Reality Mining

Capturing Detailed Data on Human Networks and Mapping the Organizational Cognitive Infrastructure

Nathan Eagle and Alex Pentland

To unobtrusively glean a detailed map of an organization's cognitive infrastructure

Who is helping whom?

What is the optimum organizational structure?

Who should connect with whom?

Who are the gatekeepers?

Who knows what?

Who influences results?

Which people work well together?

How will communication change after the merger?

Where is the expert?

Features

- **Static**

  Name: Joan N. Peterson

  Job Title: Research Assistant

  Training: Modeling Human Behavior, Organizational Communication, Kitesurfing

- **Dynamic**

  Conversation Keywords: 802.11, wireless, waveform, microphone, cool edit, food trucks, chicken, frequency,

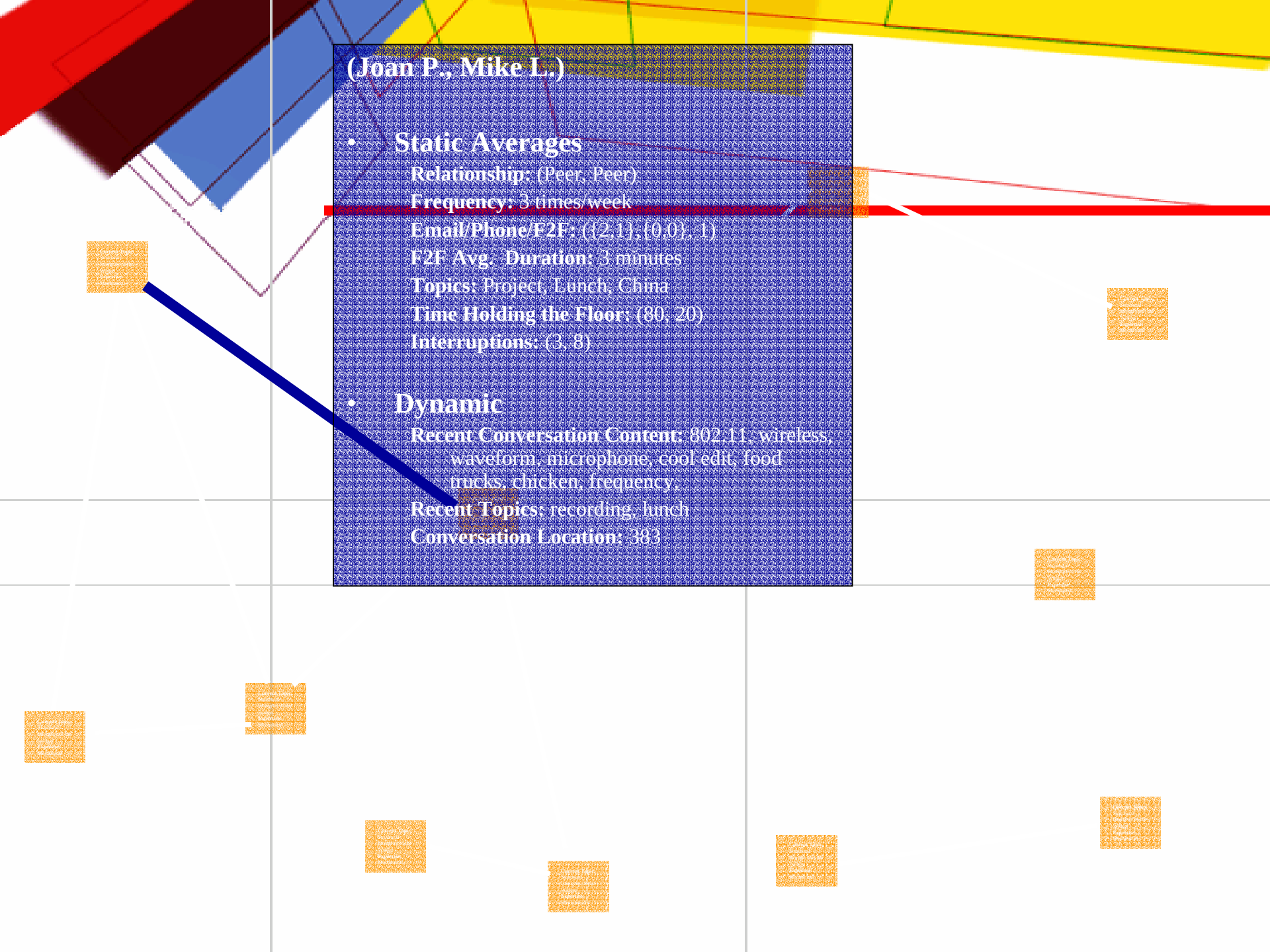
  Topics: recording, lunch

States

Talking: 1

Walking: 0

Activity: ?


(Joan P., Mike L.)
Static Averages
Relationship: (Peer, Peer)
Frequency: 3 times/week
Email/Phone/F2F: ((2,1),(0,0),.1)
F2F Avg. Duration: 3 minutes
Topics: Project, Lunch, China
Time Holding the Floor: (80, 20)
Interruptions: (3, 8)
Dynamic
Recent Conversation Content: 802.11, wireless, waveform, microphone, cool edit, food trucks, chicken, frequency,
Recent Topics: recording, lunch
Conversation Location: 383

Outline

- **The Reality Mining Opportunity**
  - 20th Century vs. 21st Century Organizations
  - Simulations vs. Surveys
  - Reality Mining Overview
- **Mining the Organizational Cognitive Infrastructure**
  - Previous Inference Work
    - Nodes: Knowledge / Context
    - Links: Social Networks / Relationships
  - Details of Proposed Method
- **Applications & Ramifications**
  - SNA, KM, Team formation, Ad Hoc Communication, Simulations
  - Probabilistic Graphical Models
  - …

Physical to Cognitive Infrastructure

20th Century Organization

21st Century Organization

Physical Infrastructure

slowly changing environment –

development of infrastructures to carry out well described processes.

