Collecting Network Data in Organizations

Agenda

- Data sources

- Getting access

- Ethics

- Questionnaire design choices

- Ego network designs

Data Sources

- Surveys
- Project billing data
- E-mail traffic
- Listservs
- Memberships
  - Distribution lists, task forces, interest groups etc.
- Direct observation
  - Location-based vs. time allocation

Getting Access

- Wide variation in acceptance

- Easiest in tech companies

- Usually need to offer quid pro quo

- Persistence pays

Ethical Challenges

- Lack of anonymity & confidentiality
  - Especially if management wants to see data
- Non-response does not equal non-participation
- Deceptive power of network analysis
  - Resps don't see the risks
  - Data maybe used against them
- Managers in love with networks fail to see the larger picture
  - Fire employees based on network position
- Practicing therapy without a license

Protecting Respondents

- Major tools
  - Management Disclosure Contract (MDC)
  - Truly Informed Consent form (TIC form)
- Additional tools
  - Anonymization and aggregation
  - Un-coerced participation
  - Personalized respondent give-back
  - Pre-survey consent
  - True opt-out option

Questionnaire Design Choices

- Paper or plastic?

- Roster vs. open-end

- Tick vs. rate

- Serial vs. parallel (grids)

Media

- With paper, you can call meeting of the department and fill out questionnaires on the spot

- Response rates lower for web surveys?

- Fancy web programming can simplify questionnaires

Roster vs. Open-End

- Roster Advantages
  - produce cleaner data – less decoding of what respondent intended
    - Is this "bob" the same as that "bob"?
  - lack of tie is less ambiguous
  - Doesn't limit number of choices a person can make
- Roster Disadvantages
  - Need clear sense of research boundary
  - Cumbersome with large datasets
- Web surveys can provide open-end with look-up

Tick vs. Rate

- Tick advantages
  - *Much* faster than ratings
  - Produces more believable data

- Tick disadvantages
  - Doesn't work with truly continuous variables
    - But some ordinal scales can be made into multiple tick questions

Recasting Ordinals as Ticks

- Rate:
  - How often do you see this person?
    - 5=every day; 4=every week; 3=every month; 2=every year; 1=never

- Multiple Ticks:
  - Which of the following people do you see at least once a year?
  - Which do you see at least once a month?
  - Which do you see at least once a week?
  - Which do you see at least once a day?

Serial vs. Parallel

- On paper, is efficient to ask questions on one page and have responses entered in grid on next page – all questions on one grid

- On web, one can ask each question one at a time, presenting new roster each time

Social Network Questionnaire

Thanks for participating. Please note that the data generated in this survey are NOT anonymous and are NOT confidential. The results will be used in the workshop in Washington. **Important note: you <u>must</u> enter your name in Question 0.**

When you're done, press the "Submit" button. Thanks for your help.

Q0. What is **your name**: [   ]

Q1. Using the checkboxes below, please indicate **who you have heard of or know about** among the participants of the workshop.

Q2. Check off the names of the **people you know**. By "know" I mean that you can attach a name to a face, you have spoken to each other at least once, and the other person is also likely to put you down.

Q3. Check off the names of people you **have worked with** on a paper or other academic/administrative project.

Q4. Check off the the names of a selected set of people whom you don't know but **would like to know**, based on things you've heard, or their interests, etc.

Name	Q1. Heard of them	Q2. Know them	Q3. Worked with	Q4. Want to know
Allata, Joan	☐	☐	☐	☐
Baer, Justin	☐	☐	☐	☐
Baker, Ted	☐	☐	☐	☐
Bercuwitz, Rick	☐	☐	☐	☐
Branzei, Oana	☐	☐	☐	☐
Brooks, Scott	☐	☐	☐	☐
Brower, Ralph	☐	☐	☐	☐

Separate consent form is better

<u>Not</u> anonymous

- Design elements
  - aided
  - grid method
  - web medium
  - 1/0 check-offs

Fill-in-the-Blank

1. If you wanted to get something improved or done on behalf of a customer who would you contact? (write as many names as you like in the spaces provided)

   _____ _____
   _____ _____
   _____ _____
   _____ _____

2. If you wanted to get a true reading on where [company name] was headed as an organiza-tion, who would you talk to?

   _____ _____
   _____ _____
   _____ _____
   _____

- Design elements
  - unaided
  - one relation at a time
  - paper & pencil
  - 1/0 data

Hybrid Questionnaire

1. If you wanted to get something improved or done on behalf of a customer who would you contact?

Name   (index no.)

