Network Data Collection

Steve Borgatti

MGT 780 Spring 2008

Sources

- Secondary (often 2-mode)
  - Memberships in groups
    - Facebook “networks”
  - Participation in events
    - Listserv threads;
    - DGG deep south data
    - Voting records, e.g. supreme court data
  - Text analyses
    - Weiss, copdab, KEDS
    - Crawdad, automap
  - Other
    - Email records, purchase/sale records, marriage records, etc
    - Emily’s data

Primary Data

- Experiments
  - Rumor planting; milgram small world
- Observation
  - Western-Electric Hawthorne plant studies
  - Ethnographic studies
    - Gary alan fine story telling; whyte street corner etc
- Surveys
  - Telephone, web, paper, etc.

Ego vs Full Network Surveys

- Egonet surveys
  - Randomly sample respondents (egos) and ask about their contacts (alters)
    - The alters are not interviewed
    - One ego's alters are not matched up with other egos or their alters
  - Collect lots of (perceived) info on the alters
  - Analyze homophily, network composition, etc.
- Full network surveys ("regular" sna)

Bounding and Sampling Issues

- Type of sampling*
  - Fixed probability (e.g., random sampling)
  - Adaptive samples (e.g., snowball samples)
  - Population (e.g., all members of frame)
- Type of bounding criteria
  - Attributes (IBM top management team)
  - Relations (anyone engaged in needle-sharing)
  - Combination (anyone in Hartford who injects with anyone in Hartford)
- Stances
  - Nominalist / etic (least delusional approach)
  - Realist / emic (best used for true groups)
  - Combination

Note: Dimensions are not independent

*Sampling of actors. Sampling of ties is also possible, but rarely done in surveys.

	Etic / Nominalist	Emic / Realist
Random sample	Random sample of persons matching researcher needs e.g., random sample of Dem and Rep voters	
Snowball sample	Interview any qualifying actor with a tie to any actor already selected, up to K waves e.g., ask each person who they inject drugs with, then interview those people. Repeat twice more times	Select alters of existing egos until few new names appearing e.g. start self-identified members of group. Ask them for other members. Keep going until it starts petering out
Census	All persons matching researcher criteria e.g., all members of the Anthropology dept.	Get list of "members" from somebody in group e.g., locate gang member, obtain list of members, interview all

Keep in mind …

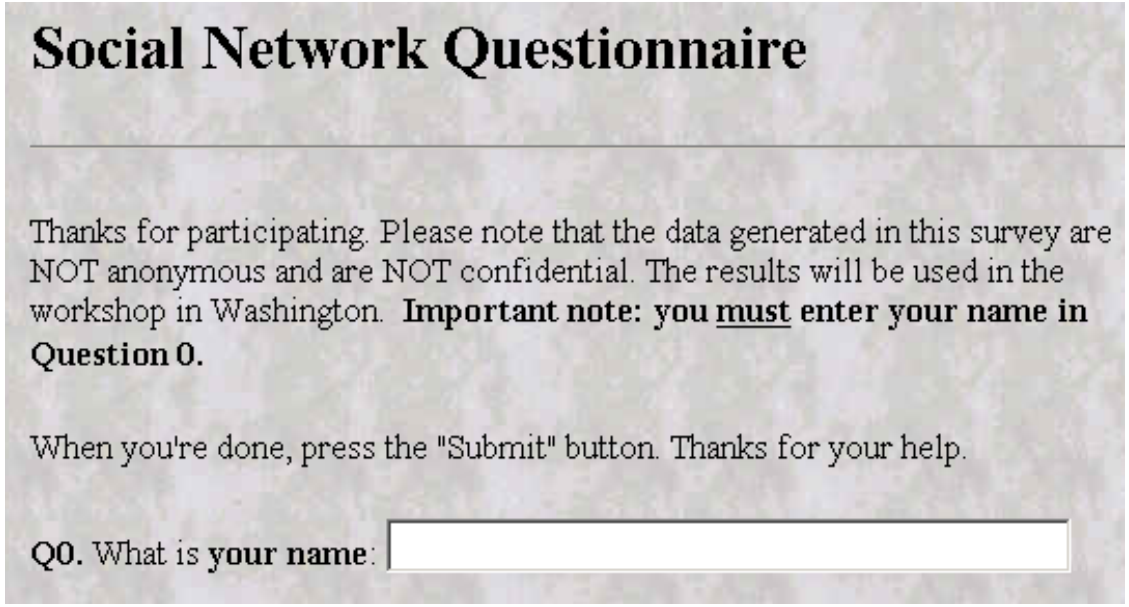
- You get to study whomever you want.
  - The friendship network among redheads at UK
- Only groups have boundaries.
- Bounding is determined by
  - the research question
    - E.g., Adoption influences versus comparative cohesion
  - the analytic technology you will use
- Realism is almost never that