Cognitive Infrastructure

flexibility, adaptation, robustness, speed –

guided and tied together by ideas, by their knowledge of themselves, and by what they do and can accomplish

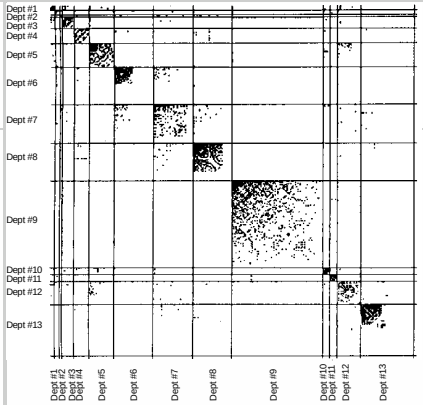
Simulations vs. Surveys

Agent-Based Simulations

Epstein & Axtell, Axelrod, Hines, Hammond, AIDS Simulations

Survey-Based Analysis

Allen, Cummings, Wellman, Faust, Carley, Krackhart



- Lots of synthetic data

- Sparse real data

From Sugarscape: http://www.brook.edu/sugarscape

Allen, T., *Architecture and Communication Among Product Development Engineers*. 1997, Sloan School of Management, MIT: Cambridge, p 33.

Bridging Simulations and Surveys with Sensors

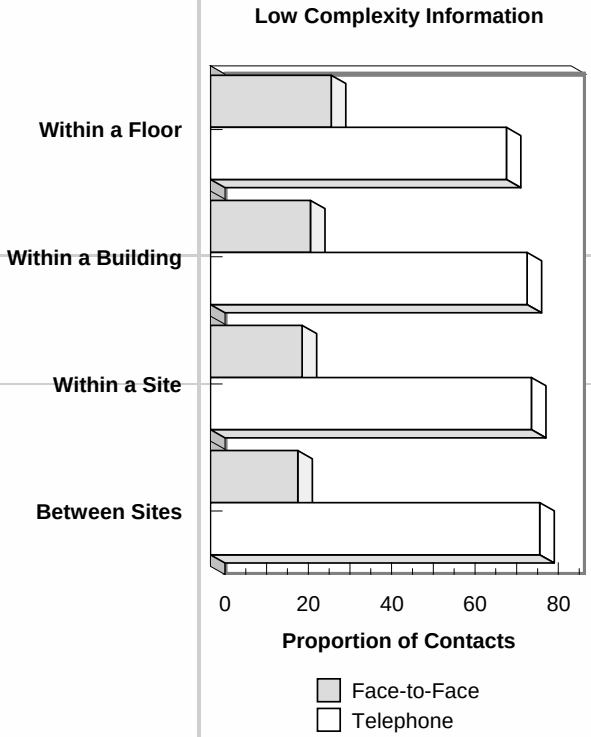
REALITY MINING

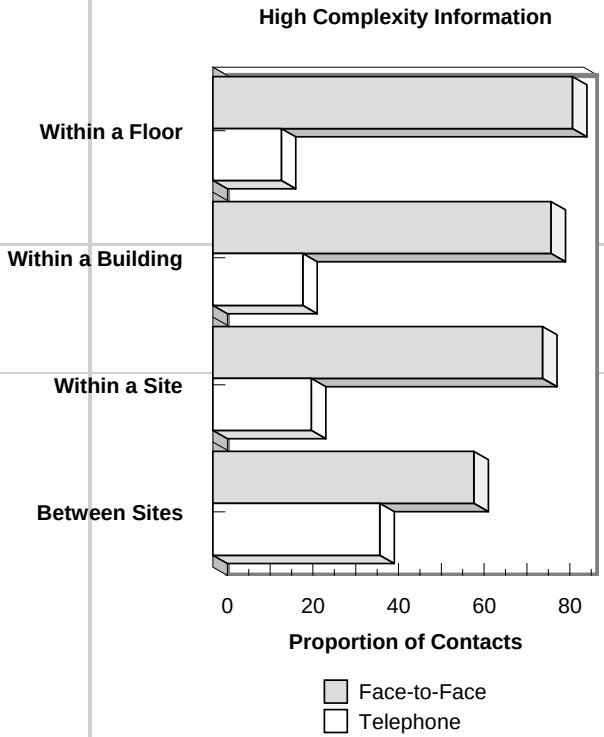
- **Hardware**
  - Linux PDAs (with WLAN)
  - Microphones
- **Data**
  - Audio
  - Local Wireless Network Information
- **Analysis**
  - Situation
    - Type / Recognizing activity patterns
  - Conversation Mining
    - Topic Spotting / Distinctive Keywords / Sentence Types
  - Conversation Characterization
    - who, what, where, when, how
- **Machine Learning**
  - Parameter Estimation, Model Selection, Prediction

(Bluetooth) Microphone / Headset

Sharp Zaurus

Why F2F Networks?





Allen, T., *Architecture and Communication Among Product Development Engineers*. 1997, Sloan School of Management, MIT: Cambridge, p 33.

Outline

- **The Reality Mining Opportunity**
  - 20th Century vs. 21st Century Organizations
  - Simulations vs. Surveys
  - Reality Mining Overview
- **Mining the Organizational Cognitive Infrastructure**
  - Previous Inference Work
    - Nodes: Knowledge / Context
    - Links: Social Networks / Relationships
  - Details of Proposed Method
- **Applications**
  - SNA, KM, Team formation, Ad Hoc Communication, Simulations
  - Probabilistic Graphical Models
  - …

Inference on Individuals : Previous Work

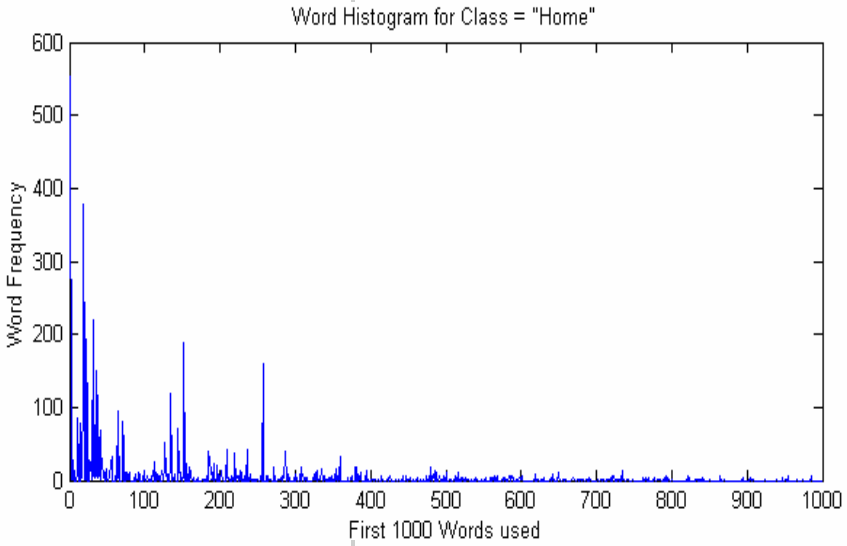
- **Knowledge Inference**
  - Self-Report: Traditional Knowledge Management
  - Email / Intranet: Shock (HP), Tacit, others?
- **Context Inference**
  - Video: iSense (Clarkson 01)
  - Motion: MIThrill Inference Engine (DeVaul 02)
  - Speech: OverHear (Eagle 02)

