



Social Network Analysis: Data collection II

Song Yang, Ph.D.

University of Arkansas

Data
collection
process

Name generator

Positional generator

Resource generator

Informant bias and data reliability

Name generator

The most well-known name generator is *the important matter discussion generator*, implemented in the 1985 and 2004 General Social Surveys (GSS). The name generator prompt is as follows: “From time to time, most people discuss important matters with other people. Looking back over the last six months, who are the people with whom you discussed matters important to you? Just tell me their first names or initials.”

Name
interpreter

SOCIAL NETWORK IN ACTION: COLLECTING ALTERS' INFORMATION IN EGOCENTRIC NETWORK STUDIES

Questions about the alter's characteristics"

Is (NAME) Asian, Black, Hispanic, White, or something else?

ASK FOR EACH NAME

How old is (NAME)?

ASK FOR EACH NAME

Questions about ego-alter ties:

How close do you feel to this person? Please describe how close you feel on a scale from 1 to 5—1 means not close, 5 means very close.

Questions about the alter relational structures:

Please think about the relations between the people you just mentioned. Some of them may be total strangers, in that they would not recognize one another if they bumped into each other on the street. Others may be especially close, as close to each other as they are to you.

First, think about (NAME 1) and (NAME 2),

A: Are they total strangers?

IF YES, PROCEED TO NEXT PAIR

B: Are they especially close?

REPEAT FOR EACH PAIR OF NAMES

Do Americans become loners?

Table 1. Size of Discussion Networks, 1985 and 2004^b

Network Size	Total Discussion Network		Kin Network ^a		Non-Kin Network ^a	
	1985	2004	1985	2004	1985	2004
0	10.0%	24.6%	29.5%	39.6%	36.1%	53.4%
1	15.0%	19.0%	29.1%	29.7%	22.4%	21.6%
2	16.2%	19.2%	21.0%	16.0%	18.1%	14.4%
3	20.3%	16.9%	11.7%	9.4%	13.2%	6.0%
4	14.8%	8.8%	5.8%	4.0%	6.8%	3.1%
5	18.2%	6.5%	2.8%	1.3%	3.4%	1.4%
6+	5.4%	4.9%	—	—	—	—
Mean	2.94	2.08	1.44	1.12	1.42	.88
Mode	3.00	.00	.00	.00	.00	.00
SD	1.95	2.05	1.41	1.38	1.57	1.40

Note: N (1985) = 1,531; N (2004) = 1,467.

Other network features

Age Heterogeneity (standard deviation of age of alters)		
<5	25.8%	29.1%
5–<10	24.6%	19.7%
10–<15	24.3%	23.9%
>15	25.3%	27.3%
Mean	10.35	10.34
SD	6.96	8.1
Population Age Heterogeneity	20.89	18.37
Education Heterogeneity (standard deviation of alters' educations)		
0–1	31.9%	34.7%
>1–2.5	41.0%	45.2%
>2.5	27.0%	20.1%
Mean	1.77	1.48
SD	1.52	1.38
Population Educ Heterogeneity	3.59	3.17
Race Heterogeneity (Index of Qualitative Variation) ^c		
0	91.1%	84.5%
>0	8.9%	15.4%
Mean	.05	.09
SD	.18	.26
Population IQV	.34	.53
Sex Heterogeneity (Index of Qualitative Variation)		
0	23.8%	24.2%
.01–.90	39.9%	37.6%
>.90	36.3%	38.1%
Mean	.67	.68
SD	.43	.46
Population IQV	.99	1.00

Issues with the 2004 GSS SNA

Interview effects

- Interviewers untrained, undertrained, or uninterested to collect network data

Online network not mentioned

- People switch from offline to online, which is not reflected in the surveys.

Positional generator

SOCIAL NETWORK IN ACTION: USING POSITIONAL GENERATOR IN EGO-CENTRIC NETWORK STUDIES

Among your relatives, friends, or acquaintances, are there people who have the following jobs?

- a. High school teacher
- b. Electrician
- c. Owner of small factory
- d. Nurse

For each job for which the respondent answers "yes," ask what is his or her relationship to you?

