-SIZE	-0.62	0.55
EXTERNAL IN-MSG-SIZE	0.28	0.54
INTERNAL EMAILS WITH ATTACHMENT	1.45	0.57
PEMAIL TIME	-0.42	0.54
NO OTHER VARIABLE INCLUDED		0.57

Table 4 – Using alternate explanatory variables, instead of INTERNAL IN-NET

In terms of Table 4's results, is success more highly correlated with sending or with receiving email? We find evidence favoring the latter variable. When we replace INTERNAL IN-NET with INTERNAL OUT-NET, we find that the internal outgoing network is statistically significant (t=2.34), but the predictive power of the model falls (to an adjusted R-squared of 0.62 from 0.68). Perhaps more importantly, when we include both variables in the regression, INTERNAL IN-NET is statistically significant (t=2.24), while INTERNAL OUT-NET is not significant (t=0.43).[25] This suggests that the output of an individual is more highly correlated with the information flowing to the individual, than with the information flowing from the individual. This stands in interesting contrast to the results of Ahuja et. al (2003), who found that people who contributed more to a discussion group had higher publication rates than people who sought information. We interpret their finding to be a signal of expertise in a public forum. In contrast, our data include *all* email communication, which might be more indicative of information consumed as input to an individual white-collar worker's production function.

A variant on the directionality issue explores the *volume* of email sent and received, and whether this communication is by persons inside or outside the firm. Higher internal email volumes, for example, might be consistent with task delegation. Substituting INTERNAL IN-VOL for INTERNAL IN-NET, the estimated coefficient is positive and statistically significant (t=2.46), but the predictive power of the regression falls; the adjusted R-squared is 0.62. Further, when we include both INTERNAL IN-NET and the INTERNAL IN-VOL in the regression, INTERNAL IN-NET is statistically significant (t=2.13), while the INTERNAL IN-VOL is not

[25] The regressions with INTERNAL IN_NET and the other variables are not shown in Table 4. They are available on request.

significant (t=0.47). This suggests that output is more highly correlated with internal network size, than it is with the volume of email.

Substituting INTERNAL OUT-VOL for INTERNAL IN-NET, the estimated coefficient is positive, but insignificant (t=0.91), and the predictive power of the regression falls significantly. (The adjusted R-squared is 0.55.) Including both INTERNAL IN-NET and the INTERNAL OUT-VOL in the regression, INTERNAL IN-NET is statistically significant (t=3.17), while the INTERNAL OUT-VOL is not significant (t=-0.26).

Another possibility is that the volume of external information received matters. Receiving external information might loosely be interpreted as conferring an advantage based on having more frequent or more recently-updated news. To explore this possibility, we substituted EXTERNAL IN-VOL; this leads to an insignificant relationship with success (t=-0.80) and the predictive power falls; the adjusted R-squared is 0.55. Similarly Table 4 shows that external network size (measured by both EXTERNAL IN-NET and EXTERNAL OUT-NET) is insignificant. This suggests that internal network size matters more than the volume of external information flowing to or from individuals in the firm.

If, instead of volume, we substitute message size, there is no meaningful relationship with output, regardless of whether this is internal incoming message size (t=-0.62) or external incoming message size (t=0.28). We also examined the effect of included attachments. This might represent information either in template form, or in alternative presentation formats. If we replace INTERNAL IN-NET with the "number of email messages with an attachment," this variable has a positive but insignificant relationship with success (t=1.45), and the adjusted R-squared is lower than the regression with INTERNAL IN-NET. When we put both of those variables in the regression, INTERNAL IN-NET is significant (t=3.09), while the "number of internal email messages received with an attachment" is not significant (t=0.93). That is, the internal network size matters more than does message size or whether the email has an attachment.

Finally, when we replaced INTERNAL IN-NET with the declared percent time spent on email (PEMAIL TIME), the latter was insignificant (t= -0.42). When we put both of those variables in the regression, INTERNAL IN-NET is statistically significant (t=3.45), while the (declared) percent of time spent on email is not significant (t=-0.89). Controlling for the internal network size, declared time spent on email is, if anything, negatively correlated with success.