What network questions to ask?

- i.e., which relations to measure
  - Implicit is often the assumption that there is a kind of true network that we are trying to reveal by asking the best relational questions
    - This is like asking in a regular survey of attitudes: which attitudes are the best ones to ask about?
- Answer is: it depends on what the research question is
  - And you are allowed to study whatever you want

Questionnaire elements

- Confidentiality reminder (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**:

Questionnaire Formats

- Aided (rosters) vs unaided (open-ends)
- Ratings, rankings, forced-choice and checkboxes
- Across (grids) or down (separate questions)
- Electronic, paper or other media

Closed-Ended vs Open-Ended

Roster of names or just blank lines?

- Closed-ended (aided)
  - Requires bounded list
  - Can be impractical for large networks
  - Each alter has ~equal chance of choice
- Open-ended (unaided)
  - Subject to recall errors
  - Can limit number of choices made (more effort, limited space)
- Bottom line:
  - I prefer rosters when practical
  - Hybrid designs when not

Name	Q1. Heard of them
Allata, Joan	☐
Baer, Justin	☐
Baker, Ted	☐
Bercuwitz, Rick	☐
Branzei, Oana	☐
Brooks, Scott	☐
Brower, Ralph	☐

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)*

_____ _____

_____ _____

_____ _____

_____ _____

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?

	( )
	( )

Hybrid designs are useful in large networks

- Lookups, dept walk-throughs, etc.

Paper version uses separate booklet containing name directory

Web version uses drop-down menus

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 of f 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 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	☐	☐	☐	☐

Tick or Rate?

- Ask resp for yes/no decisions or quantitative assessment?
  - Yes/no are cognitively easier on resp (therefore reliable, believable),
  - Yes/no *much* faster to administer
  - But yes/no provides no discrimination among levels – ratings provide more nuance
- 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?
- Forced-choice/rankings usually horrible

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

Dillman Design Considerations

- Network questionnaires can be fun but are usually time-consuming and generate anxiety
- Providing value
- Treating resp with respect
- Attractive formatting
- Cloak in authority and importance

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
  - Don't make find distinctions
    - Liking, friendship, esteem, respect, feel positive towards
  - Use homogeneous samples

Multi-item Scales?

- Multiple, similar relational questions risk respondent fatigue & annoyance
  - Who do you give advice to?
  - Who do you give information to?
  - Who do you give guidance to?
  - Who do you counsel?
- Aggregating to larger categories, such as affective & instrumental can work well

Access and Response Rates

- Dillman rules apply
- Significance, prestige and quality
- Giving back to the informant & organization
- Tireless, relentless, unremitting callbacks
- Best organizations / respondents
  - techies
- Minimum response rates
  - Reality or “journality”?
  - Depends on the research question / analysis
  - Also the pattern of non-response

Ethical & Strategic Issues

- What makes the network case 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

