Diffusion, Network Structure & Information Advantage

Marshall Van Alstyne
Researcher Scholar, MIT Sloan School
Professor, Boston University
T: 617-253-0768
E: marshall@mit.edu or mva@bu.edu

* Joint work with Sinan Aral and Erik Brynjolfsson
* Special thanks to Jun Zhang and Nat Bulkley

© 2007 MIT Center for Digital Business. All rights Reserved.
Information Economics Research

Products & Network Effects

Communications Markets (Anti-Spam)

Information & Productivity

IP Law & Policy

[Images of software, cartoon, network diagram, and figure]
Information Diffusion is Fundamental

- Many theories in finance, marketing and innovation rely on assumptions about *how information moves through social groups*.

  - **Innovation Diffusion**: Rogers (1995: 17-18) “the essence of the diffusion process is the information exchange through which an individual communicates a new idea to one or several others.”

  - **Dynamic Trading**: Hirshleifer et. al. (1994) Asymmetric access to information cascades creates abnormal profits and seemingly irrational trading equilibria (“herding” or “follow the leader” strategies).

  - **Viral Marketing**: Word of mouth and influence in networks rely on information spread (Dellarocas 2003, Domingos & Richardson 2001).

  - **Information Worker Productivity**: Information is a primary input for over 70% of the labor force; Networks affect productivity (Aral et. al. 2006).

- But little research on information diffusion itself.
Motivation

- Can you measure results of information work?
- Can you measure information diversity?
- If so, does greater access to diverse information predict results?
- If information predicts results, who gets it first?
Why Information Workers?

Production in Developed Economies is Changing…

U.S. Manufacturing and Information Services Employment:
1985-2005


IT & Information
Arguably Most
Important Here

Information, Services
70% of labor force;
60% of GDP of US

Neo-Classical Production Theory
& Task Level IT Productivity Studies
Focused Here

Manufacturing

Year

Annual Employment (M)
“In the physical sciences, when errors of measurement and other noise are found to be of the same order of magnitude as the phenomena under study, the response is not to try to squeeze more information out of the data by statistical means; it is instead to find techniques for observing the phenomena at a higher level of resolution. The corresponding strategy for economics is obvious: to secure new kinds of data at the micro level.”

—Herb Simon
Agenda

- Study Context
- Tools, Visualization & Social Network Theory
- Network Structure & Information Advantage
The Study: Executive Recruiting

The Data: Unusually Measurable Inputs and Outputs
- A medium sized E.R. firm, 14 locations throughout the U.S.
- 15 semi-structured interviews across org. levels
- ~1300 total projects over 5 years (January 1, 2000 to March 20, 2005)
- A survey of employees’ IT and information seeking practices (87% response rate)
- ~125,000 email messages over 10 months
- Over 1.5 million words!
- Analyses at the Individual Level (Individual Productivity) and Project Level (Task Level Intermediate Outcomes)

Several Measurable Outputs
i. Revenues per person and per project
ii. Number of completed projects
iii. Duration of projects
iv. Number of simultaneous projects
v. Compensation per person
The Setting:
- Executive Search Process

Recruiters access information in two ways:

**Technological**
1. Executive Search System
2. External DBs
3. Internet (e.g. HotJobs, Monster etc)

**Social**
- Phone
- eMail
- F2F

Capture Requirements
Initial Search / Create Initial Pool
Vet Candidates
Create Interview Pool / Interview Internally
Create Final Pool / Facilitate Client Interviews (~ 6)
Tools & Technology

The view from the E-Mail Microscope
To: Marshall Van Alstyne <mvanalst@umich.edu>
Subject: Re: YOUR PROPOSAL
Date: Sun, 17 Nov 2002 09:54:23 -0500
From: Ann < averhey@umich.edu >
Cc: averhey@umich.edu, Geoffrey Parker <gparker@tulane.edu>
X-Originating-IP: 68.41.189.43

Ok, i will look for all the pieces today then and try to get everything in Fastlane tonight.