In summary, incoming messages matter more for success than outgoing messages do. The message's size, or whether the message has an attachment, is not correlated with success. Sending or receiving a large message volume is much less important for success than the size of one's internal contact network within the firm.

4.2 Internal vs. External Network

One counterintuitive result is that the internal network seems to be more important than the external network. At first glance, this might seem surprising, since one might think that a recruiting firm searching for suitable candidates would especially benefit from an external contact network. However, this result is likely explained by IT substitution occurring more for external contacts.

First, consider corroborating evidence from the survey. The empirical results are consistent with supplemental data on perceived relative values of information sources. The average responses to the statement "In terms of relative value, the best information comes from these sources" are as follows.

People within the firm 35.8%
Our internal database: 26.4%
People outside my firm: 20.1%
Public access Web pages: 9.2%
News or Trade press: 5.6%
External proprietary databases: 2.9%

In the past, it was likely that less information of value came from sources other than word-of mouth, i.e., the comments of people either inside or outside of the firm. But given the rapid growth of the Internet, as the above numbers show, more valuable information likely comes both directly from Web pages and indirectly from the Internet, in the form of databases and easily accessed news/trade press. These sources are likely substituting effectively for contacts outside of the firm.

Further, the *number* of external contacts is probably not a good measure of the value of outside information. It is likely that one or two key external contacts do make a difference for an internal consultant, but this would not show up in the data.[26]

In a similar vein, when we add a variable measuring the number of people who are included in a personal Rolodex or Palm Pilot on the right-hand side to the regressions in Table 3, the estimated coefficient is negative, suggesting that – controlling for internal network size and position in the information flow (betweenness) – an increase in the number of personal contacts does not lead to increased productivity.

4.3 The effect of perceptions

When we add three perception variables to the first regression in Table 3 (participants' own perceived skill with SEARCHTOOLS,[27] perceived value of face-to-face contacts, and perceived value of telephone use), we obtain the second "preferred" regression in Table 3. The adjusted R-squared increases to 0.73 relative to the model without the perception variables. The second regression, in Table 3, shows that the effect of perceived skill with SEARCHTOOLS on output is positive, but not significant (t=1.13). The regression also shows that high values placed on face-to-face (FTF) interactions are associated with higher output, but the effect is also not significant (t=0.93).[28]

[26] Thus, our results do not necessarily contrast with those of Granovetter (1973, 1974) which showed the importance of weak (indirect) ties. This is because some external contacts undoubtedly form a bridge to a new community, with fresh information on job opportunities, while others provide information of negligible additional value.

[27] One could argue that SEARCHTOOLS should be included in the first regression because it is perception of skill, rather than value. Our results are robust to this change.

[28] The theoretical literature (Seeley 2001) suggests that FTF contact is a good starter for a relationship, but that email can then sustain the relationship. (Email does not sustain a relationship as well, however, if it is used as a starter.) If we eliminate perceived value from telephone use from the second

When we exclude INTERNAL IN-NET and BETWEENNESS, but include all three of the perception values discussed above, as well as the perceived value of email use and PROJECTS, we find that the coefficients associated with PFTF VALUE and PEMAIL VALUE are both positive and statistically significant at the 10% level (t=1.76 and t=1.70 respectively), while the estimated coefficient on PTEL VALUE is positive, but not statistically significant (t=0.68). The adjusted R-squared of the regression is only 0.38.

4.4 Robustness to functional form

In terms of model specification, we find that the linear model performs better than models with different functional forms. This finding mirrors that of Ichniowski, Shaw & Prennushi (1997), who found that the linear form better described blue-collar productivity on steel-finishing lines. The results are qualitatively similar, with the exception that BETWEENNESS is no longer statistically significant. Taking the natural logarithm of all of the quantitative variables, the adjusted R-squared of the "log/log" model, using the first regression in Table 3, is slightly lower than that of the linear model (0.68 vs. 0.66), and BETWEENNESS is no longer statistically significant (t=1.03). See Table 5. Similar results obtain when one employs the log/log model using the second regression in Table 3. The adjusted R-squared is lower than that of the linear model (0.73 vs. 0.70), and BETWEENNESS is not statistically significant (t=0.33).