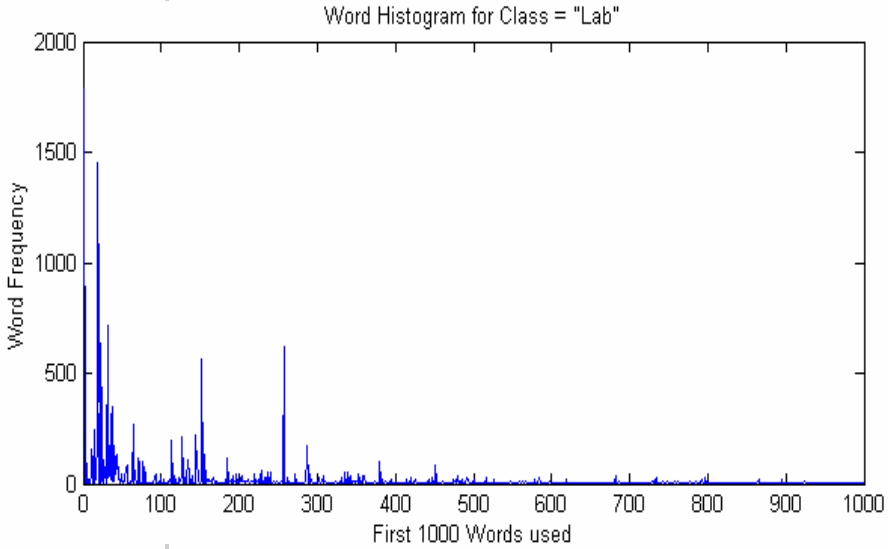
OverHear : Data Collection

- **2 months / 30 hours of labeled conversations**
- **Labels**
  - location
    - home, lab, bar
  - participants
    - roommate, colleague, advisor
  - type/topic
    - argument, meeting, chit-chat

OverHear : Classifier

- **Distinct Signatures for Classes?**



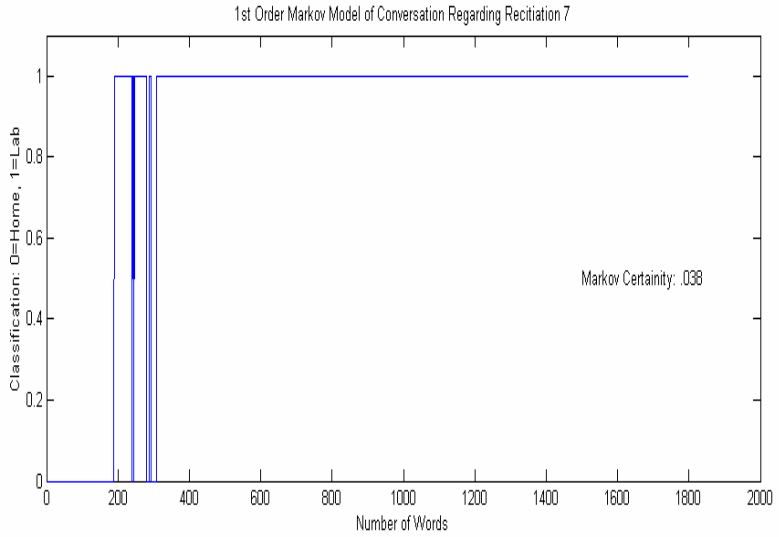


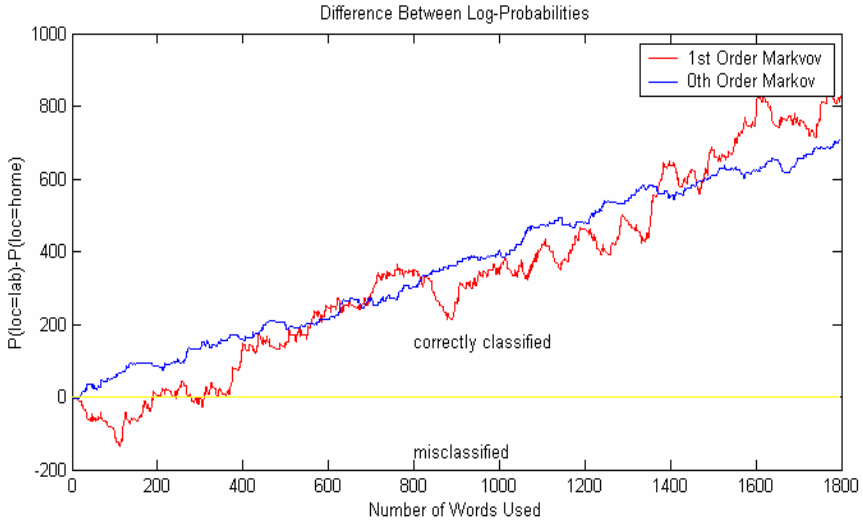
- **Bi-grams : 1st Order Modified Markov Model**

$$\sum_{j=1}^{n-1} \log(count2(stream(j), stream(j+1)) * conf(j)^{q} * conf(j+1)^{q}$$

OverHear : Initial Results

- **Accuracy highly variant on class**
  - 90+% Lab vs. Home (Roommate vs. Officemate)
  - Poor Performance with similar classes
- **Increasing model complexity didn't buy much**
- **Demonstrated some speaker independence**