Meeting is up to you. I have to go to DRDA first thing in the morning to hand them all the PAFs so they can process all the proposals. The meeting is to give you one last chance to view the entire proposal package before DRDA pushes the "Send" button. We could also try to do this virtually so neither of us has to travel to the other site.

As far as footers go, let's not worry about it as long as you are page numbering each section individually. I usually add more information to the footer but I don't have time to worry about this detail.

Ann
Information Loss is Small

This plots the entropy of hashed text relative to that of plaintext using the Enron data.

\[
y = 2.0592x - 6.8973 \\
R^2 = 0.8802
\]
Information Content & Overlap

Example content for a single inbox.

Example overlap for 12 people.
Topology

Comprehending the Social Networks
Who gets news first? Does it Matter?
A 40 year old assumption

- Network structure is associated with productivity and performance.
  - Productivity of information workers (Aral, Brynjolfsson, Van Alstyne 2006)
  - Productivity of R&D teams (Reagans & Zuckerman 2001)
  - Wages, Promotion (Burt 1992), Job Placement (Granovetter 1973, Burt 1992)
  - Innovation (Burt 2004)

- Key theoretical mechanism: access to information.
Information Diversity Depends on Structure

Structural diversity is higher for A

Structurally diverse networks have low “structural equivalence.” A’s contacts are disjoint while B’s contacts are connected to each other and to the same alters.
But Diversity also Depends on Bandwidth

Total Novelty *Flow* Is Higher for B

Although A’s contacts are disjoint, B’s narrower contacts communicate vastly more frequently.
A Need for Skepticism

1. Diversity associated with weak ties => lower bandwidth, frequency, topical dimension and detail, complexity.

2. Awareness enables differentiation. (Coleman 1988)

3. Non-Information Resources from Diverse Networks
   - Power
   - Negotiating Leverage
   - Access to Capital

   - Decision Relevance
   - Disparate Information => Unconnected Ideas
Network Structure & Information Advantage

Who gets informed first & does it matter?!?
Do diverse network structures provide access to diverse information?

Does access to diverse information drive productivity?

Network Diversity $\Rightarrow$ Information Diversity $\Rightarrow$ Productivity
Information Advantage

Value of information comes from its uneven distribution across local network neighborhoods.

Connection to diverse neighborhoods gives access to novel pools of information.

Novel information is valuable due to its local scarcity.

Actors with scarce, novel information can
✓ broker opportunities, engage in information arbitrage or
✓ apply information to problems that are intractable given local information (innovation).
Measuring Information Diversity

Information clusters in an inbox.

Represent each message as a multi-dimensional vector of term frequencies.

In-boxes and Out-boxes represent collections of vectors.

Measure ‘variance’ of the vectors in someone’s in-box or out-box, e.g. cosine distance.

Validate diversity measurement using public data from Wikipedia.org.
Wikipedia Categories

<table>
<thead>
<tr>
<th>Computer science</th>
<th>Geography</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Artificial intelligence</td>
<td>+ Climate</td>
<td>+ Robotics</td>
</tr>
<tr>
<td>+ Machine learning</td>
<td>+ Climate change</td>
<td>+ Robots</td>
</tr>
<tr>
<td>+ Natural language processing</td>
<td>+ History of climate</td>
<td>+ Robotics competitions</td>
</tr>
<tr>
<td>+ Computer vision</td>
<td>+ Climate forcing</td>
<td>+ Engineering</td>
</tr>
<tr>
<td>+ Cryptography</td>
<td>+ Cartography</td>
<td>+ Electrical engineering</td>
</tr>
<tr>
<td>+ Theory of cryptography</td>
<td>+ Maps</td>
<td>+ Bioengineering</td>
</tr>
<tr>
<td>+ Cryptographic algorithms</td>
<td>+ Atlases</td>
<td>+ Chemical engineering</td>
</tr>
<tr>
<td>+ Cryptographic protocols</td>
<td>+ Navigation</td>
<td>+ Video and movie technology</td>
</tr>
<tr>
<td>+ Computer graphics</td>
<td>+ Exploration</td>
<td>+ Display technology</td>
</tr>
<tr>
<td>+ 3D computer graphics</td>
<td>+ Space exploration</td>
<td>+ Video codecs</td>
</tr>
<tr>
<td>+ Image processing</td>
<td>+ Exploration of Australia</td>
<td>+ Digital photography</td>
</tr>
<tr>
<td>+ Graphics cards</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Diversity Measures are Consistent