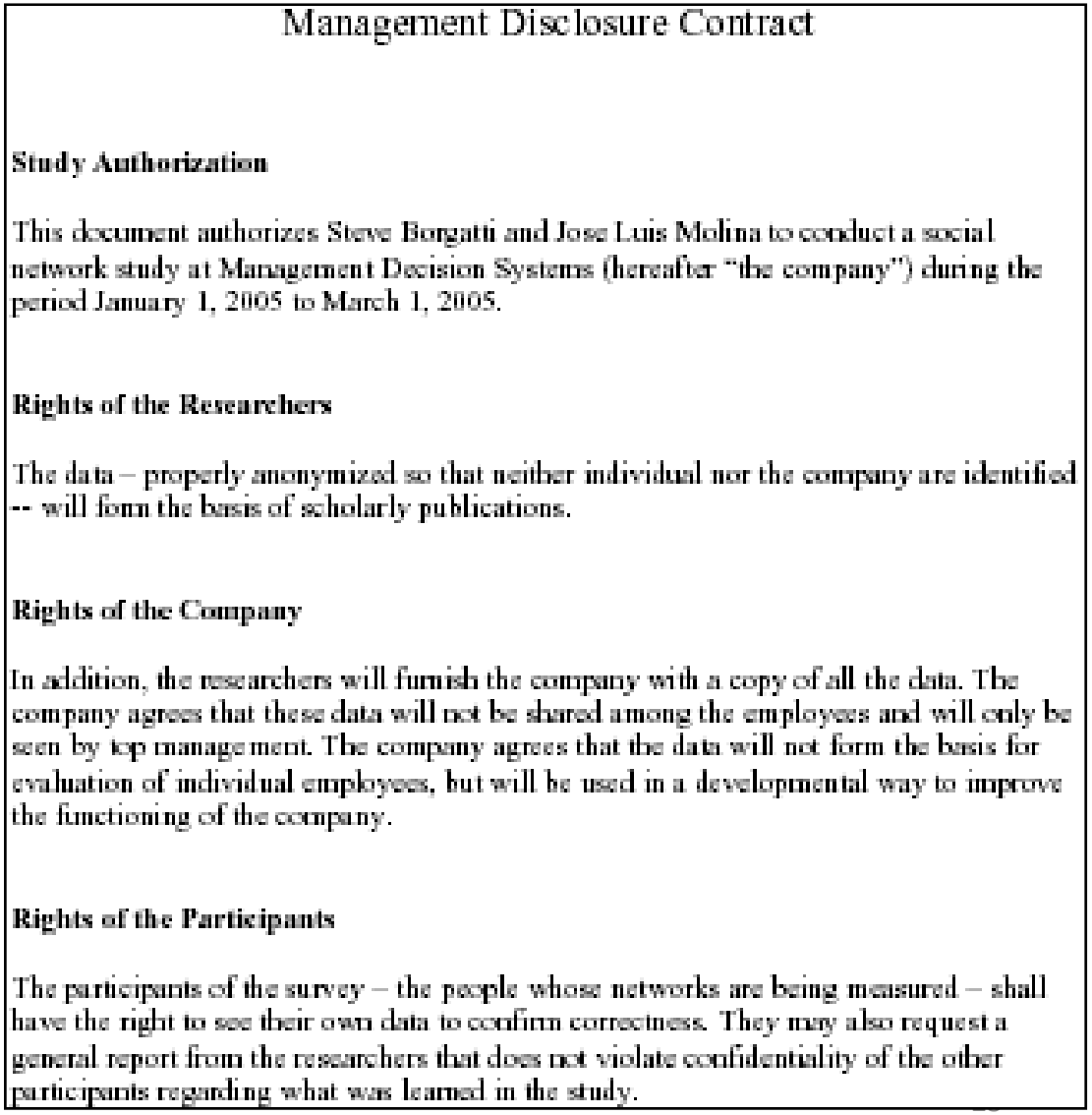
- Respondents cannot be anonymous
- Missing data are troublesome
  - Creating incentive to downplay dangers
  - Results may be wrong (cf use of polygraphs by courts)
- Non-participants still included
  - And participants are like informers
- Outputs ideally show individual level data
- Pushes boundary of the professional
- Deceptively powerful
  - is still unknown; looks like research
- Quid pro quo arrangements with research sites
  - Management is hiring/firing based on “research” results

Ethical Issues

- Consent forms
- Anonymizing
- Non-participation
- Aggregating & categorizing
- Quid pro quos
- Managers' network dashboard

3-Way Disclosure Contract

- For research done in organizations
- Signed by management, the researchers, and each participant
- Clearly identifies what will be done with the data



Management Disclosure Contract

Study Authorization

This document authorizes Steve Borgatti and Jose Luis Molina to conduct a social network study at Management Decision Systems (hereafter "the company") during the period January 1, 2005 to March 1, 2005.