Dept. Variable L_REVENUES	Regression 1: log/log model			Regression 1: log/linear model	
Indep. Variables	Coefficient	T-stat	Indep Variables	Coefficient	T-stat
CONSTANT	8.05	6.94	CONSTANT	11.06	29.22
L_INTERNAL IN-NET	0.89	3.18	INTERNAL IN-NET	.014	3.08
L_BETWEENNESS	0.021	1.03	BETWEENNESS	.00021	1.80
L_PROJECTS	0.39	5.06	PROJECTS	.071	3.87
PARTNER	0.37	2.46	PARTNER	0.35	2.13
CONSULTANT	0.62	3.97	CONSULTANT	0.62	3.64
N of observations	33			33	
R-squared	0.72			0.67	
Adjusted R-squared	0.66			0.61	

Table 5: Regressions Using Alternative Functional Forms[29]

In the case of a "log/linear" model -- i.e., a model taking the natural logarithm of the dependent variable, but using the explanatory variables in levels -- we find that the adjusted R-squared of the log/linear model using the first regression in Table 3 is lower than that of the linear model (0.68 vs. 0.61). In this case, BETWEENNESS is

regression in table 3, the coefficient on PFTF VALUE is positive and statistically significant. We do not have sequencing here, but this result would be consistent with the theoretical literature on this issue.

[29] A variable with an "L_" in front of it is the natural log. For example, L_INTERNAL IN-NET is the natural log of INTERNAL IN-NET.

again statistically significant (at the 90% level of confidence.) Hence, the results are qualitatively similar to those of the linear model. When employing the log/linear model using the second regression in Table 3, the adjusted R-squared is lower than that of the linear model (0.73 vs. 0.64). The results are again qualitatively similar, with the exception that BETWEENESS is no longer statistically significant (t=1.15).

4.5 Robustness to alternative dependent variables

We now examine the results using an alternative dependent variable. We find that most of the variables in the two preferred regressions in Table 3 are statistically significant when we use COMPLETED CONTRACTS as the measure of output (or success) instead of using REVENUES as the measure. The predictive power of the regressions in Table 6 is lower, however, than the predictive power of those in Table 3; the first model in Table 6 has an adjusted R-squared of 0.55 (vs. 0.68 for the corresponding model in Table 3). When we add the perception values, the results in Table 6 are similar to the results in Table 3. In this case, all of the variables in the preferred regression are statistically significant and SEARCHTOOLS is statistically significant as well. The other perception variables have the same signs as in Table 3, and are again statistically insignificant.

	Preferred Regression Without Perception Variables		Preferred Regression With Perception Variables	
Independent Variables	Coefficient	T-stat	Coefficient	T-stat
CONSTANT	-2.71	-0.98	-7.56	-2.35
INTERNAL IN-NET	0.042	1.21	0.79	2.29
BETWEENNESS	0.0021	2.46	0.0016	1.95
PARTNER	1.86	1.55	2.49	1.96
CONSULTANT	4.03	3.23	4.80	3.59
PROJECTS	0.49	3.64	0.37	2.80
SEARCH TOOLS			0.0059	2.05
PFTF VALUE			0.022	1.20
PTEL VALUE			-0087	-0.37
N of observations	33		33	
R-squared	0.62		0.71	
Adjusted R-squared	0.55		0.62	

Table 6: Completed Contracts as a Measure of Success (or Output)

Finally, the importance of having an output variable such as REVENUES, rather than having a factor price such as salary, can be illustrated by running the preferred regressions in Table 3 with SALARY as the dependent variable. In regressions identical to those in Table 3, but with SALARY as the dependent variable, the estimated coefficients on INTERNAL IN-NET are positive, although not statistically significant; the estimated coefficients on BETWEENNESS are completely insignificant; and the adjusted R-squared values are lower than the corresponding regressions in Table 3. This is not surprising, given that the correlation between SALARY and REVENUES is only 0.26. This result, and the relatively low

correlation, both illustrate why it is important to use a "true" measure of success (REVENUES), rather than a proxy for success (SALARY).

Taken together, the robustness results in sections 4.5 and 4.6 suggest that INTERNAL IN-NET is very robust to alternative functional forms, as well as employing Completed Contracts as a measure of success. While the result regarding BETWEENNESS is robust to employing Completed Contracts as a measure for success, the result regarding betweenness is not robust to some alternative functional forms.