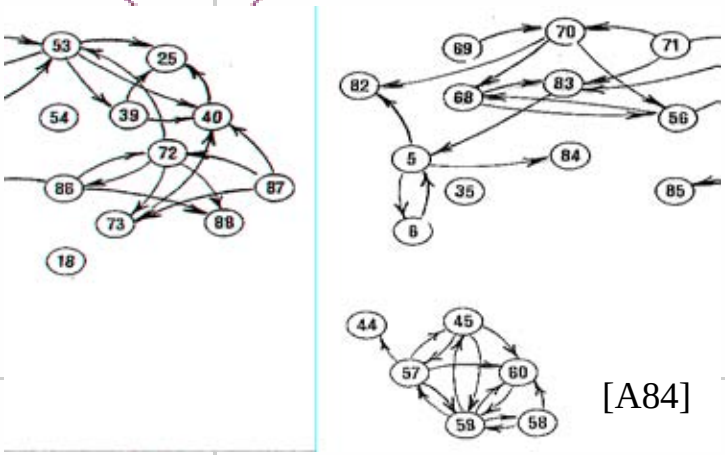




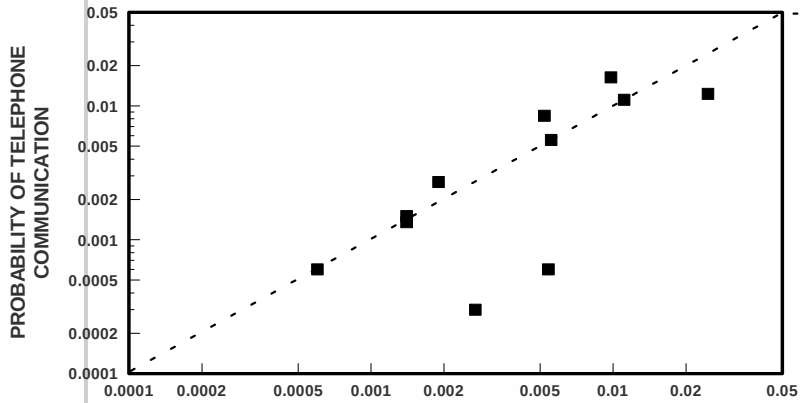
Relationship Inference : Previous Work

- **Relationship Inference / Conversation Analysis**
  - Human Monitoring: (Drew, Heritage, Zimmerman)
  - Speech Features: Conversation Scene Analysis (Basu 02)
- **Social Network Inference**
  - Surveys: Traditional Social Network Analysis
  - IR Sensors: ShortCuts (Choudhury02, Carley99)
  - Affiliation Networks
    - Email Lists, Board of Directions, Journals, Projects
  - Theoretical: Small World / Complex Networks
    - Kleinberg: Local Information
    - Problems within Social Navigation Models

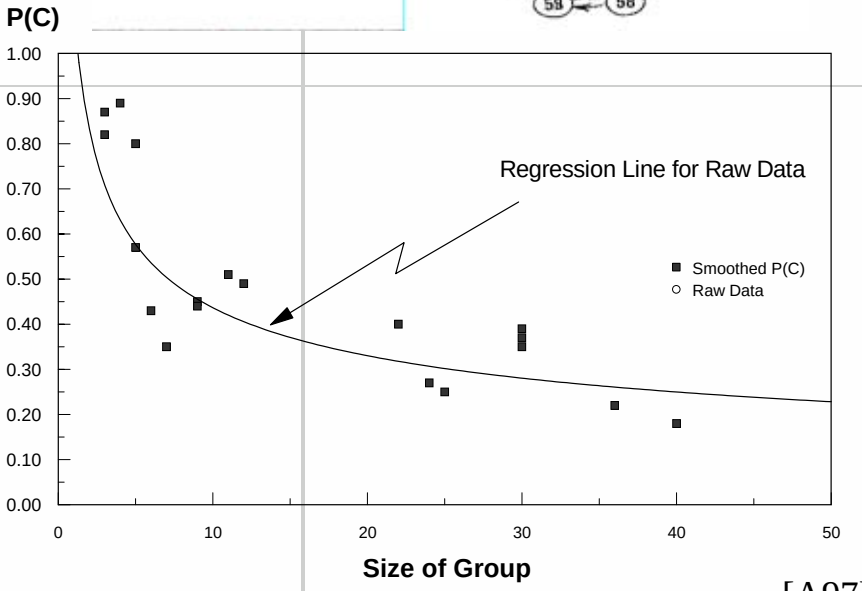
Allen's Studies in the 20th Century



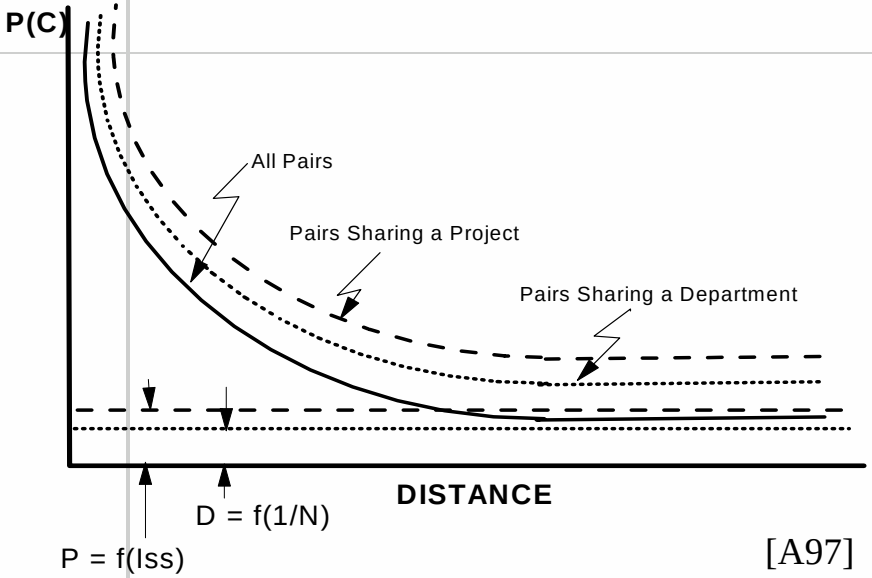
[A84]



[AH87]

