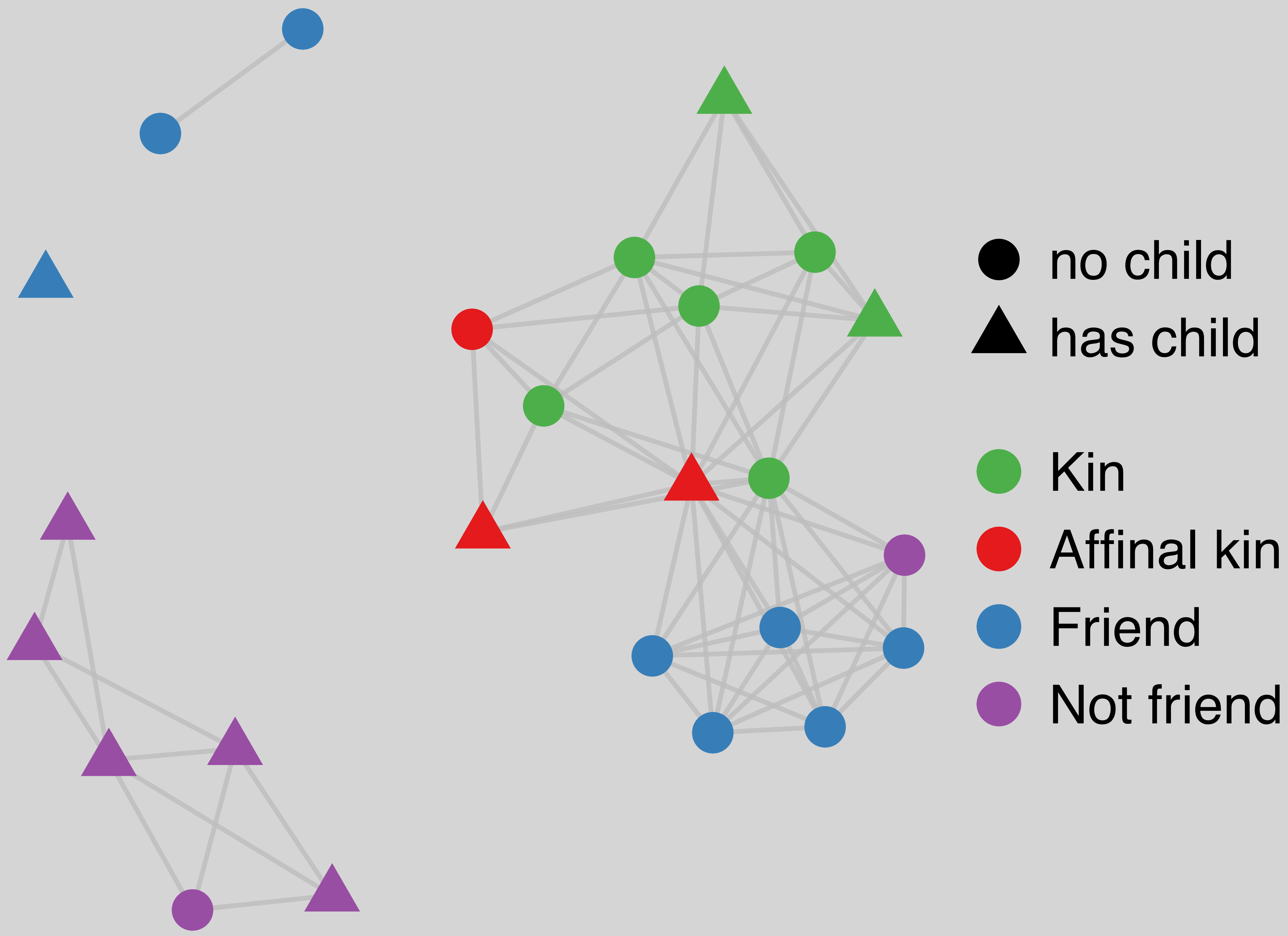
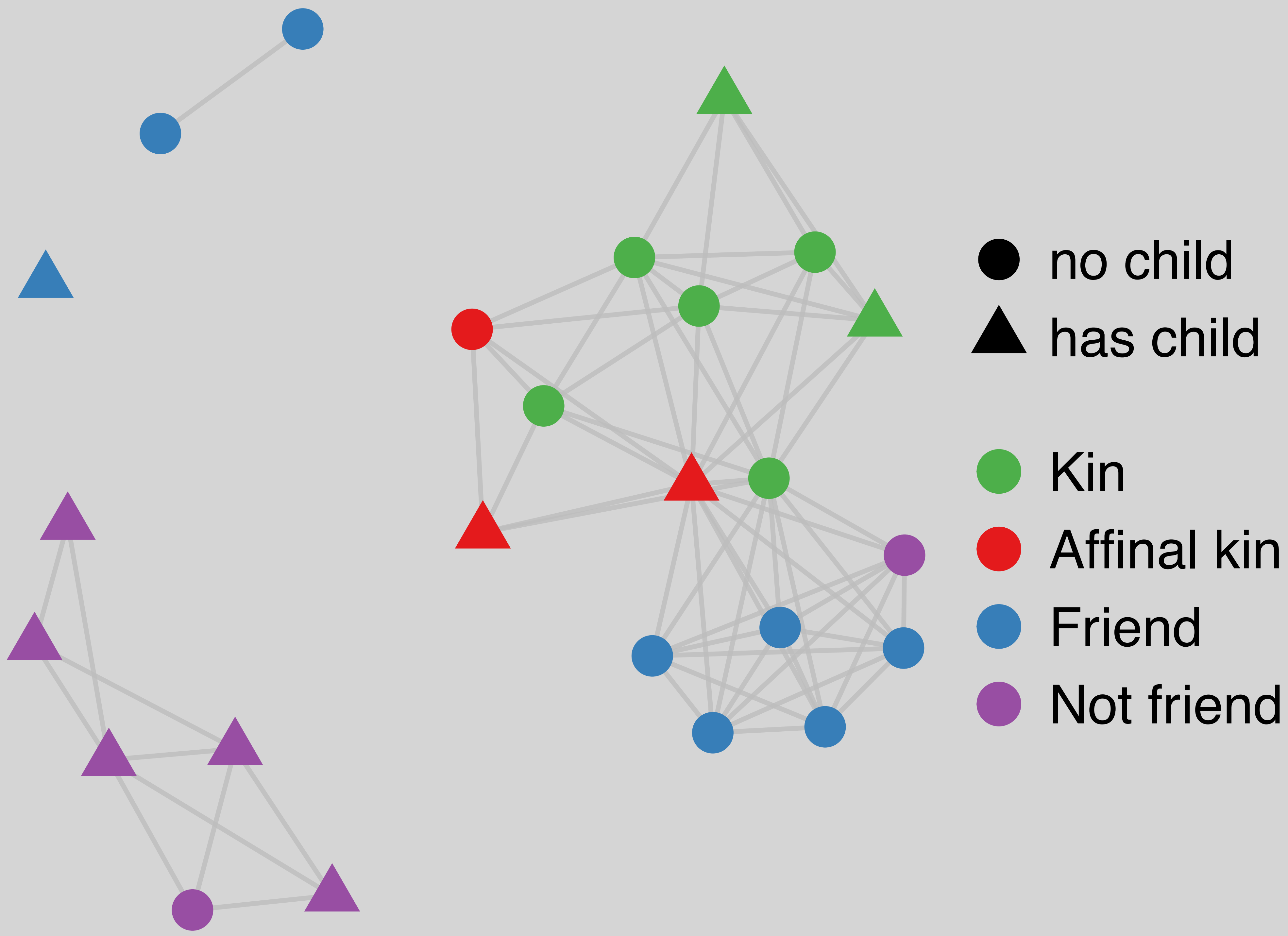