Rights of the Researchers

The data – properly anonymized so that neither individual nor the company are identified -- will form the basis of scholarly publications.

Rights of the Company

In addition, the researchers will furnish the company with a copy of all the data. The company agrees that these data will not be shared among the employees and will only be seen by top management. The company agrees that the data will not form the basis for evaluation of individual employees, but will be used in a developmental way to improve the functioning of the company.

Rights of the Participants

The participants of the survey – the people whose networks are being measured – shall have the right to see their own data to confirm correctness. They may also request a general report from the researchers that does not violate confidentiality of the other participants regarding what was learned in the study.

Truly Informed Consent Form

Truly Informed Consent Form

Introduction

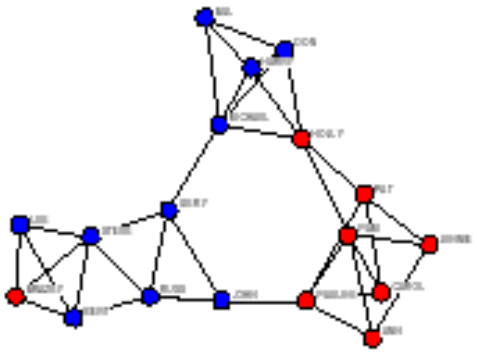
This is a social network study in which we will try to map out the communication network of the organization.

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 this 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).

Page 1

Truly Informed Consent Form

Risks & Costs

Since management will see the results of this study, there is a chance that someone in management could consider your set of communication contacts to be inappropriate for someone in your position, and could think less of you. Please note, however, that the researchers have obtained a signed agreement from management stipulating that the data will be used for improving communication in the company and will not be used in an evaluative way.

Individual Benefits

We will provide you with direct, individualized feedback regarding your location in the social network of the organization.

Withdrawal from the Study

You may choose to stop your participation in this study at any time. If so, you will not appear on any of the social network maps and no metrics will be calculated that involve you. Note that management has agreed that participation in the study is voluntary.

Confidentiality

As explained above, your participation will not be anonymous. In addition, all of top management will be able to see results of the study that include your name. Outside of top management, however, the data will be kept confidential. Any publicly available analyses of these data will not identify any individual by name, nor identify the organization.