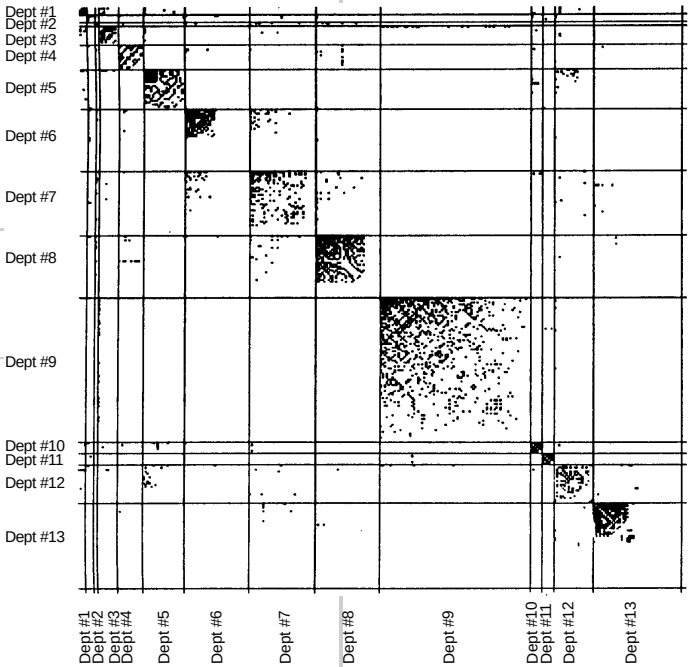


[A97]



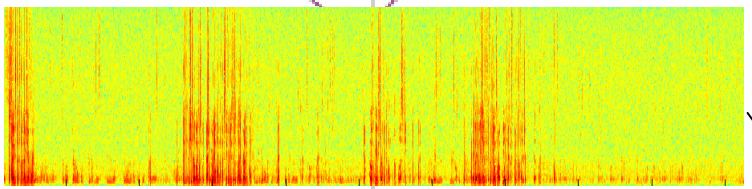
[A97]

Future Organizational Studies?



→ ?

Individuals : Reality Mining



Audio Spectrogram

okay -7 0 second -9 609 aesthetician -13 948 time -3 1417 ASEAN 6 9910 networks
9 9439 to -1 11075 some -7 11734 of 9 11963 Brown's -10 14527 speech -4 14966
that -1 15226 yesterday's -14 17530 and -2 18319 run -15 22659 a -4 22919 helmet
-13 22958 that -4 23357 is -4 23597 the -5 23806 networks -5 23896 and -2 24415
markup -9 24505 models 3 25043 about -19 25442 homeless -10 26969 the -7 27817
change -15 27947 will 0 28396 when 0 28855 Ms. -14 29304 C -6 29912 E -7 30202 C
1 30521 You -10 30760 know -3 30910 what -9 31090 to -5 31319 the -9 31419
Business -16 31549 Network -14 31878 for 14 32227 one -11 37595 day -5 37764

Computer Transcription

(HASABILITY "microphone" "record sound")
(HASREQUIREMENT "record something" "have microphone")
(HASUSE "microphone" "amplify voice")

Common Sense Topic Spotting

wlan0 IEEE 802.11-DS ESSID:"media lab 802.11" Nickname:"zaurus"
Mode:Managed Frequency:2.437GHz Access Point: 00:60:1D:1D:21:7E
Link Quality:42/92 Signal level:-62 dBm Noise level:-78 dBm

Wireless Network Information

Features

- **Static**
  - **Name:** Nathan N. Eagle
  - **Office Location:** 384c
  - **Job Title:** Research Assistant
  - **Expertise:** Modeling Human Behavior, Organizational Communication, Kitesurfing
- **Dynamic**
  - **Conversation Content:** 802.11, wireless, waveform, microphone, cool edit, food trucks, chicken, frequency,
  - **Topics:** recording, lunch
  - **Current Location:** 383

States

- **Talking:** 1
- **Walking:** 0
- **Emotion:** ?

Social Network Mapping

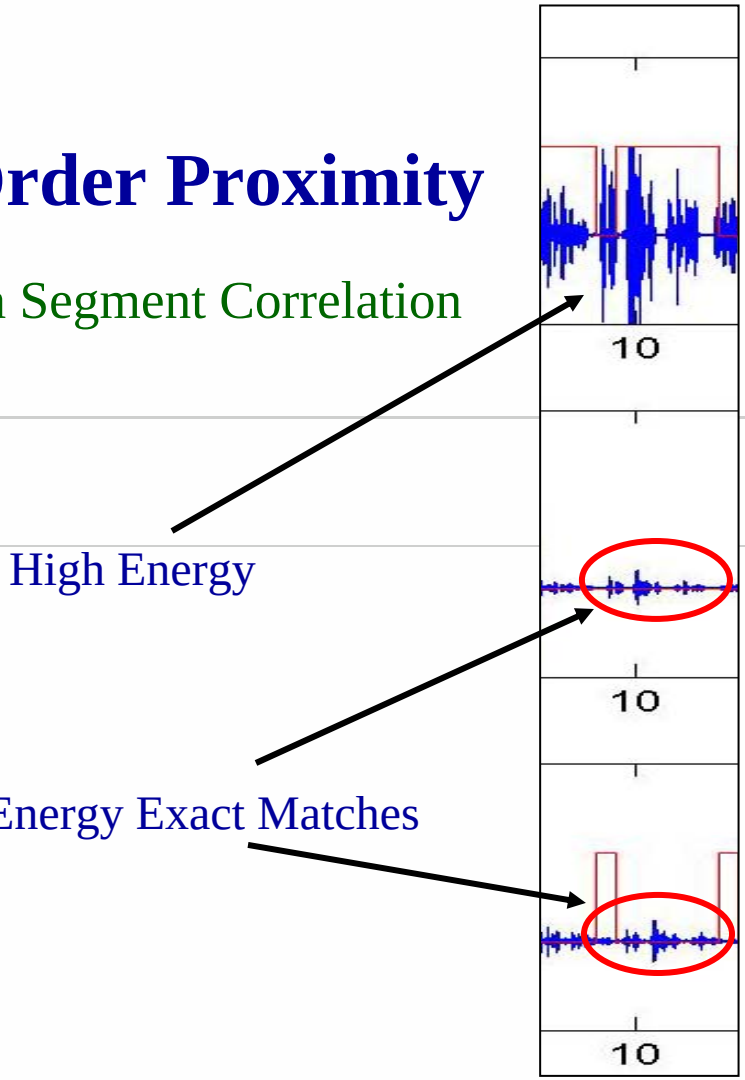
First-Order Proximity

- 802.11b Access Point Check

```
Status
  Found IP 159.139.90.1 for <no ssid>::00:04:76:BB:A7:04 v
  Found IP 159.139.90.1 for <no ssid>::00:04:76:BB:A7:04 v
  Found IP 159.139.90.1 for <no ssid>::00:04:76:BB:A7:04 v
  Found IP 159.139.120.13 for <no ssid>::00:B0:D0:DE:60:E3
Battery: AC charging 100% 0h0m0s
```