5 Conclusions and Further Discussion

The goal of this research was to determine whether the increased use of information technology and better network attributes correlate with individual white-collar output. Does time spent on email, communications volume, sending messages or receiving them, or more frequent use of attachments best predict knowledge worker performance? Is the size of the external network, the size of the internal network, or a more favorable position within the network the strongest predictor of output?

Using data on management recruiting, we find evidence that, controlling for the number of projects and job rank, individual success is positively correlated with (i) the number of unique internal contacts, and (ii) betweenness, although the result regarding betweenness is not significant in all alternative specifications. Our results can be measured in real dollar revenues and in completed projects. Surprisingly, revenues are more highly correlated with the size of an individual's internal email network than with standard demographic and human capital measures. Also noteworthy is the fact that internal social networks are better predictors of output than are external social networks. This is consistent with employee perceptual data. Further, these two network metrics are better predictors than the number of messages, email attachment frequency, employees' declared time spent on email, and employees' declared time spent in face-to-face meetings. These results are consistent with theories that white-collar output increases with advantageous network position. Finally, we note that information received predicts output better than does information sent.

Our findings are based on a unique data set that affords precise and objective measures of individual performance, information flows, and contact networks; in other words, it has measurable inputs and outputs at the individual level. The data set is also unique because of the necessity of developing tools to gather all computer mediated communications at the individual level. The software tools we developed addressed data capture, privacy preservation, and security maintenance without incident. To the best of our knowledge, this is the first study to successfully measure individual message-level communication and correlate that measure with individuals' dollar-denominated output.

There are two caveats to these findings: (i) We have data on only 33 individuals; and (ii) we cannot (and do not) attach a causal interpretation to our results. Regarding (i), it is true that we only have data from one firm. The data we do have, however, are incredibly rich and unique. We have all of the email messages sent and received by these individuals and we also know, among other things: (a) whether messages were

sent from or received by people inside/outside the firm; (b) the size of an individual's internal network, in terms of unique email contacts inside the firm; (c) the size of the individual's external network in terms of unique email contacts outside of the firm; (d) the topological structure of information flows; and (e) the size of email messages sent and received. In addition to the wealth of data on employees' use of information technology, we also have precise measures of individual success in the form of (1) billing revenues generated on specific contracts, and (2) the number of contract completions attributed to each person.

Regarding point (ii) above, the main goal of our project is to determine whether a favorable network position or a larger contact network predicts success, rather than to establish causality (which is typically very hard to show empirically). These results are strong, and have the potential for replication in other settings. We believe that the results are interesting because they show (i) which measures of information-technology use (i.e., internal network size and not email volume) are most highly correlated with success; and (ii) that (controlling for contact network size) the social-network measure of betweenness is associated with success as well. Hence, although the small sample size and the inability to determine causation are limitations, we believe that the unique aspects of the data set are such that they outweigh these limitations.

While one cannot draw strong general inferences from one firm's data, results based on management recruiting may be reasonably representative of various white-collar occupations. Professions with similar project work include sales, accounting, fund- raising, law, medicine, real estate, and consulting. These professions each involve case-based problem solving and an extensive process of search and deliberation in order to complete time bounded white-collar tasks. In such cases, professional networks can prove to be a useful resource (Rangan 2000; Barr 2000). Thus, while the magnitudes of coefficient estimates are likely to be industry specific, results regarding correlation between IT use, information flows, and success might generalize to various similar white-collar industries. We anticipate that future research will help establish stronger causal relationships, and will help identify further factors governing the important connections among individual technology use, social networks, and information flows.

6 References

Ahuja G. , (2000), "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study," *Administrative Science Quarterly* 45: 425-455.

Ahuja, M. K., Galleta, D. F., & Carley, K. M. (2003),. "Individual centrality and performance in virtual R&D groups: An empirical study,". *Management Science*, 49, 21-38.

Barr, A. (2000) "Social Capital and Technical Information Flows in the Ghanaian Manufacturing Sector;" *Oxford Economic Papers*, 52:3, pp. 539-559.

Bulkley, N. & Van Alstyne, M. (2004). "Why Information Influence Should Productivity". (2004) *The Network Society: A Global Perspective*; Manuel Castells (ed.). Edward Elgar Publishers. pp: 145-173.