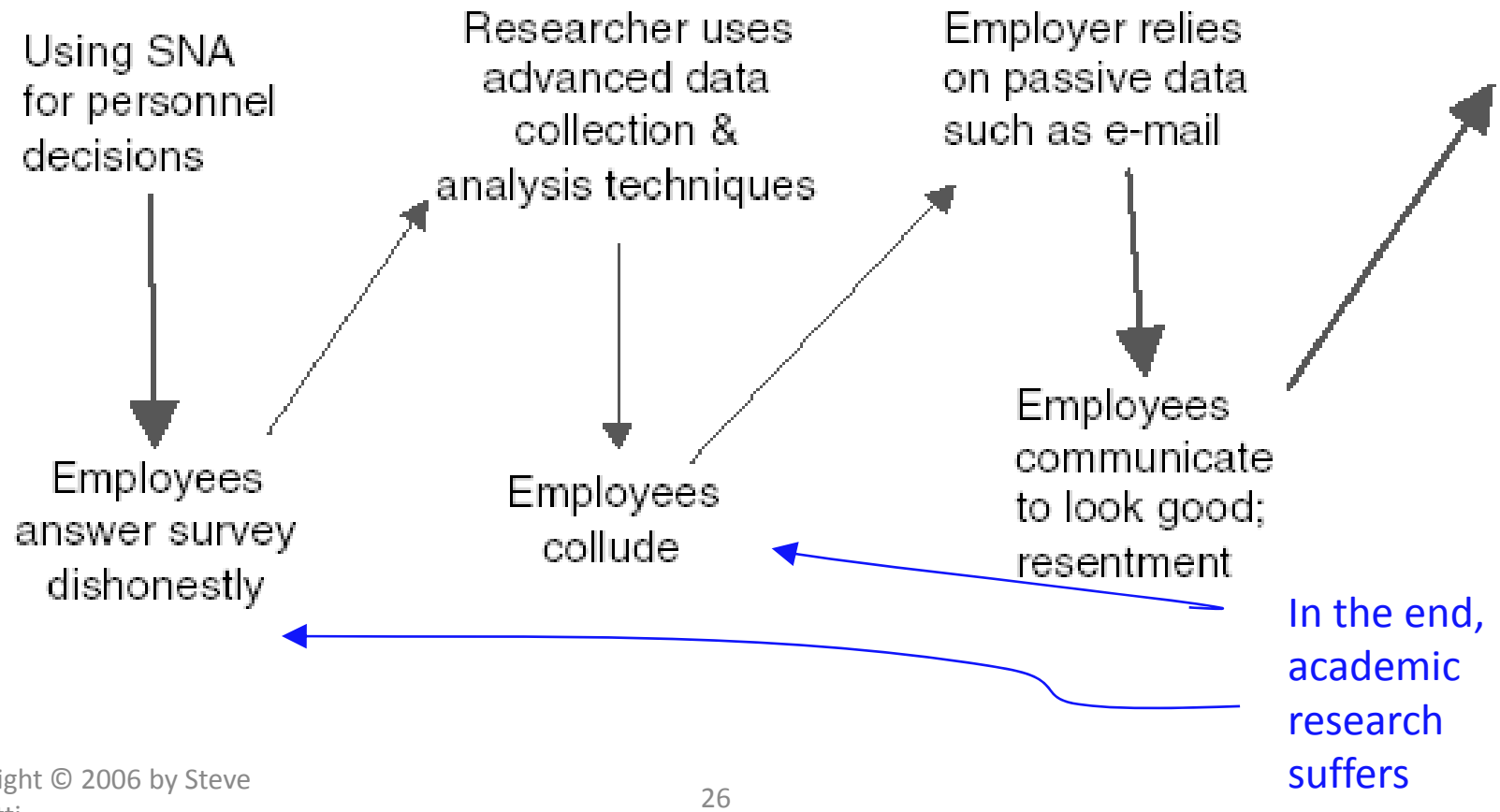
Participant's Certification

I have read and I believe I understand this Informed Consent document. I believe I understand the purpose of the research project and what I will be asked to do. I understand that I may stop my participation in this research study at anytime and that I can refuse to answer any question(s). I understand that management and only management will see the results of this research with individuals identified by name.
I hereby give my informed and free consent to be a participant in this study.

Signatures:

Page 2

The Dialectics of Data Collection



Coping with common data problems

Unexpected Asymmetry

- M claims to have sex with B, but B does not claim to have sex with M
  - The relation is logically symmetric, but empirically asymmetric
  - errors of recall; strategic response
- Sometimes asymmetry is the point
- 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

- 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

- Quick and dirty
  - For logically symmetric relations
    - if Xij is missing, substitute Xji
    - 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 Aij = Bji
    - i.e., missing row is replaced with column of the inverse relation
- Bayesian imputation methods

Krackhardt CSS

Q1. How well the members of each pair know each other:									
	Response scale: Blank = They have never met. 1 = They are merely								
Knowledge	Aaron	Ali	Dan	Dave	David	Ed	George	Greg	Howard
Aaron									
Ali									
Dab									
Dave									
David									
Ed									
George									
Greg									
Howard									

Krackhardt CSS

- Data cube
- Aggregations
  - Row las
  - Col las
  - Intersection LAS
  - Majority rule
- Romney Weller and Batchelder consensus method
  - Class data

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 resps as collaborators