Second Order Proximity

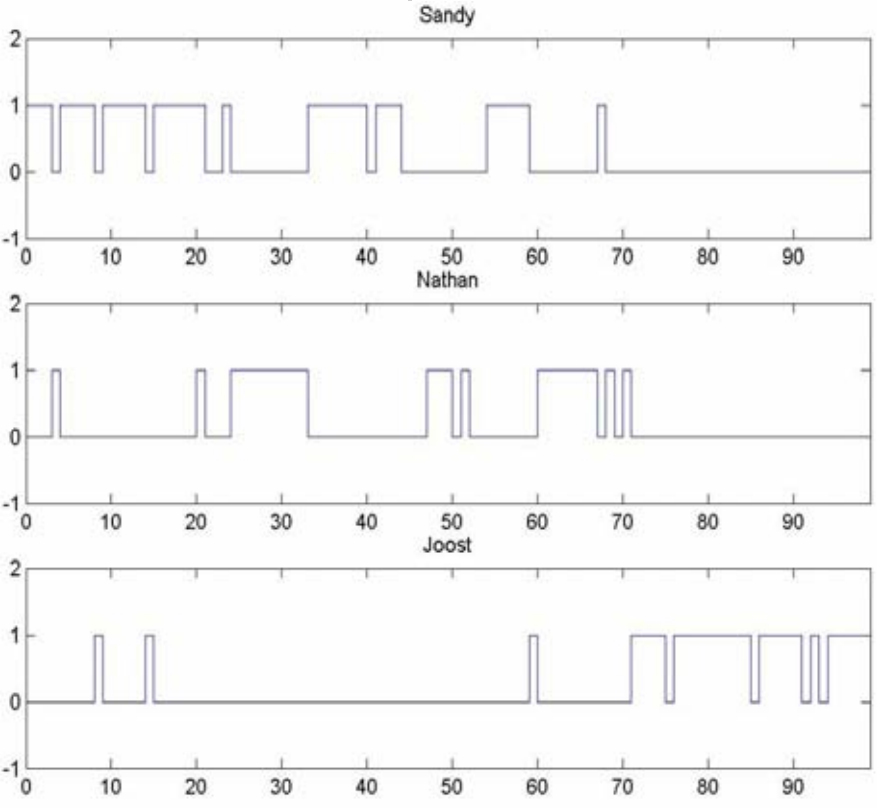
- Waveform Segment Correlation



Social Network Mapping

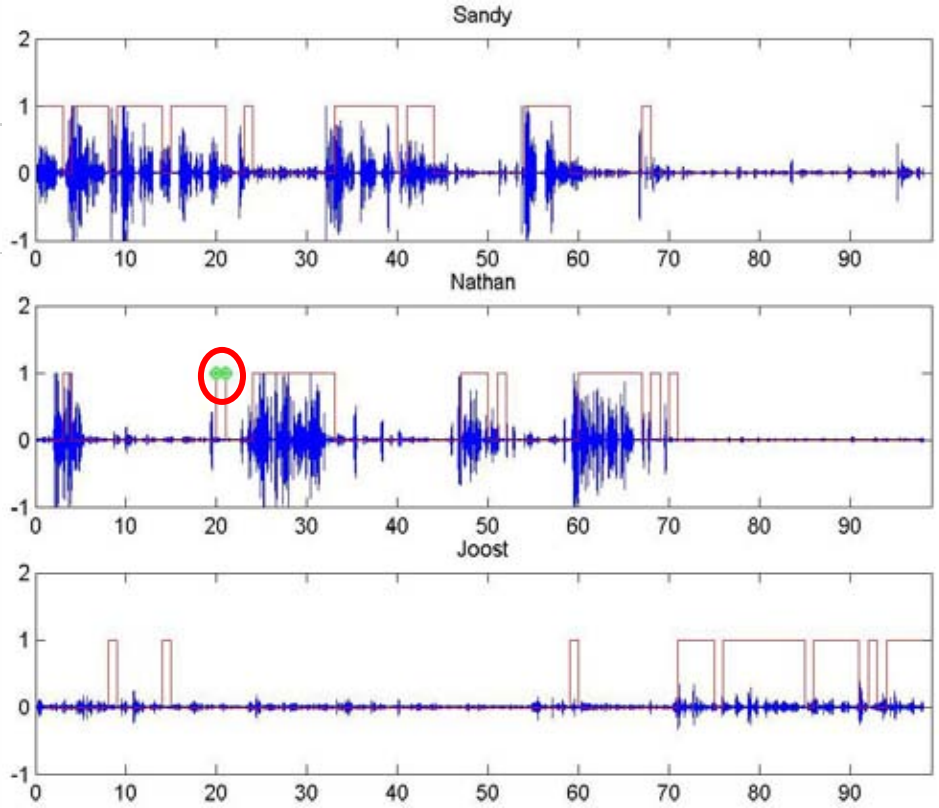
Pairwise Conversation

- Mutual Information [B02]



Interruption Detection

- Non-Correlation + Speaker Transition



Sample Data

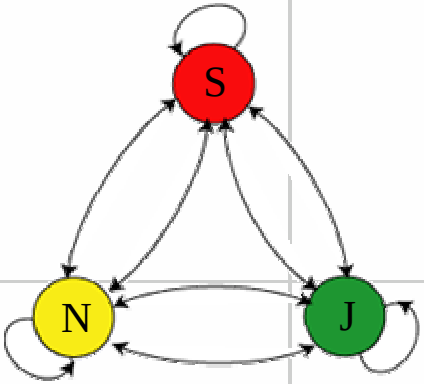
Group

	Relationship	Speaking	Transitions	Interruptions	Duration	Email/F2F	Freq	Proximity	Group Topics
Joost	(∅ , peer, prof)	20%	(.1, .4, .5)	(∅, 2, 0)	15 min	(.1, .9)	1/wk	5%	class, digital, PDA, Zaurus, approval, serial numbers, students, speakers
Nathan	(peer, ∅ , advisor)	27%	(.4, .1, .5)	(2, ∅, 1)					
Sandy	(grad, advisee, ∅)	53%	(.1, .3, .6)	(2, 1, ∅)					