#### Correlations Between the Five Measures of Information Diversity

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. VarCosSim</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. VarDiceSim</td>
<td>0.9999</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. AvgCommon</td>
<td>0.9855</td>
<td>0.9845</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. AvgCommonIC</td>
<td>0.9943</td>
<td>0.9937</td>
<td>0.9973</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>5. AvgBucDiff</td>
<td>0.9790</td>
<td><strong>0.9778</strong></td>
<td>0.9993</td>
<td>0.9939</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Basic Diversity Models

\[ ID_{it}^I = \gamma_i + \beta_1 E_{it}^I + \beta_2 NS_{it} + \beta_3 NS_{it}^2 + \beta_3 ND_{it} + \beta_4 SE_{it} \]
\[ + \sum_j B_j HC_{ji} + \sum_m B_m Month + \varepsilon_{it} \]

\[ P_{it} = \gamma_i + \beta_1 NRI_{it}^I + \beta_2 ND_{it} + \sum_j B_j HC_{ji} + \sum_m B_m Month + \varepsilon_{it} \]

E-Email volume, NS-Network Size, ND-Network Diversity, SE-Structural Equivalence, HC-Human Capital, NRI-Non Redundant Info

Multiple Specifications and Functional Forms
Network Structure & Information Diversity

Table 3. Network Structure & Access to Diverse Information

<table>
<thead>
<tr>
<th></th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable:</strong></td>
<td>Information Diversity</td>
<td>Information Diversity</td>
</tr>
<tr>
<td><strong>Specification</strong></td>
<td>Fixed Effects</td>
<td>OLS-c</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>Age, Gender, Education, Industry Experience, Partner, Consultant</td>
<td></td>
</tr>
<tr>
<td>Total Email Inbound</td>
<td>-.001 (0.001)</td>
<td>.001 (0.001)</td>
</tr>
<tr>
<td>Network Size</td>
<td>.474*** (.114)</td>
<td>.296* (.138)</td>
</tr>
<tr>
<td>Network Size-Squared</td>
<td>-.272** (.089)</td>
<td>-.240* (.139)</td>
</tr>
<tr>
<td>Network Diversity</td>
<td>.128** (.052)</td>
<td>.268*** (.072)</td>
</tr>
<tr>
<td>Structural Equivalence</td>
<td>-.005 (.033)</td>
<td>.062 (.096)</td>
</tr>
<tr>
<td>Constant</td>
<td>.128* (.075)</td>
<td>.016 (.634)</td>
</tr>
<tr>
<td>Temporal Controls</td>
<td>Month</td>
<td>Month</td>
</tr>
<tr>
<td>F-Value (d.f.)</td>
<td>5.61*** (13)</td>
<td>5.03*** (19)</td>
</tr>
<tr>
<td>R²</td>
<td>.14</td>
<td>.24</td>
</tr>
<tr>
<td>Obs.</td>
<td>540</td>
<td>434</td>
</tr>
</tbody>
</table>

Access to Diverse Information is Positively Correlated With Network Diversity

Relationship between Network Size & Information Diversity is Non-Linear

Marginal Increase in Information Diversity Approaches 0 after ~ 12 contacts.
Diversity Slightly Increasing (But Fairly Flat) By Age, Industry Experience

Human & Demographic capital predict information diversity less well than social capital
### Table 6. Non-Redundant Information and Individual Performance