- 1. Relative
- 2. Friend
- 3. Acquaintance

Positional generator

Assumption: the formal position of one's social contacts serve as one's social resources through social ties. Positions of not the oneself, but his/her contacts (families, friends, acquaintances) are proxies of social resources that can potentially be mobilized to facilitate one's personal gains/goals.

In Taiwan, social resources acquired through knowing somebody in high prestigious position help men to obtain high position and income. But they do not help women, who have to rely on human capital for the career advancement.

What
positional
generator
can
measure

Extensiveness: the number of different positions generated by the respondents.

Upper reachability: the highest occupation status report among ego's social contacts

Range: the distance between the most prestigious position and the lowest position among ego's social contacts.



Resource generator

SOCIAL NETWORK IN ACTION: USING RESOURCE GENERATORS TO COLLECT EGO-CENTRIC NETWORK DATA

Do you know anyone who

- A: can repair a car, bike, etc.?
- B: can visit socially?
- C: repairs household equipment?
- D: knows a lot about government regulations?
- E: can give a good reference when you are applying for a job?

(Continued)

(Continued)

For each item to which the respondent answers “yes”:

What is his/her relationship to you?

- A: family members
- B: friend
- C: acquaintance

Informant bias

Informant bias: discrepancy between self-reported and actual behavioral data

Early 1930s, Richard LaPierre and a Chinese couple travel across the US, eating at 184 restaurants and staying in 66 hotels, were refused by only one hotel.

LaPierre later sent a questionnaire to the same establishment, asking them “whether they would accept members of Chinese race.” 92 percent replied “no.”

Sources of informant bias

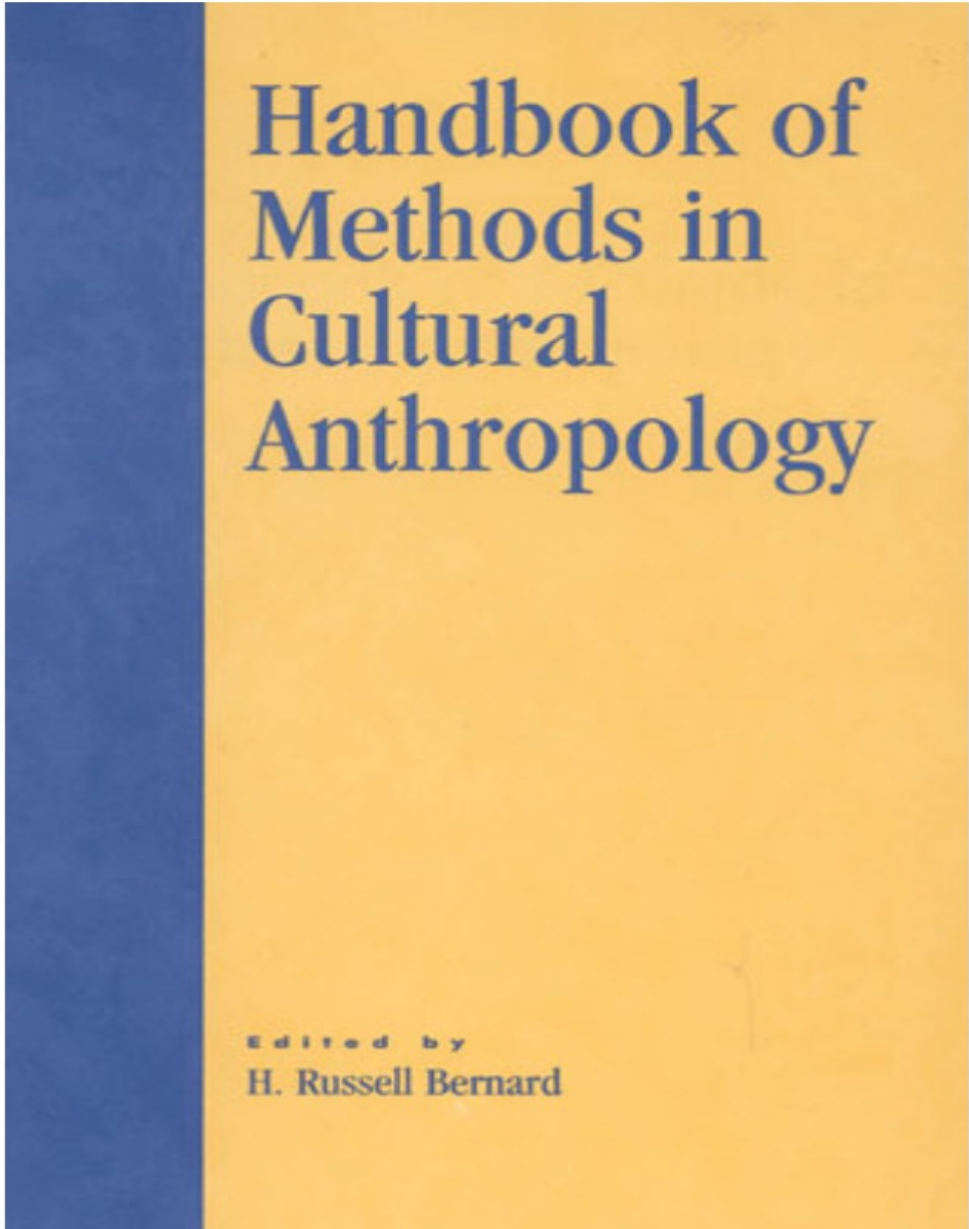
People can not cognitively handle the large amount of data required to report their behaviors accurately.