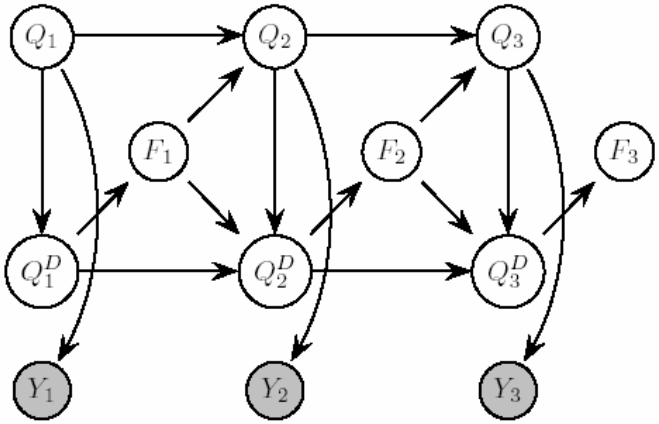
Pairwise

	Relationship	Speaking	Transitions	Interruptions	Duration	Email/F2F	Freq	Proximity	Group Topics
Joost	(∅ , peer)	65%	(.7, .3)	(∅, 6)	30 min	(.6, .4)	3/wk	25%	capital, entrepreneurship, management
Nathan	(peer, ∅)	35%	(.6, .4)	(2, ∅)					

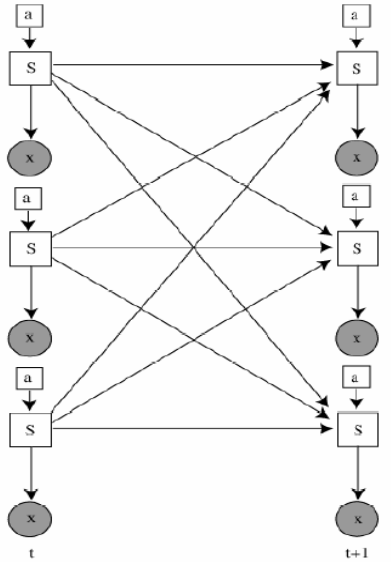
Networks Models



Conversation Finite State Machine



Variable-duration (semi-Markov) HMMs [Mu02]



The Influence Model with Hidden States [BCC01]

Initial Study : Project-based Class

- 10-15 MIT graduate students
- 2-3 hours/week, diverse team projects
- Email and F2F interactions recorded
- Interactions captured over three months

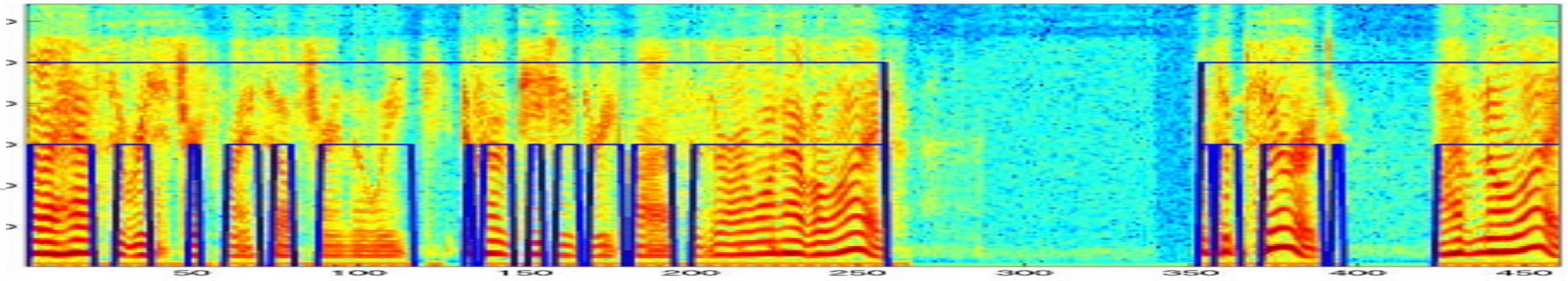
Outline

- **The Reality Mining Opportunity**
  - 20th Century vs. 21st Century Organizations
  - Simulations vs. Surveys
  - Reality Mining Overview
- **Mining the Organizational Cognitive Infrastructure**
  - Previous Inference Work
    - Nodes: Knowledge / Context
    - Links: Social Networks / Relationships
  - Details of Proposed Method
- **Applications**
  - SNA, KM, Team formation, Ad Hoc Communication, Simulations
  - Probabilistic Graphical Models
  - …

Reality Mining : The Applications

- **Knowledge Management**
  - Expertise Finder
  - High-Potential Collaborations
- **Social Network Analysis**
  - Additional tiers of networks based on content and context
  - Gatekeeper Discovery / *Real* Org Chart
- **Team Formation**
  - Social Behavior Profiles
- **Architectural Analysis**
  - Real-time Communication Effects
- **Organizational Modeling**
  - Org Chart Prototyping – global behavior
  - Discovery of unique sensitivities and influences
- ….