Burt, R. S. (1992). Structural Holes: The Social Structure of Competition. Networks and Organizations. N. Noria and R. G. Eccles. Cambridge, Mass., Harvard University Press: 57-91.

Calvo-Armengol A., E. Patacchini and Y. Zenou (2008), "Peer effect and social networks in education" *Review of Economic Studies*, forthcoming.

Fershtman, C., and N. Gandal, 2008, "R&D Spillovers: The 'Social Network' of Open Source Software," available at http://www.tau.ac.il/~gandal/Research.htm.

Freeman, L. (1979), "Centrality in Social Networks: Conceptual Clarification." *Social Networks*, 1: 215-239.

Goyal, S. (2007), *Connections: An Introduction to the Economics of Networks*, Princeton University Press.

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

Granovetter, M. (1974), "Getting a Job: A Study of Contacts and Careers." Cambridge, MA: Harvard University Press.

Ichniowski, C., K. Shaw, Prennushi, G. (1997). "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." American Economic Review (87:3): 291-313.

Jackson, M.O., (2006), "The Economics of Social Networks," In *Proceeding of the 9th World Congress of the Econometric Society* (ed. R. Blundell, W. Newey and T. Persson). Cambridge University Press.

Jackson, M.O. (2008), "Social Networks in Economics", forthcoming in the *Handbook of Social Economics* (edited by Benhabib, Bisin and Jackson), Elsevier.

Katz M, Shapiro C. 1985. Network externalities, competition, and compatibility. *American Economic Review* **75**: 424-440.

Leibowitz, Stan J., Stephen E. Margolis. 1994. Network externality: an uncommon tragedy, *Journal of Economic Perspectives*, Vol. 8, No. 2, 133-150.

Rangan, S. (2000), "The Problem of Search and Deliberation in Economic Action: When Social Networks Really Matter." *Academy of Management Review* 25(4): 813-818.

Seeley, C. (2000), "Change Management: A base for knowledge-sharing," *Knowledge Management Review*, 3(4):24-29

Stiroh, K. (2002), "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say**?**," *American Economic Review*, 92: 1559-1576.

Van Alstyne, M. & Zhang, J. (2003) "EmailNet: A system for mining social influence and network topology in communication." North American Association for Computational Social Science (NAACSOS). Pittsburgh, PA.

7 Appendix

	Regression 1: Preferred Model With Human Capital	
Independent Variables	Coefficient	T-stat
CONSTANT	-291,308.6	-0.95
INTERNAL IN-NET	6,505.9	2.83
BETWEENNESS	85.5	1.57
PROJECTS	26,377.8	3.34
PARTNER	236,096.5	2.34
CONSULTANT	334,201.1	4.11
GENDER	-36,608.2	-0.96
AGE	-1,586.3	-0.64
EDUCATION	-1,744.3	-0.11
EXPERIENCE	-1,270.5	-0.44
N of observations	32	
R-squared	0.75	
Adjusted R-squared	0.65	

Table A1: Preferred Regression #1 with Demographic and Human Capital Variables.[30]

[30] We are missing demographic data on one individual. Hence, this regression has 32 observations.

	Revenues	Internal In-Net	Betweenness	Partner	Consultant	Projects
Revenues	1.00					
Internal In-Net	0.43	1.00				
Betweenness	0.43	0.46	1.00			
Partner	-0.20	0.21	0.11	1.00		
Consultant	0.30	-0.36	-0.16	-0.89	1.00	
Projects	0.54	0.28	0.13	0.07	-0.09	1.00