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>Revenue</td>
<td>Revenue</td>
<td>Revenue</td>
<td>Completed Projects</td>
<td>Completed Projects</td>
<td>Completed Projects</td>
<td>Project Duration</td>
<td>Project Duration</td>
<td>Project Duration</td>
</tr>
<tr>
<td>Fixed Effects Within</td>
<td>Between Estimator</td>
<td>OLS-C</td>
<td>Age, Gender, Edu, Ind Exp., Partner, Consultant</td>
<td>Between Estimator</td>
<td>OLS-C</td>
<td>Age, Gender, Edu, Ind Exp., Partner, Consultant</td>
<td>Between Estimator</td>
<td>OLS-C</td>
<td>Age, Gender, Edu, Ind Exp., Partner, Consultant</td>
</tr>
<tr>
<td>Non-Redundant Information</td>
<td>3835.47*** (1198.08)</td>
<td>4725.47* (2844.25)</td>
<td>8815.06** (2958.88)</td>
<td>.101*** (.023)</td>
<td>.085 (.059)</td>
<td>.192*** (.046)</td>
<td>-16.47** (5.45)</td>
<td>-35.24 (25.58)</td>
<td>-31.56** (15.67)</td>
</tr>
<tr>
<td>Constant</td>
<td>35301.03*** (1397.09)</td>
<td>8071.08 (10839.51)</td>
<td>56717.38** (19898.69)</td>
<td>.668*** (.027)</td>
<td>.139 (.225)</td>
<td>.884** (.418)</td>
<td>284.10*** (6.35)</td>
<td>118.34 (97.50)</td>
<td>-39.28 (196.34)</td>
</tr>
<tr>
<td>Temporal Controls</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
<td>Month / Year</td>
</tr>
<tr>
<td>F-Value (d.f.)</td>
<td>2.40*** (9)</td>
<td>5.30 (7)</td>
<td>2.52** (15)</td>
<td>3.34*** (9)</td>
<td>3.80*** (7)</td>
<td>4.43*** (15)</td>
<td>2.60*** (9)</td>
<td>1.78 (7)</td>
<td>3.09** (15)</td>
</tr>
<tr>
<td>R²</td>
<td>.05</td>
<td>.46</td>
<td>.22</td>
<td>.08</td>
<td>.38</td>
<td>.27</td>
<td>.06</td>
<td>.22</td>
<td>.27</td>
</tr>
<tr>
<td>Obs.</td>
<td>420</td>
<td>420</td>
<td>320</td>
<td>420</td>
<td>420</td>
<td>320</td>
<td>420</td>
<td>420</td>
<td>320</td>
</tr>
</tbody>
</table>
We also find diminishing marginal returns to novel information.

Access to novel information beyond a point is associated with flattening, then declining revenue generation.
Observations

- *Bigger and More Diverse* networks provide access to more diverse information.

- Diverse information is associated with higher productivity, and...

- Diverse networks provide additional benefits.
Diffusion

So who gets news first?

Is it demographics (age, gender), human capital (education, experience), task character (project co-work, same industry), or organizational hierarchy (job level)??
Word characterizations appear standard

Distribution of word frequencies follows a standard Zipf’s law distribution. Distribution of common word usage.
If critical news enters the firm, can network position affect who becomes aware of it? Aware of it sooner?

We categorized words by type by usage statistics:

Network position affects *who* sees information and *when*.
Diffusing Discussion Topic
Hypothesized Several Drivers

Demography

✓ Demographic Dissimilarity Between Originator & Recipient (Age Difference, Gender Difference, Education Difference)
✓ Individual Characteristics (Age, Gender, Education)

Organizational Hierarchy

✓ Managerial Level Difference - Lateral vs. Vertical Diffusion
✓ Managerial Level – If Vertical, is it up or down?