Bias is not random, rather some informants are better than others at providing accurate accounts of their actions.

Familiarity with the interview topic can boost informant recall of events.

However, although informants with great familiarity with the topic forget a little, they tend to falsely recall – reporting on nonexistent members.

People do not know, with any acceptable accuracy, to whom they talk over any given period of time ... We are now convinced that cognitive data about communication can NOT be used as proxy for the equivalent behavioral data.



Handbook of Methods in Cultural Anthropology

Edited by
H. Russell Bernard

Research findings on informant accuracy

Highly knowledgeable informants produce unbiased data about long-term repeated pattern.

Highly knowledgeable informants also tend to produce consensus answers to questions, which indicates greater validity.

Individual's perceptions are biased toward portraying themselves as more central in a network than does the systematic aggregation of all network actor's perception.

Reliability

Reliability measures the extent to which a particular instrument, when applied repeatedly to the same subject, yields an identical result every time.

E.g. ask the respondents to name their friends with whom they discuss important matters.

Three months later, ask the same set of respondents the same questions.

Reliability

A complete correlation between the two set of responses – all respondents enumerate the same sets of alters both times, indicates the greatest reliability.

A complete disjoint answers, in which a set of informants choose a completely different set of alters each time, suggest no reliability.

A problem with such measure of reliability?

A test-retest reliability measure

Jaccard's coefficient

$$= \frac{\text{number of persons nominated by an informant at both times}}{\text{Total number of unique nominations for both test}}$$

	Second Interview	
First Interview	Names checked	Not checked
Names checked	A=17	B=5
Not checked	C=8	D=15

What's Jaccard coefficient for such data?

Note that

- 1) Jaccard coefficient ranges from 0, indicating no reliability, to 1, indicating complete reliability.
- 2) Jaccard coefficient is an individual characteristics, a sample/group level Jaccard coefficient can be the average of the individual coefficients.

Informant reliability and validity

- Romney and Weller (1984) reanalyzed Bernard's dataset,
 - informant validity is positively correlated with its reliability – valid informants are also reliable.
 - Informants with high validity produce consensus answers with each other than would informants with low validity.
 - informants should be weighted according to their reliability and validity.

Cues and prompts

Does providing a checklist with cues and prompts help informants to achieve higher validity?

Brewer and Webster's (1999) study

- Interview 217 students living in a dormitory building
- First interview: write down all your friends living in the hall
- Second interview: provide informants with a roster of all the residents in the dorm and ask them to identify all their friends
- On the average, the second interview elicits 20% more friends than does the first interview.
- Informants tend to forget more distant – less intimate alters without the cues.