

Data Collection & Entry

Issues in Data Collection

- Gaining access
 - Difficulty varies widely
 - Quid pro quo improves chances but adds ethical issues
- Ethics
 - Various issues, including lack of anonymity
- Informant accuracy
 - Systematic biases: status, long-term
- Ego net vs complete network data
- Data sources
 - Survey questionnaires
 - Records
- Sampling & bounding
- Questionnaire formats
 - media (web, emailed Word/Excel, paper)
 - aided vs unaided
 - 1/0 vs ordinal ratings

Ego Network Data

- Can be a random sample or census
- Typically ask:
 - Respondent's alters (name generator)
 - Attributes of the alters (gender, age, etc.)
 - Possibly, ties among the alters.

Direct Observation

- Hawthorne approach
 - Permanently post observer in back of room to record all interactions
 - No direct access cognitions and affect
- Time allocation (event sampling)
 - Researchers arrive at random times and record interactions
 - % of occasions that i is seen talking with j is index of amount of time i spends with j

Questionnaire Choices

- Adjacency construction strategy
 - Row-based, row & column-based or CSS
- Question formats
 - Closed-ended vs. open-ended
 - Check-off (binary) vs. rating (valued)
 - Multiple grids vs. multi-relational grid
- Dillman total design considerations
 - Providing value, attractive format, etc.
- Informed consent

Construction Strategy

- Row-based
 - Each informant questionnaire corresponds to one row in the network adjacency matrix
 - Issues of comparability, accidental asymmetry
 - For undirected (symmetric) relations, construct data matrix:
 - Intersection rule: $X_{ij} = 1$ if $X_{ij} = 1$ and $X_{ji} = 1$
 - Union rule: $X_{ij} = 1$ if $X_{ij} = 1$ or $X_{ji} = 1$
- Row & Column based (directed relations)
 - A_{ij} : Who do you give advice to?
 - B_{ij} : Who do you get advice from?
 - $X = A \cup B'$ (intersection criterion)
 - $X_{ij} = 1$ iff $(A_{ij} = 1) \text{ AND } (B_{ji} = 1)$
 - i.e., i gives advice to j if i says i gives advice to j and j says they receive advice from i
 - Problem with cognitive & affective relations – resp is expert

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:

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.

Not anonymous

Separate consent form is better

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

Name	Q1. Heard of them	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"/>
Bercuwitz, Rick	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Branzei, Oana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brooks, 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"/>

Closed-Ended vs Open-Ended

- i.e., do you provide a roster of names or just blank lines?
- Closed-ended ...
 - eliminate recall errors
 - Require bounded list
 - Can be impractical for large networks
- Open-ended ...
 - Can limit number of choices made (more effort, limited space)

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:

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.

Not anonymous

Separate consent form is better

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

Name	Q1. Heard of them	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"/>
Bercuwitz, Rick	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Branzei, Oana	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brooks, 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"/>

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 organization, who would you talk to?

_____	_____
_____	_____
_____	_____
_____	_____

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

Tick or Rate?

- i.e., resp asked for yes/no decisions or quantitative assessment?
- Yes/no are easier, faster, more valid, but give cruder data
- A series of binaries can replace 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?

Multiple Grids vs Multi-Relational Grid

- Multiple grids:
 - Each question has separate response area (e.g., roster of names, each with checkbox)
 - Response grid immediately follows each question
- Multi-relational grid:
 - Response grid has multiple columns, one for each relation (question)
 - Question text separate from response grid

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
 - if either A or B mentions the other, it's 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

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

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

Informed Consent

- Questionnaire section
 - explaining the study
 - Detailing who will have access to data and in what level of detail
 - How it will be used
 - Requesting signature indicating understanding
- In roster studies, can separate consent from questionnaire and administer ahead of time

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

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

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

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 empirically 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

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 be done?

- Should consent be obtained to be mentioned in questionnaire?
 - i.e., should open-ends be forbidden?
- Should all network research be prohibited?

Data Entry

- Paper & pencil
 - type responses into text file via special format
 - import into network software package
- Emailed Xcel or Word documents
 - cut and paste from each document into master
 - Then cut and paste into network package
- Web forms
 - may require some pre-processing in Excel
 - Then cut & paste into network package



Data Formats for Text Files

<p>DI n = 5 Format = fullmatrix Labels embedded Data:</p> <table><thead><tr><th></th><th>billy</th><th>john</th><th>jill</th><th>mary</th></tr></thead><tbody><tr><th>billy</th><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><th>john</th><td>1</td><td>0</td><td>1</td><td>0</td></tr><tr><th>jill</th><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><th>mary</th><td>1</td><td>0</td><td>1</td><td>0</td></tr></tbody></table>		billy	john	jill	mary	billy	0	1	0	0	john	1	0	1	0	jill	0	0	0	1	mary	1	0	1	0	<p>DI n = 5 Format = nodelist Labels embedded Data:</p> <p>Billy jill john jim jane Jill billy bob bertha Dick jane Jim bob billy brenda</p>	<p>DI n = 5 Format = edgelist Labels embedded Data:</p> <p>Billy jill Billy john 6.3 Dick jane Jim bob 2.5</p>
	billy	john	jill	mary																							
billy	0	1	0	0																							
john	1	0	1	0																							
jill	0	0	0	1																							
mary	1	0	1	0																							

(Values optional)

(No values possible)

(Values optional - assigned 1
if omitted)

Sampling & Bounding

- Sampling & bounding not a problem for ego networks
- Sampling for complete networks is in its infancy
 - Snowballing especially problematic
- Bounding
 - Node attributes, such as dept
 - Relations, such as does drugs with
 - Combination

Issues with boundaries

- When the dependent variable is attitude formation or other actor transformations, need to bound correctly
 - Organic farmers
- For autocorrelation studies, bounding is not critical
- For snowball samples, must exclude edge