Tie & Network Characteristics

✓ Strength of Tie, Path Length, Betweenness Centrality, Constraint

Task Characteristics

✓ Project Co-Work
✓ Same Work Specialization? (Medical, High Tech, by Region)
✓ Industry Tenure Difference
Two Diffusion Models

Logistic model is better for probabilistic discrete outcomes (receive word or not?)

\[
\ln\left(\frac{p(y_i=1)}{1-p(y_i=1)}\right) = \beta_i + \sum_{j} X_{ij} \beta_j
\]

Hazard rate model is better for conditioning on prior events (receive news given not yet received?)

\[
R(t) = r(t)e^{\beta X}
\]

R(t)-baseline news completion Rate, t-Time, \(\beta\)-vector of coefficients, \(X\)-vector of independent vars.

Multiple Specifications and Functional Forms
See Handout

Is it demographics (age, gender), human capital (education, experience), task character (project co-work, same industry), or organizational hierarchy (job level)??
### Demographic Distances Matter

- Gender Dummy (Male = 1)
- Age Difference
- Gender Difference
- Education Difference

### Project Co-Work Matters

- Geographic Proximity (Same Office = 1)

### Social Ties Matter

- Prior Project Co-Work
- Industry Tenure Difference
- Same Area Specialty

### Information Type Matters

- Communication Volume (Total Email)
- Strength of Tie
- Communication Volume (Total Email)
- Path Length
- Friends in Common
- Betweenness Centrality

### Table 3. Drivers of Access to Information

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable:</td>
<td>Word Received</td>
<td>Rate of Receipt</td>
</tr>
<tr>
<td>Specification (Coefficient Reported)</td>
<td>Logistic (Odds Ratio)</td>
<td>Hazard Model (Hazard Ratio)</td>
</tr>
<tr>
<td>Demography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender Dummy (Male = 1)</td>
<td>1.551 (.219)***</td>
<td>1.236 (.167)***</td>
</tr>
<tr>
<td>Age Difference</td>
<td>.986 (.004)***</td>
<td>.996 (.004)***</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>.869 (.014)***</td>
<td>1.009 (.010)***</td>
</tr>
<tr>
<td>Education Difference</td>
<td>.906 (.023)***</td>
<td>.971 (.020)***</td>
</tr>
<tr>
<td>Geographic Distance</td>
<td>.857 (.088)</td>
<td>.865 (.078)</td>
</tr>
<tr>
<td>Prior Project Co-Work</td>
<td>1.042 (.016)***</td>
<td>1.031 (.012)**</td>
</tr>
<tr>
<td>Task Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry Tenure Difference</td>
<td>.996 (.006)</td>
<td>1.002 (.006)</td>
</tr>
<tr>
<td>Same Area Specialty</td>
<td>.883 (.080)</td>
<td>.983 (.067)</td>
</tr>
<tr>
<td>Organizational Hierarchy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial Level Difference</td>
<td>.951 (.038)</td>
<td>.997 (.033)</td>
</tr>
<tr>
<td>Partner Dummy</td>
<td>.933 (.188)</td>
<td>1.062 (.168)</td>
</tr>
<tr>
<td>Consultant Dummy</td>
<td>.870 (.184)</td>
<td>1.118 (.207)</td>
</tr>
<tr>
<td>Descriptive Network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication Volume</td>
<td>1.0002 (.0002)***</td>
<td>1.000 (.000)</td>
</tr>
<tr>
<td>Strength of Tie</td>
<td>1.002 (.001)***</td>
<td>1.000 (.000)</td>
</tr>
<tr>
<td>Path Length</td>
<td>.711 (.047)***</td>
<td>.828 (.033)***</td>
</tr>
<tr>
<td>Friends in Common</td>
<td>.954 (.007)***</td>
<td>.992 (.005)</td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>1.005 (.002)***</td>
<td>1.004 (.002)***</td>
</tr>
<tr>
<td>Word Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Information</td>
<td>3.209 (.056)***</td>
<td>2.292 (.065)***</td>
</tr>
<tr>
<td>Discussion Topics</td>
<td>.081 (.008)***</td>
<td>.025 (.002)***</td>
</tr>
<tr>
<td>Log Pseudolikelihood</td>
<td>-224204.48</td>
<td>-1694852.4</td>
</tr>
<tr>
<td>Wald χ² (d.f.)</td>
<td>6264.80 (19)***</td>
<td>8878.76 (19)***</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>.28</td>
<td>-</td>
</tr>
<tr>
<td>Observations</td>
<td>543308</td>
<td>462422</td>
</tr>
</tbody>
</table>

