

Slide 1



Data Collection & Ethics

The goal of this session is to familiarize everyone with issues of social network data collection. We will discuss issues such as questionnaire design, types of survey media, methods to cope with missing data, and ethical issues of social network analysis.

Objectives:

After this section, you should be able to:

- Describe different methods of collecting Social Network Data
 - Paper vs. Plastic
 - Roster vs. Open-ended
 - Tick vs. Rate
 - Row based / Row & Column Based / Matrix Based
 - Describe ethical issues of Social Network Research

1. Identify the Population: Bounding the Study

- 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, no matter where we cut it
- One key is to isolate when bounding matters
 - Yes: Interpersonal influence studies
 - No: Selection studies

Types of Boundaries

- Attribute-based
 - Top management team at Enron
 - Members of the same work unit
 - Crystal Meth users in Lexington
- Relation-based
 - Snowballing out from seed sample until few or no new names (i.e., exhaust current component)
- Mixed criteria
 - Friendship ties among high school students in Providence
- Theoretical criteria

Slide 5

Gaining Access

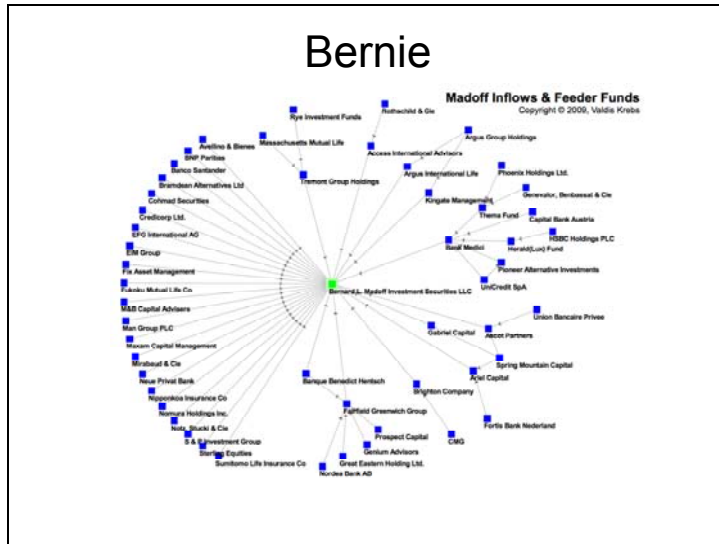
- A little harder than for ordinary studies
 - Strong preference for complete data
 - Respondent fears
 - Length of interview

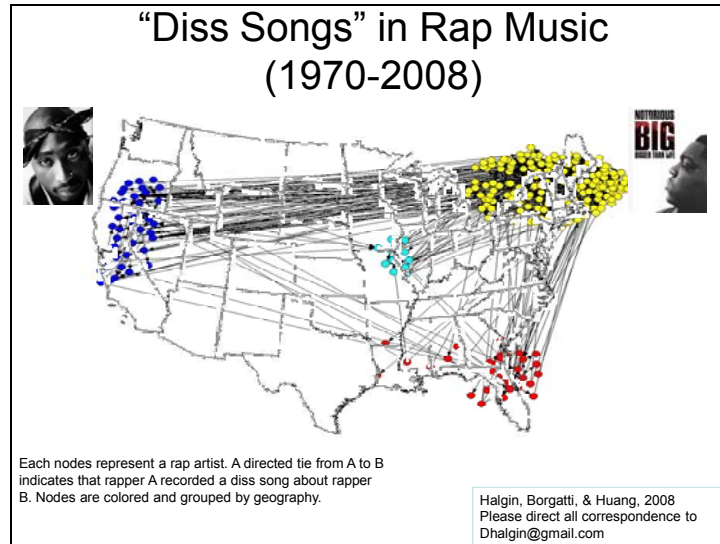
- Quid pro quo helps but muddies the ethical waters

Step 2: Determine Data Sources

- Archival data
- Interviews
- Observations
- Surveys

Slide 7





Step 3: Collect the Data

- What questions to ask?
- How to format your survey?

What Questions to Ask?

- **IT DEPENDS!!!**
 - A relation is just a variable. “Giving advice” is to network analysis what “attitude toward gun-control” is to survey research.
- It is the researcher who defines the relations of interest. 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

Ethnographic Sandwich

- Ethnography at front end helps to ...
 - Select the right questions to ask
 - Word the questions appropriately
 - Create enough trust to get the questions answered
- Ethnography at the back end helps to ...
 - Interpret the results
 - Can sometimes use respondents as collaborators

Survey Design Issue

- Paper or Plastic?
- Close-ended (Roster) vs. Open-ended
- Repeated Roster vs. MultiGrid
- Tick vs. Rate

Paper or Plastic?

- Paper medium
 - Reliable
 - Reassuring to respondents
 - Errors in data entry
 - Data entry is time-consuming
- Electronic
 - Span distances, time zones
 - Harder to lose
 - Fewer data handling errors
 - Lower response rate
 - Emailed documents vs survey instruments

Closed-Ended vs Open-Ended

Roster of names or just blank lines?

