methodological challenges and theoretical opportunities of collecting large personal networks in large samples
Disclaimer

LARGE NETWORKS

25

LARGE SAMPLES

700
one kind of social interaction, informal conversations with networks of relatives, friends, and neighbours, was important for historical change in bedroom behavior

WATKINS 1995
Social Influence & Fertility

historical evidence

Spatial Analysis of the Causes of Fertility Decline in Prussia

Joshua R. Goldstein
Sebastian Klüsener

convenience samples

Does Fertility Behavior Spread among Friends?

Nicoletta Balbo and Nicola Barban

American Sociological Review
2014, Vol. 79(3) 412–431
© American Sociological Association 2014
DOI: 10.1177/0003122414531596
http://asr.sagepub.com

qualitative studies

Channels of social influence on reproduction

Laura Bernardi
Max Planck Institute for Demographic Research

social learning
social contagion
social pressure
social support
quantifying social influences on fertility behaviour using personal network data
Bigger Is Better (?)

weak ties
structure characteristics
Data Collection Worries

Number of alters

Number of alter-questions

0 5 10 15 20 25 30 35 40 45 50

0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300

1 question

13 questions

25 questions

25, 325
Data Collection Worries

Number of alters vs. Number of alter-questions

- 1 question
- 13 questions
- 25 questions

Number of alters:
- 0
- 5
- 10
- 15
- 20
- 25
- 30
- 35
- 40
- 45
- 50

Number of alter-questions:
- 0
- 100
- 200
- 300
- 400
- 500
- 600
- 700
- 800
- 900
- 1000
- 1100
- 1200
- 1300

Points:
- (25, 325)
- (25, 300)
Does the online collection of ego-centered network data reduce data quality? An experimental comparison

Uwe Matzat, Chris Snijders

YES
GENSI: A new graphical tool to collect ego-centered network data

Tobias H. Stark\textsuperscript{a,}*, Jon A. Krosnick\textsuperscript{b}

\textsuperscript{a} Utrecht University\textsuperscript{c}, Padualaan 14, 3584 CH Utrecht, The Netherlands
\textsuperscript{b} Stanford University, 450 Serra Mall, Stanford, CA 94305, United States
compared to standard survey-methods, people who used GENSI:
- enjoyed the survey more
- thought the survey was more interesting
- said they were more willing to participate in a future survey
“A practical limitation for future research with GENSI is that the tool is only suitable for small ego-centered networks. When the number of alters exceeds seven or eight, it gets visually challenging to see all circles in a network.”
Collecting large personal networks in a representative sample of Dutch women, using GENSI
Longitudinal Internet Studies for the Social sciences
True probability sample of households drawn from the population register.
Respondents participate in monthly Internet surveys.
Extensive background information available on respondent
High retention rates (e.g., 70 %)

All women between 18 - 40 asked (N = 1322)
N = 758 responded (57%); age: 29 (± 6)
Incentive: 12.50 euro
Period of 1 month (~ March)
Methodology

**Ego**

- Detailed fertility intentions

**Alters (25)**

- Sex
- Age
- Education
- Relationship type
- Closeness
- Frequency of contact F2F
- Frequency of other contact

- Number and age of children
- Friend
- Wants children
- Does not want children
- Help with children
- Talk about children
- Relationship with other alters
Please list 25 names of individuals 18 years or older with whom you have had contact in the last year. This can be face-to-face contact, but also contact via phone, internet, or email. You know these people and these people also know you from your name or face (think of friends, family, acquaintances, et cetera). You could reach out to these people if you would have to. Please name your partner in case you have one.
How close are you to these people?
Als het gaat om ANNE

Met wie heeft ANNE contact? Met contact bedoelen we alle vormen van contact, zoals face-to-face contact, contact via (mobiele) telefoon, post, email, SMS, en andere manieren van online en offline communicatie.