Social Network Analysis



Knowledge Management

okay -7 0 second -9 609 aesthetician -13 948 time -3 1417 ASEAN 6 9910 networks 9 9439 to -1 11075 some -7 11734 of 9 11963 Brown's -10 14527 speech -4 14966 that -1 15226 yesterday's -14 17530 and -2 18319 run -15 22659 a -6 22919 helmet -13 22958 that -4 23357 is -6 23597 the -5 23806 networks -5 23896 and -2 24415 markup -9 24505 models 3 25043 about -19 25442 homeless -10 26969 the -7 27817 change -15 27947 will 0 28306 when 0 28955 Ms. -14 29304 C -6 29912 E -7 30202 C 1 30521 You -10 30760 know -3 30910 what -9 31090 to -5 31319 the -9 31419 Business -16 31549 Network -14 31878 for 14 32227 one -11 37595 day -5 37764 you'll -4 37924 get -4 38114 this 3 38393 statistician -10 38642 and 2 39391 yet -14 46355 it -9 46614 is -6 46953 clear 2 47143 how -10 47532 such -8 56612 a -6 56771 question 1 56901 for -5 57300 all -6 60014 on 4 60114 our 15 62947 get -14 63815 out -6 64085 get -11 64404 this 5 64713 even -6 75799 Jean -6 76596 and 1 84049 do -9 84678 what -7 84977 did -5 85107 it -3 85296 in -4 87041 fact -6 87960 hit -12 86092 me -15 96311 yet -1 96471 and -6 96740 if -10 98546 if 2 98656 the 0 99005 NC -6 99093 be 7 101490 the 15 107875 way -13 108813 the -2 109132 state -9 109362 has -1 109601 sustained -5 109920 agency -16 110559 also 15 111049 know -5 115917 the -6 116396 date -7 116495 the 2 116905 events -11 122043 via -16 122532 Van -17 124229 de 3 124517 and -1 125425 a -3 125565 vote 1 125625 the -9 130104 and -8 130723 it -4 130863 is -4 130953 the -4 131142 death -5 132499 the 9 133627 and 2 136380 mean -6 138326 that -2 142047 in -4 142187 more -7 142367 of -8 142556 a 0 145969 Nobel -11 146018 be 9 149660 even -4 149989 their -11 152484 baby -7 152663 the -13 156944 then -15 157013 the -2 157323 bomb -4 157492 it -12 158590 to -6 158749 be 13 159039 a -2 159657 HA 0 162980 finale -16 165043 ya -13 166312 via -15 168956 a -7 169445 book -5 169724 that -6 170034 the -7 170124 steps -13 170453 in -6 171061 the -8 171161 new -10 172099 war -6 172299 they've -13 172628 been -3 173157 well -16 180979 be 0 181139 this -4 181328 leave -21 181697 a -15 181917 few -6 182246 Quebec -19 183304 thought -13 183962 a -10 184231 bit -7 184261 G -5 187534 flung -12 190278 their 0 190647 only -10 192114 visible -16 192433 and 2 197352 they -2 197521 a -12 197721 bad -9 197871 flu -15 198170 while -7 198539 said -5 210472 he -9 210901 would -8 210991 take 1 211150 the -7 211380 ancestors -4 211510 of 3 212109 all 0 212249 three -1 212497 get -12 213405 up -4 213695 in -1 213924 a -3

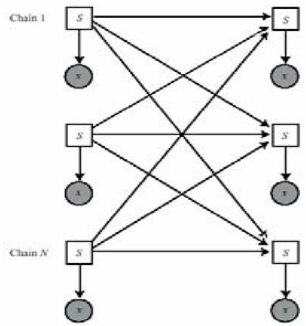
=

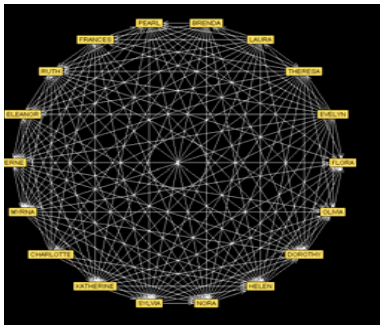
Collaboration & Expertise

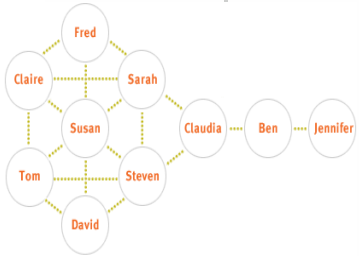
- **Querying the Network**
  - Nodes with keywords&questions
  - Directed Graph = Web Search
- **Clustering Nodes**
  - Based on local links and profile
- **Team Formation**
  - Social Behavior Profiles
- **Ad hoc Communication**
  - Conversation Patching

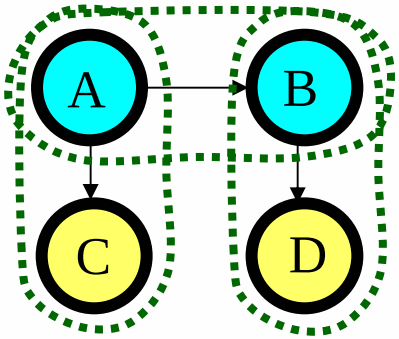
Organizational Modeling

- **Organizational Disruption Simulation**
- **Understanding Global Sensitivities in the Organization**
- **Org-Chart Prototyping**









Privacy Concerns

- **Weekly Conversation Postings**
  - Topic Spotting, Duration Participants
  - User selects Public / Private
- **10 Minute Delete / Mute Button**
- **Low Energy Filtering**
- **Demanding Environments**
  - Fabs, Emergency Response

Anticipations for Reality Mining

- **Positive**
  - Recognition of key players, gate keepers,
  - Recognition of isolated cliques, people,
  - Group dynamics quantified
- **Negative**
  - Big Brother Applications
  - Seeing the data as ground truth
- **Bottom Line**
  - This is going to happen whether we like it or not, anticipating the repercussions needs to be thought about now, rather than later.

Conclusions

- **There is an opportunity to deploy sociometric applications on the growing infrastructure of PDAs and mobile phones within the workplace**
- **Details from this data can provide extensive information of an organization's cognitive infrastructure.**

[BCC01] Sumit Basu, Tanzeem Choudhury, Brian Clarkson and Alex Pentland. *Learning Human Interactions with the Influence Model*. MIT Media Lab Vision and Modeling TR#539, June 2001.

[Mu02] Murphy, K. *Modeling Sequential Data using Graphical Models.* Working Paper, MIT AI Lab, 2002

[AH87] Allen, T.J. and O. Hauptman. *The Influence of Communication Technologies on Organization Structure: A Conceptual Model for Future Research*. Communication Research 14, 5, 1987, 575-587.

[A97] Allen, T., *Architecture and Communication Among Product Development Engineers*. Sloan School of Management, MIT: Cambridge, 1997, p 33.

[A84] Allen, T.J., 1984 (1st edition in 1977), *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*, MIT Press, Mass.