- Closed-ended (aided)
 - Requires bounded list
 - Can be impractical for large networks
- Open-ended (unaided)
 - Subject to recall errors
 - Can limit number of choices made (more effort, limited space)

Name	Q1: Head of Desk
Allard, John	<input type="checkbox"/>
Ayer, Jordan	<input type="checkbox"/>
Beane, Ted	<input type="checkbox"/>
Burrows, Rick	<input type="checkbox"/>
Byrnes, Owen	<input type="checkbox"/>
Byrnes, Scott	<input type="checkbox"/>
Bowers, Ralph	<input type="checkbox"/>

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

Copyright © 2006 by Steve Borgatti

Repeated Roster vs MultiGrid

Q1. Please indicate which of the following you had met or been aware of before coming to this workshop.

- Allata, Joan
- Baer, Justin
- Baker, Ted
-

Q2. Check off the names of the people you know. By "know" I mean that you have spoken to each ...

- Allata, Joan
- Baer, Justin
- Baker, Ted
-

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 make a name to a face, you have spoken to each other at least once, and the other person is also likely to get 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 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 Name	Q2: Know them	Q3: Worked with.	Q4: Want to Know
Allata, Joan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Baer, Justin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Baker, Ted	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brandt, Susan	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brunks, Scott	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brower, Ralph	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Copyright © 2006 by Steve Borgatti

Tick or Rate?

- Ask respondent for yes/no decisions or quantitative assessment?
 - Yes/no are cognitively easier on respondent (therefore reliable, believable),
 - Yes/no ***much*** faster to administer
 - But yes/no provides no discrimination among levels
- A series of binaries can replace one quant rating:
 - Instead of "How often do you see each person?"
 - 1 = once a year; 2 = once a month; 3 = once a week; etc.
- Use three questions (in this order):
 - Who do you see at least once a year?
 - Who do you see at least once a month?
 - Who do you see at least once a week?

Question Wording Issues

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

Issues with SNA data

- Asymmetry
- Missing data
- Ethics

Unexpected Asymmetry

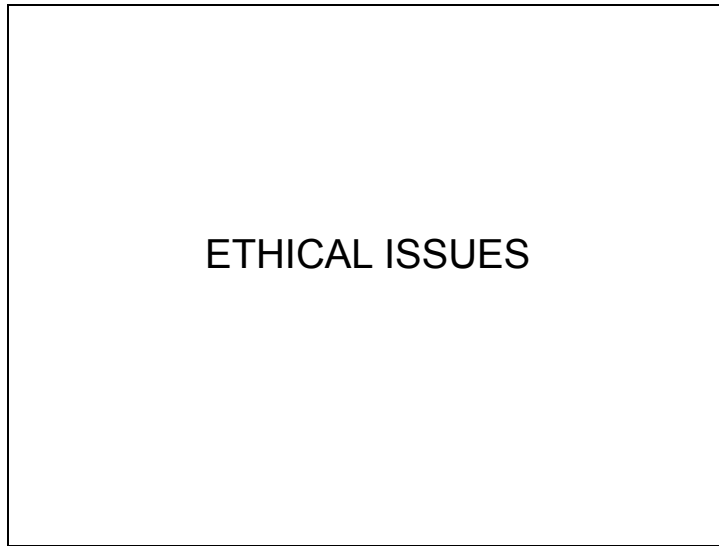
- A claims to have sex with B, but B does not claim to have sex with A
 - The relation is logically symmetric, but empirically asymmetric
 - Errors of recall; strategic response
- Sometimes finding asymmetry is the point
 - Can reveal status differences
- Logically symmetric data may be symmetrized
 - If either A or B mentions the other, it's a tie
 - Only if each mentions the other is it a tie

Non-symmetric Relations

- 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 logically symmetric relations
 - if X_{ij} is missing, substitute X_{ji}
 - If whole row missing, substitute corresponding column
- For logically non-symmetric relations, ask questions both ways (who do you give advice to, who gives advice to you)
 - set $A_{ij} = B_{ji}$
 - i.e., missing row is replaced with column of the inverse relation



Ethical Issues

- What makes network research especially challenging ethically?
- What are the dangers & to whom?
 - In academic setting
 - In management setting
 - In mixed situations
 - In national security setting
- What can we do about it?

Ethical Issues

- Respondents cannot be anonymous
- Non-respondents are still included
- Missing data can be powerful
- Has the potential to be mis-used by Management

Confidentiality Reminder

- This is in addition to consent form

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 must enter your name in Question 0.**

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

Q0. What is your name:

Truly Informed Consent Form

Truly Informed Consent Form

Introduction

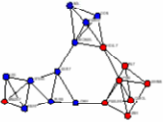
This is a social network study in which we will try to map out the communication network of the organization.

Goals

The academic goal of this study is to understand the factors that determine who talks to whom. We want to understand what factors hinder communication, and which ones facilitate communication. The organization's goal in this study is to improve communication in areas that need it.

Procedures

You will be asked to fill out an online survey about who you interact with regularly, along with background information about yourself, such as training, department you're in, and so on. It should take about 30 minutes to complete. In order to map out who talks to whom, we will need you to give us your name when filling out the survey. Once the data have been collected, we will construct social network maps like the one:



Note that the maps contain each person's name. These maps will be shown to management (specifically, all officers in the organization), but will not be shown to others in the organization. In addition, we will calculate network metrics such as calculating the "degrees of separation" between pairs of people (i.e., the length of the network paths from one person to another).

Summary

- There are three steps to getting started on a social network study
 - Identify the population
 - Determine data sources
 - Collect data
- In addition there are a number of issues that must be considered such as response bias, missing data, unexpected asymmetry, and ethical considerations