Selecteer de personen die contact met elkaar hebben door met de muis op het bolletje te klikken. Er zal een lijn ontstaan die aangeeft dat de personen contact met elkaar hebben. Druk nogmaals op het bolletje om de lijn weer te laten verdwijnen, als de personen geen contact met elkaar hebben.
THE TASK
coming up with 25 names,
answering 16 questions about all alters,
evaluating 300 alter-alter ties

THE RESULT
50% within 21 minutes
97% hardly any missing values
Listing first alter took about 30 seconds

N = 654
Listing the last alters took about twice as long as listing the first alters, but still only about 10 seconds.
Listing all 25 alters took around 3.5 minutes

N = 654
N = 654

- Naming 25 alters took 3.5 minutes
- Listing all alter-alter-ties took 3.5 minutes
- Responding to all alter-question took 15 minutes
Collecting large personal networks feasible
Not too time-consuming
Little missing data
Data quality?
Quantity ≠ Quality

- Partner
  - via partner
  - via partner

- Mother
  - Father
  - Brother
  - Sister
Quantity ≠ Quality

89% of all possible ties were reported

87% of all possible ties were reported
Conclusion

Collecting large personal networks feasible
Not too time-consuming
Little missing data
Data quality?

GENSI useful for large(r) networks
Improved user experience?
65% of the respondents enjoyed filling in the survey, whereas 10% did not enjoy it so much

Did you enjoy answering the questions?

- Definitely not: 4%
- 2: 6%
- 3: 26%
- 4: 32%
- Definitely yes: 33%

N = 691
Collecting large personal networks feasible
Not too time-consuming
Little missing data
Data quality?

GENSI useful for large(r) networks
Improved user experience?

Valuable data
Programmed in JavaScript

Pros
- “Light” (1 Mb)
- Works with any device with browser
- Can be implemented in other surveys
- Free

Cons
- Not ideal for mobile phones (currently)
- Answers can’t be saved for later use
- Being able to “go back” requires considerable coding

www.tobiasstark.nl/GENSI
www.gertstulp.nl/GENSI
Alternatives

Network Canvas  https://www.networkcanvas.com/

OpenEddi  https://github.com/jfaganUK/openeddi3

GENTLE  https://www.gentle.eu/
Balancing Bias and Burden

Scientific interest
Weak ties
Network structure
Network composition

Respondent burden
Time
Boredom
Poor(er) response
Quantifying Bias

evaluating two strategies to reduce burden by lowering number of alters

1. dropping alters

2. random subset
network structure
Density
Proportion of Isolates
Maximum Degree
Degree Centralisation
Betweenness Centralisation
Mean Betweenness Centrality
Maximum Betweenness Centrality
Closeness Centralisation
Mean Closeness Centrality
Maximum Closeness Centrality

network composition
Average and SD of:
Alter age
Closeness
Frequency of F2F contact
Frequency of other contact
Education

Proportion of:
Female Alters
Friends
Kin
https://socialsciencemethods.shinyapps.io/BalancingBiasAndBurden
Conclusions

Lowering number alters increases bias
15-20 ‘sufficient’ for most measures

Randomly sampling alters superior to dropping alters
More consistent, less bias

More bias in structural versus compositional measures
Huge variation
A potentially useful strategy:

1) Eliciting large number of alters
2) Alter-alter-ties for random sample
3) Alter attributes for smaller subsample

Results can serve as guide for novel data collection

https://socialsciencemethods.shinyapps.io/BalancingBiasAndBurden

Carefully examine outcome
Amount of bias versus time gains
Time gains through different type of questions
Practical Guide

A potentially useful strategy:

1) Eliciting large number of alters
2) Alter-alter-ties for random sample
3) Alter attributes for smaller subsample

Results can serve as guide for novel data collection

https://socialsciencemethods.shinyapps.io/BalancingBiasAndBurden

Carefully examine outcome
Amount of bias versus time gains
Time gains through different type of questions

Results May Vary
"representative" survey experience paid well
LARGE NETWORKS
LARGE SAMPLES

PART I
PART II
PART III

Vera Buijs
Friends, Family, Family Friends

family of choice
close
seen often
long-term
“Friends”