_____Denny Terio_____ (**169**)

_____Eric Estrada_____ (**27**)

_____ (   )

_____ (   )

2. If you wanted to get a true reading on where [company name] was headed as an organization, who would you talk to?

_____ (   )

_____ (   )

- Design elements
  - hybrid aided/unaided
  - one relation at a time
  - paper & pencil (w/ lookup form)
  - 1/0 data

Web version uses drop-down menus

Unexpected Asymmetry

- A claims to have sex with B, but B does not claim to have sex with A
  - The relation is logically symmetric, but empirically asymmetric
  - errors of recall; strategic response
- Sometimes asymmetry is the point
- Logically symmetric data may be symmetrized
  - Low standard: if either A or B mentions the other, it's a tie
  - High standard: it's only a tie if both A and B mention the other

Non-Symmetric Relations

- Gives advice to
- Can't symmetrize logically non-symmetric relations, except by changing meaning of tie
- Unless you ask question both ways:
  - Who do you give advice to?
  - Who gives advice to you?
- Two estimates of the A$\rightarrow$B tie, and two estimates of the A$\leftarrow$B tie

Missing Data

- For symmetric relations
  - if X_{ij} is missing, substitute X_{ji}
  - If whole row missing, substitute corresponding column
- For non-symmetric relations collected using row & column based questionnaire,
  - set $A_{ij} = B_{ji}$
  - i.e., missing row is replaced with column of the inverse relation

CSS Method

- Each respondent asked about relations among all pairs of persons in group, not just those involving self

  - Yields network matrix $C(k)$ for each respondent

- Aggregate respondent matrices using choice of rules

  - Local: $X_{ij} = 1$ if $C(i)_{ij}$ and $C(j)_{ij}$
  - Global: $X_{ij} = 1$ if $C(k)_{ij} = 1$ for most k

Krackhardt CSS

Q1. How well the members of each pair know each other:									
Knowledge	Aaron	Ali	Dan	Dave	David	Ed	George	Greg	Howard
Aaron									
Ali									
Dab									
Dave									
David									
Ed									
George									
Greg									
Howard									

Dillman Design Considerations

- Network questionnaires can be fun but are usually time-consuming and get generate anxiety

- Providing value

- Treating resp with respect

- Attractive formatting

- Cloaked in authority and importance

Explaining Questions

- "Friendship" does not mean the same thing to everyone
  - Especially across national cultures
- Mitigating practices
  - Use one word label plus two or three sentence description, plus have full paragraph detailed explanation available
  - Homogeneous samples

Sampling

- Local measures are not a problem
  - Ego-network stuff
- Global network measures like regular equivalence or eigenvector centrality are more of a problem
  - Robustness currently being studied
- Statistical corrections for snowball sampling now being studied

Bounding

- Extremely vexing to beginners and outsiders
  - Network concept would seem to argue against boundaries
  - Empirical research makes clear we are all connected
  - Even if distant links don't matter, some people in the sample will be at the edge

- One key is to isolate when bounding matters
  - Yes: Interpersonal influence studies
  - No: homophily studies

Types of Boundaries

- Realist (emic) vs nominalist (etic)
- Attribute-based
  - Top management team at Enron
  - Drug injectors in Hartford
- Relation-based
  - Snowballing out from seed sample until few or no new names (i.e., exhaust component)
  - But is component a real boundary?
- Mixed criteria
  - Sexual ties among residents of Nang Rong

Which relations to measure?

- IT DEPENDS!!!
  - A relation is just a variable. "giving advice" is to network analysis what "attitude toward gun-control" is to survey research.
    - In survey research, do you ask what questions you should ask??
- What's relevant for the phenomena in question?
  - HIV diffusion: sexual ties and needle-sharing are directly involved
    - Other ties like acquaintanceship can potentially turn into sex and sharing ties
- It is the researcher who defines the relations of interest
  - But measuring emically non-salient relations can be challenging
    - Check off the people who send Christmas cards to your friends
    - Who are the people whose bodies are similar to your own?
- Which questions tap "the" social network of the group?
  - Looking for validated "social network scale"

Response scales

- Some respondents positively biased
  - Give big numbers in general when rating strength of tie or frequency

- Row-based approach yields matrices in which each row potentially has different measurement scale
  - Can create asymmetry when none "exists"

- For valued data can normalize by rows
  - Z-scores, euclidean norms, maximum, marginals

Informant Accuracy

- Bernard, Killworth et al compared observed with recalled interaction data
  - Ham radios, deaf TTYs
  - About half of the cells in the adjacency matrix were wrong

- Romney & Faust noted that structural analyses didn't seem so far off

- Freeman, Romney & Freeman

Krackhardt CSS

- Many sources of inaccuracy
  - Recall and exaggeration of ties with high status people
  - Idiosyncratic understanding of the question
- Take "average" of everyone's perception of given dyad's relationship
  - Capitalize on social cognition (see Dawes)
 - Great for deliberately hidden relationships