Notes: Age, Edu, Industry Tenure not significant. * p < .05; ** p < .01; *** p < .001.
Results by Information Type

- Demographic distances still matter
- Stronger ties aid discussion
- Longer paths hurt news, really hurt discussion
- Project co-work predicts discussion
- Job gaps matter, consultants discuss more

### Table 4. Drivers of Access to Discussion Topics & Event News

<table>
<thead>
<tr>
<th></th>
<th>NEWS Model 1</th>
<th>NEWS Model 2</th>
<th>DISCUSSION Model 3</th>
<th>DISCUSSION Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification (Coefficient)</td>
<td>Logistic (Odds Ratio)</td>
<td>Logistic (Odds Ratio)</td>
<td>Hazard Model (Hazard Ratio)</td>
<td>Hazard Model (Hazard Ratio)</td>
</tr>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (Male=1)</td>
<td>1.544 (.227)**</td>
<td>1.073 (.137)</td>
<td>1.332 (.228)*</td>
<td>1.075 (.162)</td>
</tr>
<tr>
<td><strong>Demographic Distance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Difference</td>
<td>.992 (.004)**</td>
<td>.981 (.007)**</td>
<td>.998 (.004)</td>
<td>.994 (.007)</td>
</tr>
<tr>
<td>Gender Difference</td>
<td>.902 (.017)**</td>
<td>.814 (.069)**</td>
<td>1.007 (.012)</td>
<td>1.092 (.110)</td>
</tr>
<tr>
<td>Education Difference</td>
<td>.925 (.022)**</td>
<td>.832 (.034)**</td>
<td>.966 (.024)</td>
<td>1.013 (.037)</td>
</tr>
<tr>
<td><strong>Tie &amp; Network Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email Volume</td>
<td>1.0001 (.00007)*</td>
<td>1.0001 (.0001)*</td>
<td>1.0001 (.000)</td>
<td>1.0001 (.000)**</td>
</tr>
<tr>
<td>Strength of Tie</td>
<td>1.000 (.000)</td>
<td>1.007 (.001)**</td>
<td>.999 (.000)</td>
<td>1.006 (.001)**</td>
</tr>
<tr>
<td>Path Length</td>
<td>.732 (.041)**</td>
<td>.029 (.005)**</td>
<td>.814 (.044)**</td>
<td>.310 (.045)**</td>
</tr>
<tr>
<td>Geographic Proximity</td>
<td>.883 (.090)</td>
<td>.929 (.106)</td>
<td>.879 (.097)</td>
<td>.993 (.115)</td>
</tr>
<tr>
<td>Friends in Common</td>
<td>.972 (.005)**</td>
<td>.877 (.012)**</td>
<td>.992 (.007)</td>
<td>.969 (.012)**</td>
</tr>
<tr>
<td>Betweenness Centrality</td>
<td>1.004 (.002)</td>
<td>1.007 (.002)</td>
<td>1.006 (.002)**</td>
<td>1.002 (.002)**</td>
</tr>
<tr>
<td>Constraint</td>
<td>.186 (.213)</td>
<td>2.243 (.265)</td>
<td>.282 (.410)</td>
<td>1.664 (1.698)</td>
</tr>
<tr>
<td><strong>Task Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Project Co-Work</td>
<td>1.010 (.014)</td>
<td>1.080 (.0185)**</td>
<td>1.018 (.016)</td>
<td>1.066 (.018)**</td>
</tr>
<tr>
<td>Industry Tenure Difference</td>
<td>.996 (.006)</td>
<td>.978 (.008)**</td>
<td>.999 (.008)</td>
<td>.999 (.008)</td>
</tr>
<tr>
<td>Same Area Specialty</td>
<td>.933 (.073)</td>
<td>1.038 (.139)</td>
<td>.981 (.078)</td>
<td>1.795 (.252)**</td>
</tr>
<tr>
<td>Managerial Level Difference</td>
<td>.963 (.035)</td>
<td>1.138 (.079)*</td>
<td>.992 (.037)</td>
<td>1.097 (.089)</td>
</tr>
<tr>
<td>Partner Dummy</td>
<td>.856 (.186)</td>
<td>1.515 (.271)**</td>
<td>1.084 (.216)</td>
<td>1.411 (.232)**</td>
</tr>
<tr>
<td>Consultant Dummy</td>
<td>.798 (.177)</td>
<td>1.659 (.262)**</td>
<td>1.221 (.289)</td>
<td>1.749 (.288)**</td>
</tr>
</tbody>
</table>
Information Diffusion Results