Table A2: Correlation among Variables in first regression in Table 2

	IIN	ION	EIN	EON	IIV	IOV	EIV	EOV	IIM	EIM	IEA	PET
INTERNAL IN-NET (IIN)	1.00											
INTERNAL OUT-NET (ION)	0.75	1.00										
EXTERNAL IN-NET (EIN)	0.02	0.06	1.00									
EXTERNAL OUT-NET (EON)	0.28	0.39	0.54	1.00								
INTERNAL IN-VOL (IIV)	0.69	0.55	-0.09	0.11	1.00							
INTERNAL OUT-VOL (IOV)	0.54	0.72	0.36	0.69	0.50	1.00						
EXTERNAL IN-VOL (EIV)	0.08	0.16	0.93	0.70	-0.04	0.41	1.00					
EXTERNAL OUT-VOL (EOV)	0.40	0.49	0.45	0.89	0.30	0.72	0.59	1.00				
INTERNAL IN-MSG-SIZE (IIM)	0.09	-0.09	-0.09	-0.05	-0.33	-.26	-.01	-.25	1.00			
EXTERNAL IN-MSG-SIZE (EIM)	0.01	-0.07	-0.69	-0.38	0.02	-.28	-.68	-.31	0.17	1.00		
INTERNAL EMAILS WITH ATTACHMENT (IEA)	0.34	0.38	0.65	0.68	0.33	0.64	0.77	0.69	-.15	-.42	1.00	
PEMAIL TIME (PET)	0.32	0.22	0.45	0.65	0.19	0.38	0.57	0.60	0.09	-.41	0.49	1.00

Table A3: Correlation among Alternative Measures of Email use from Table 3

Table A4: Survey Questions Used♣

Q#	Variable	Survey Text
Q01	Age	"I was born in 19__ __"
Q02	Yrs Education	"My total years of education are *(Grammar School 1-8 yrs, High School 12, Some College 13,14,15, College Degree 16, Masters or Professional Degree 18, PhD 20,21+... yrs.)*"
Q03	Industry Experience	"My total number of years of industry experience are ____"
Q04	Industry Expertise	"My areas of specialization or expertise are *(list – please use commas to separate)*:"
Q05	Team Interdependence	"My job tasks are highly inter-dependent with other people's tasks. I must often coordinate with other team members."
Q06	Task Routineness	"My data requirements are highly routine. I could specify all I need on standard forms."
Q07	Data Routineness	"For information that *is* routine, the process of getting it has been automated."
Q08a	F2F Contacts	"How many people do you communicate with on a typical day in the following modes. *[Face-to-Face, Phone, E-Mail, Instant Messaging, Other (please specify)]*."
Q08b	Phone Contacts	"How many people do you communicate with on a typical day in the following modes. *[Face-to-Face, Phone, E-Mail, Instant Messaging, Other (please specify)]*."
Q08c	Email Contacts	"How many people do you communicate with on a typical day in the following modes. *[Face-to-Face, Phone, E-Mail, Instant Messaging, Other (please specify)]*."
Q08d	IM Contacts♠	"How many people do you communicate with on a typical day in the following modes. *[Face-to-Face, Phone, E-Mail, Instant Messaging, Other (please specify)]*."
Q09	Information Sharing	"Colleagues are always willing to share their private search information with me."
Q110	ESS Skill	"I am highly effective at using our in-house proprietary search tools. This means I know what information they contain and I can easily find, add, and modify the records I need."
Q11	Phone Skill	"I am highly effective interacting with people on the phone. This means I am both persuasive in pitching an opportunity and successful in gathering corporate intelligence."
Q12a	Relative time with project team (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12b	Relative time with non-project company colleagues (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12c	Relative time with people outside firm (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12d	Relative time spent on web (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12e	Relative time spent using ESS (Database Use) (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"

Q12f	Relative time spent using outside databases (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12g	Relative time spent searching trade news (%)	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q12h	Relative time spent in misc. sources (%)♠	"What proportion of your time do you spend gathering information from the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13a	Relative value from project team (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13b	Relative value from non-project company colleagues (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13c	Relative value from people outside firm (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13d	Relative value from using web (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13e	Relative value from using ESS (Database Use) (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13f	Relative value from using outside databases (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13g	Relative value from searching trade news (%)	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q13h	Relative value from misc. sources (%)♠	"What proportion of your value do you receive from interacting with the following sources: *[My project team, Company colleagues not on my project, People outside my company, Public access Web pages, Our internal database, External proprietary databases, News or trade press, Other (please specify)]*?"
Q14	Multitasking	"Information technology has increased my ability to handle more projects at the same time."

♣Survey questions reported for this phase of research only.

♠Fewer than four respondents provided data on "IM Contacts" or overlapping "Other" categories, preventing us from drawing statistical inferences from models that used them.