- High-quality relation
- Role relation
- Inconsistent concept

People vary in use "residual category"

Close people you want to see often

Mutual agreement role-related norms
AIM

predicting who is considered a friend among *kin* and *non-kin* using three measures of tie strength:

- closeness
- frequency of f2f contact
- frequency of other forms of contact

SETUP

1. Personal characteristics (e.g. age of respondent)
2. Alters (25 names)
3. Origin of the relationship ("What is your relationship with <name> or how do you know him/her?")
4. Relationship characteristics (e.g. closeness to alter, per alter)
5. Friendship ("Which of these people do you consider a friend?")
701 respondents reporting on 17,525 alters classified 7,331 as friends on average 10 friends (SD = 5)
Friend certainly not orthogonal to family

- High-school: 1100, 83%
- College: 1806, 74%
- Primary school: 514, 70%
- Partner: 489, 60%
- Social activity: 1717, 56%
- Partner's friends: 903, 56%
- Mutual acquaintance: 1295, 55%
- Neighbourhood: 717, 32%
- Other: 98, 31%
- Work: 2571, 29%
- Sibling: 1190, 27%
- Kin: 2485, 16%
- Parent: 1226, 15%
- In-law: 1324, 13%
Closeness strong predictor of friendship particularly in non-family, not close people also considered friends.
Frequency of face-to-face contact weaker predictor, different effect in family versus non-family

<table>
<thead>
<tr>
<th>face-to-face contact</th>
<th>% labelled friend</th>
</tr>
</thead>
<tbody>
<tr>
<td>daily</td>
<td>58%</td>
</tr>
<tr>
<td>few times a week</td>
<td>51%</td>
</tr>
<tr>
<td>few times a month</td>
<td>58%</td>
</tr>
<tr>
<td>once a month</td>
<td>60%</td>
</tr>
<tr>
<td>few times a year</td>
<td>49%</td>
</tr>
</tbody>
</table>

Family: 39%, 26%, 18%, 11%, 8%
Non-family: 58%, 51%, 58%, 60%, 49%
Frequency of other forms of contact consistently predicts friendship, but much weaker than closeness.
Prediction

Prediction accuracy of friendship based on measures of tie strength:
[closeness, frequency of f2f contact, frequency of other forms of contact]

Family

80%

Non-family

77%
Prediction accuracy of friendship based on measures of tie strength:
[closeness, frequency of f2f contact, frequency of other forms of contact]

Family
- Baseline: 80%

Non-family
- Baseline: 77%

Prediction accuracy of friendship based on measures of tie strength:
[closeness, frequency of f2f contact, frequency of other forms of contact]

Family

80%
baseline

Non-family

77%
baseline

• 3 measures of tie strength!
• No family
• Homogenous sample
“Friends”

Inconsistent concept

Role relation

High-quality relation

People vary in use “residual category”

Close people you want to see often

Mutual agreement role-related norms

Kitts & Leal 2021 [GO READ!]
Asking for a friend...

probably too vague a concept to be used in scientific research

Claude Fischer (1982)

when using name generators:
- asking for friends might give you in-laws
- asking for family might give you friends
- asking for close, frequently seen people might not give friends

when used as classification:
- friend not orthogonal to family, neighbours, colleagues
- people vary in use, some unpredictable some predictable (e.g. age, sex)
Shrinking kin-networks
Dense networks can provide more support, particularly in kin

- Consanguineal kin
- Affinal kin
- Friend
- Not friend

% Dense networks can provide more support, particularly in kin

- can ask for help with childcare
- can talk to about having children

Density
Dense networks can provide more support, particularly in kin

- Consanguineal kin
- Affinal kin
- Friend
- Not friend

- Can ask for help with childcare
- Can talk to about having children

Density %

0 25 50 75 100

0 25 50 75 100
methodological challenges and theoretical opportunities of collecting large personal networks in large samples

- Stulp, G. Collecting large personal networks in a representative sample of Dutch women. Social Networks
- Stulp, G & Barrett, L. Do data from large personal networks support cultural evolutionary ideas about kin and fertility?