First estimated access to information of all types:

1. **Gender**: Men 55% more likely to receive information of all types.
2. Men are no more productive than women.
3. **Demographic Dissimilarity** reduces likelihood.
   - Gender Difference: 13%
   - Education Difference: 10% per year difference
   - Age Difference: 1% per year difference
4. **Strength of tie** increases likelihood of receiving information. 10 additional emails $\sim 2\%$ increase.
5. **Path Length**: An additional hop $\sim 29\%$ reduction.
6. **Prior Project Co-Work**: One more prior project in common associated with 4% increased diffusion likelihood.
- Seeing 10 additional words predicts .1% additional project.

- Being 1 rank lower predicts 23% of one project less complete.

- Being 1 rank later predicts 13% of one project less complete.
- Seeing 1 additional word predicts $70 additional revenue.
- Being 1 rank lower predicts $10,000 less revenue.
- Being 1 rank later predicts $5,900 less revenue.

### Table 6. Information Diffusion & Revenues

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Revenues</td>
<td>Total Revenues</td>
<td>Total Revenues</td>
<td>Total Revenues</td>
<td>Total Revenues</td>
<td>Total Revenues</td>
<td>Total Revenues</td>
</tr>
<tr>
<td>Age</td>
<td>1127.36</td>
<td>888.59</td>
<td>720.11</td>
<td>1812.38</td>
<td>1414.45</td>
<td>2846.80</td>
<td>1525.45</td>
</tr>
<tr>
<td>Gender</td>
<td>-65152.48*</td>
<td>-65387.82*</td>
<td>-67740.8*</td>
<td>-70968.47*</td>
<td>-65451.9*</td>
<td>-64268.99</td>
<td>-71504.52*</td>
</tr>
<tr>
<td>Education</td>
<td>(36796.11)</td>
<td>(34113.54)</td>
<td>(34320.92)</td>
<td>(41507.65)</td>
<td>(37780.24)</td>
<td>(41860.22)</td>
<td>(41663.96)</td>
</tr>
<tr>
<td>Industry</td>
<td>-2517.68</td>
<td>-1744.85</td>
<td>-1599.66</td>
<td>-2061.68</td>
<td>-2365.72</td>
<td>-2648.24</td>
<td>-2222.38</td>
</tr>
<tr>
<td>Experience</td>
<td>(2771.90)</td>
<td>(2755.45)</td>
<td>(2766.55)</td>
<td>(2749.58)</td>
<td>(2789.73)</td>
<td>(2630.66)</td>
<td>(2771.76)</td>
</tr>
<tr>
<td>Partner</td>
<td>121600.4</td>
<td>129394.5*</td>
<td>133607.9*</td>
<td>171580.1*</td>
<td>125243.7*</td>
<td>171220.7*</td>
<td>167003.2*</td>
</tr>
<tr>
<td>Consultant</td>
<td>61777.68</td>
<td>72674.13</td>
<td>73515.88</td>
<td>91727.37</td>
<td>64306.19</td>
<td>93837.54</td>
<td>82969.67</td>
</tr>
<tr>
<td>Words Seen</td>
<td>70.52***</td>
<td>-10202.88***</td>
<td>-5931.05***</td>
<td>152.07**</td>
<td>64.93***</td>
<td>321.50***</td>
<td>114.96***</td>
</tr>
<tr>
<td>Mean Rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank &lt; 10%</td>
<td>152.07**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank &lt; 50%</td>
<td>64.93***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words Seen In 1 Week</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words Seen In 1 Month</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>64973.45</td>