Ego Networks

- Focal individual
  - The respondent
  - Called EGO
- Set of people they are connected to
  - Called Alters
- Optional: ties among the alters
  - Collected from ego
- Optional: attributes of the alters
  - Collected from ego
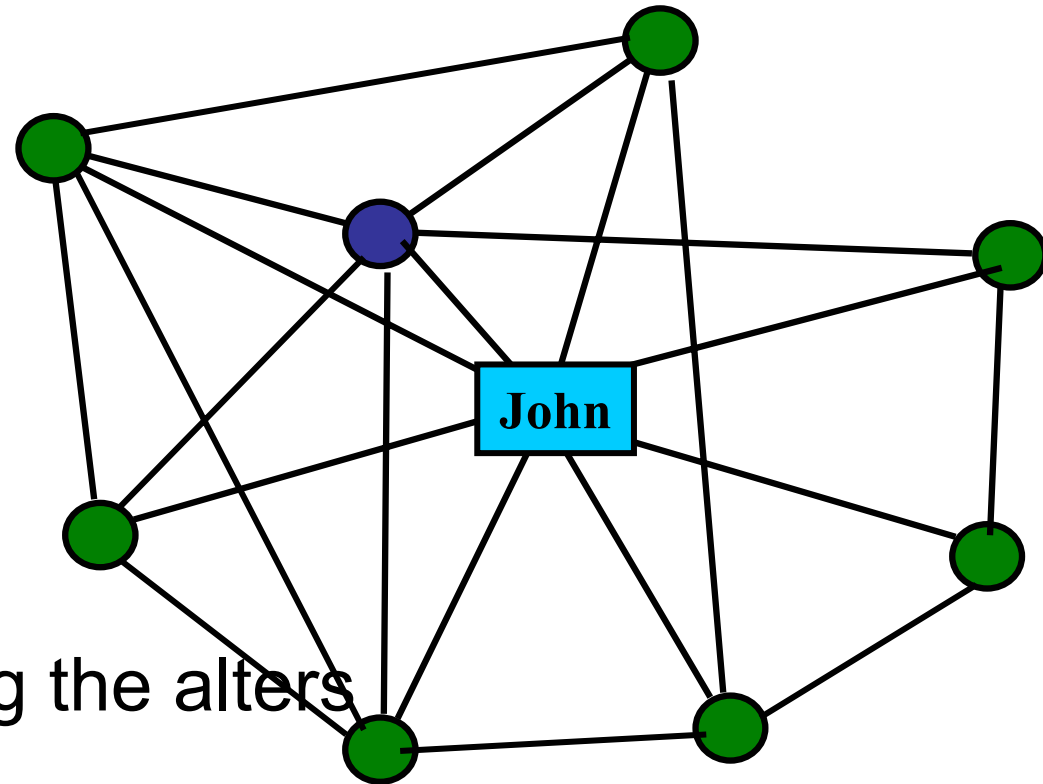
Advantages & Disadvantages of Ego Networks

- Advantages
  - Can take random samples of large population
    - Don't need complete network
  - Use standard survey techniques and programs

- Disadvantages
  - Can't compute cool concepts like graph-theoretic distance
  - Can't see the network structure
  - Inaccurate data (if collected from ego)
  - Peter marsden . Annual review of sociology

Ego Networks
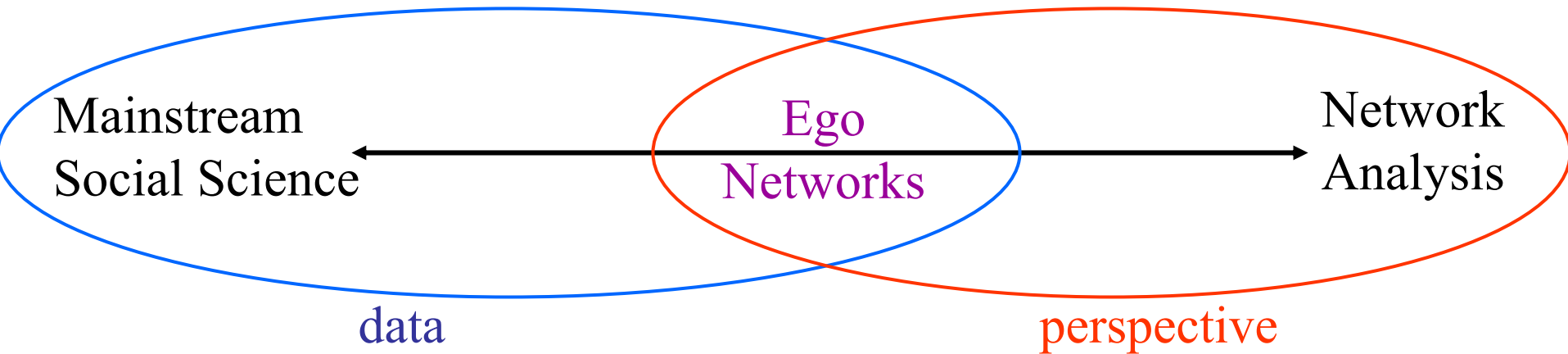
- Combine the perspective of network analysis with the data of mainstream social science

Data Collection

- Name generator to elicit alters
  - Collect distinct names into a roster
- Relationship questionnaire
  - Systematic assessment of the relation that ego has to each alter
- Attribute questionnaire
  - What are the attributes of ego
  - What are the attributes of the alters
- Structure questionnaires
  - Relations among the alters

Analysis of Ego Networks

- If you just have list of contacts for different relations:
  - Degree centrality
    - Who many ties does a person have
  - Multiplexity
- If you have attributes of the egos and alters:
  - Selection: who interacts with whom?
    - Homophily/heterophily
  - Composition
    - What kinds of people are in different kinds of people's networks?
  - Heterogeneity
    - How much diversity in a person's network?
- If you have ties among the alters:
  - Structural holes / density
    - Are ego's alters connected to each other?