

Unobtrusive Research Methods: E-mail Data (and Other Electronic Communications)

Adam M. Kleinbaum
Tuck School of Business
Dartmouth College
August 1st, 2014

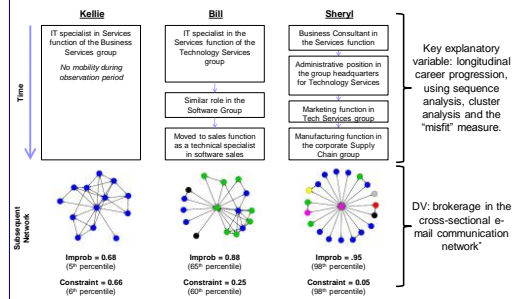
Agenda

- The Data
- What I've Done With The Data
- How (Specifically) I Did It
- Emerging Methods, Future Directions

The Data

- Data from over 30,000 employees at "BigCo," a large, multi-divisional information technology products and services firm
- Network data based on electronic communication archives, including over 250 million dyadic e-mail exchanges, spanning 4Q2006 and 1Q2008
- Demographic and HR data, such as business unit, office, salary grade, gender
 - Some of this was longitudinal

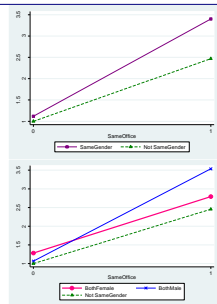
What I've Done With The Data (1)



Kleinbaum, Adam M. 2012. "Organizational Misfits and the Origins of Brokerage in Intrafirm Networks." *Administrative Science Quarterly* 57(3):407-52.

What I've Done With The Data (2)

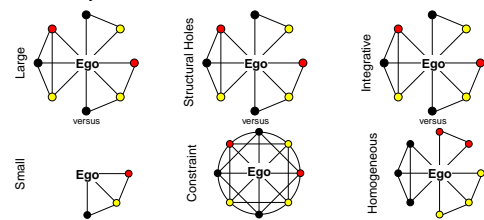
- Formal and informal structures both affect networks, but how do they interact?
- Homophily should be more prevalent where people have more discretion: within, rather than across, formal structural boundaries within the firm.



Kleinbaum, Adam M., Toby E. Stuart, and Michael L. Tushman. 2013. "Discretion Within Constraint: Homophily and Structure in a Formal Organization." *Organization Science* 24(5):1316-36.

What I've Done With The Data (3)

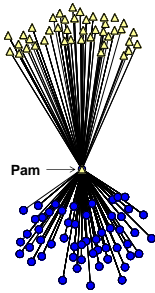
- Staff members should have networks that are: large, sparse and integrative across boundaries.
- But why? Person effects or role effects?



Kleinbaum, Adam M., and Toby E. Stuart. 2014. "Inside the Black Box of the Corporate Staff: Social Networks and the Implementation of Corporate Strategy." *Strategic Management Journal* 35 (1):2-47.

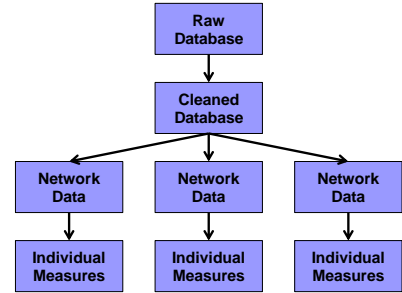
What I've Done With The Data (4)

- The rate at which networks change may be heterogeneous due to:
 - Individual-level: personality, social structure
 - Firm-level: organizational culture, incentives
- This has important strategic implications for firms' dynamic capabilities of coordination and adaptation.



Kleinbaum, Adam M., and Toby E. Stuart. 2014. "Network Responsiveness: The Social Structural Microfoundations of Dynamic Capabilities." *Academy of Management Perspectives* (Forthcoming).

How (Specifically) I Did It



How (Specifically) I Did It: Cleaning The Data

| Step | Command | Output File Name | File Lines | Job Lines | Run Time |
|--------|-------------------------|------------------|------------|-----------|----------|
| Step 1 | cat /tmp/step1_data.txt | | | | |
| Step 2 | cat /tmp/step2_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 3 | cat /tmp/step3_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 4 | cat /tmp/step4_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 5 | cat /tmp/step5_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 6 | cat /tmp/step6_data.txt | output_file_name | 7632358 | 7647474 | 11hour |

How (Specifically) I Did It: Cleaning The Data

| Step | Command | Output File Name | File Lines | Job Lines | Run Time |
|---------|--------------------------|------------------|------------|-----------|----------|
| Step 7 | cat /tmp/step7_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 8 | cat /tmp/step8_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 9 | cat /tmp/step9_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 10 | cat /tmp/step10_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 11 | cat /tmp/step11_data.txt | output_file_name | 7632358 | 7647474 | 11hour |
| Step 12 | cat /tmp/step12_data.txt | output_file_name | 7632358 | 7647474 | 11hour |

How (Specifically) I Did It: Making the Data Usable

- A network from a cross-section of e-mail data:
 - 3-month observation window
 - To and CC recipients included, not BCCs
 - Each sender → recipient pair is counted
 - Mass mails are excluded
 - Ties weighted by message count

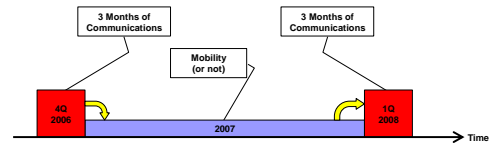
How (Specifically) I Did It: Analysis Tools

- Data cleaning and parsing: Perl, sed, grep
- Data manipulation: C
- Network analysis: igraph package in R
- Econometrics: Stata

Emerging Methods, Future Directions

- Identification
 - Best source of fine-grained longitudinal network data, enabling better causal analysis

Using Propensity Scores to Account for Endogeneity of Mobility with Networks



First Stage: Estimate models of the effect of network structure (and controls) during 4Q06 on the probability of a transition between line and staff during 2007. This estimates the propensity to transition, or "propensity score"

Second Stage: Estimate models of the effect of mobility in 2007 on 1Q08 network structure, weighting observations in the pseudo-population by the inverse of their propensity for treatment (IPTW)

Kleinbaum, Adam M., and Toby E. Stuart. 2014. "Inside the Black Box of the Corporate Staff: Social Networks and the Implementation of Corporate Strategy." *Strategic Management Journal* 35 (1):24-47.

Emerging Methods, Future Directions

- Identification
 - Best source of fine-grained longitudinal network data, enabling better causal analysis
- Data
 - Truly exploiting the real-time longitudinality of e-mail data.
 - E.g., Quintane et al. (2014) "How Do Brokers Broker?"
- Statistical/Econometric
 - Simplicity! There is a trend toward increasingly complex statistical models that readers (writers?) don't fully understand.
 - E.g., exponential random graph models, relational event models.
 - Multi-way clustering with `clus_nway` is a simple, robust method to account for non-independence in dyad models.

Thanks!

(and a shameless plug)

Toward a Network Science of Digital Data in Organizations

Monday, 9:45 - 11:15

Pennsylvania Convention Center, Room 121 B

Featuring perspectives by

Ron Burt, Nosh Contractor, David Lazer and Toby Stuart