<td>765031.8***</td>
<td>915736.5***</td>
<td>195308.1</td>
<td>85886.32</td>
<td>68776.88</td>
<td>166804.6</td>
</tr>
<tr>
<td></td>
<td>(247744.40)</td>
<td>(22344.2)</td>
<td>(231192.6)</td>
<td>(276691.3)</td>
<td>(255321.6)</td>
<td>(290924.2)</td>
<td>(272595.9)</td>
</tr>
<tr>
<td>F-Value (d.f.)</td>
<td>4.46***</td>
<td>5.39***</td>
<td>5.54***</td>
<td>2.64**</td>
<td>3.77***</td>
<td>3.56***</td>
<td>2.83**</td>
</tr>
<tr>
<td>R²</td>
<td>.39</td>
<td>.42</td>
<td>.42</td>
<td>.24</td>
<td>.36</td>
<td>.27</td>
<td>.27</td>
</tr>
<tr>
<td>Obs.</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
</tbody>
</table>
Diffusion of Different Information

Access to Event News
- ✓ Demographic Similarity
- ✓ Structural Network Characteristics

Access to Discussion Topics
- ✓ Demographic Similarity
- ✓ Structural Network Characteristics
- ✓ Tie Characteristics
- ✓ Task Characteristics
- ✓ Organizational Hierarchy

- Distance (network, demographic, task) slows diffusion

- Different types of information diffuse differently.
  - Task Distance Slows Discussion Diffusion but not News
  - Discussion travels vertically up and down the hierarchy.
  - News travels everywhere.
Conclusions

Productivity, measured as $ and completed projects
- Is predicted to improve with (1) novel words seen, (2) higher rank order, (3) faster receipt.
- Each 10 words seen corresponds to 1% of 1 project completed.
- Each 10 words seen corresponds to $700 of revenue.

Not all information diffuses equally
- News (facts?) behaves differently than discussion (procedures?)
- Distance matters: gender, education, age, co-work more than geography
- Getting information later predicts lower output
  ⇒ Consider job rotation & talking to people who aren’t like you.
Future: New Sites & IT Interventions

- Translation Firm
  - IT Interventions & Experiments:
    - Pre/post new database
    - Pre/post new biz processes
  - Also tracking:
    - Room presence
    - Telecom phone calls
    - Minutes with files

- Insurance Claims Processing
  - Natural Experiments:
    - Pre/post new database
    - Pre/post new biz processes

- Call center
  - Star Trek badges!
  - Physical location & FTF contacts
  - Voice patterns, Link frequency, Conversation dominance
van Leeuwenhoek discovered *cells* in the 1670s when he invented high powered microscopes. One consequence was the subsequent development of the germ theory of disease.
Although van Leuwenhoek developed a better microscope, it took Pasteur to prove the germ theory of disease. “Pasteurization,” his process of heating liquids such as milk to kill most bacteria and molds already present within them is still widely used today.
Thank You!

To Learn More About This And Related Research, Please Visit:

Search SSRN for “Alstyne”

http://ssrn.com/author=253298

Who says information work can’t be measured? You are sitting on an information goldmine!