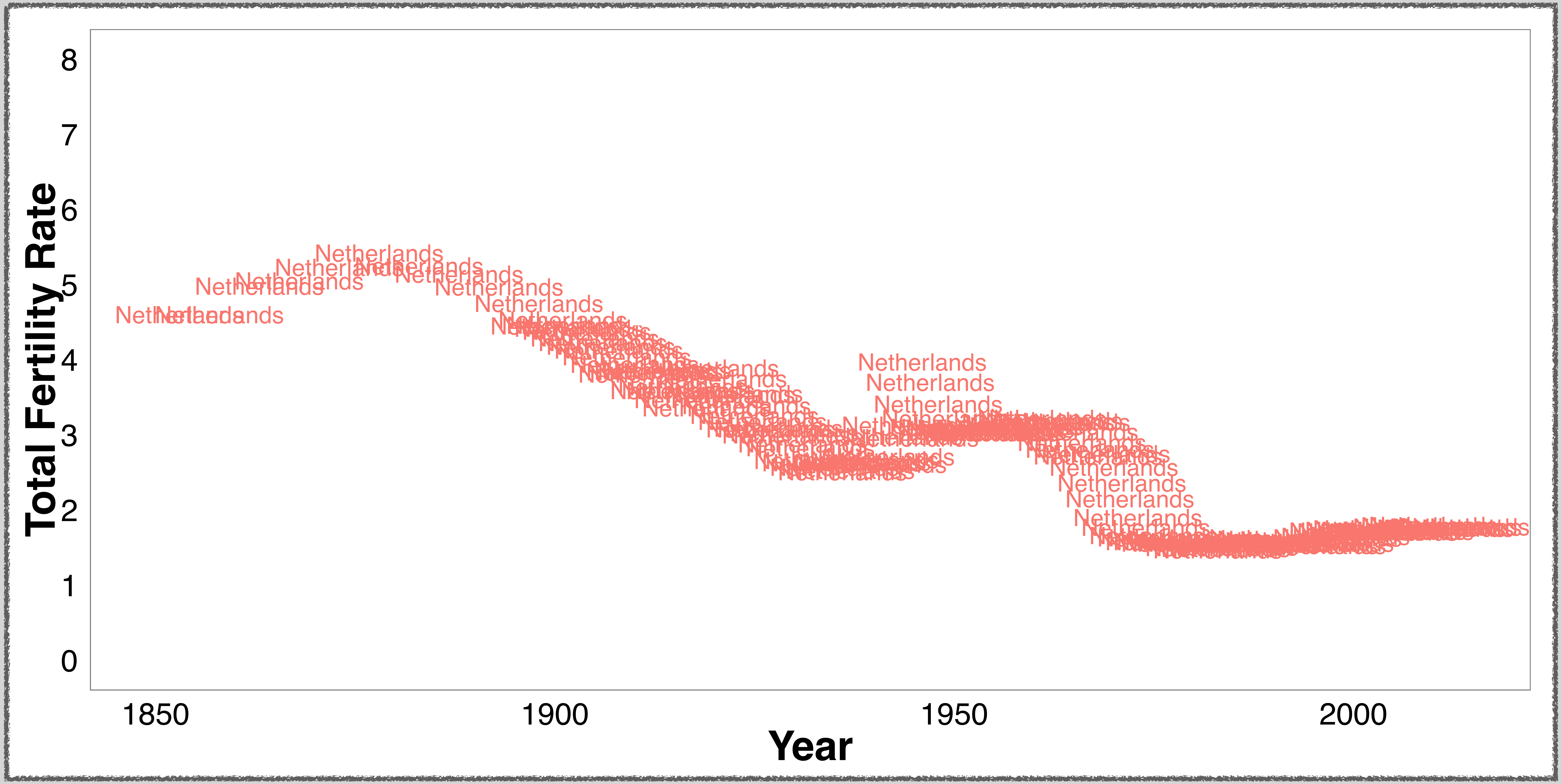
A data-driven approach shows that individuals' characteristics are more important than their networks in predicting fertility outcomes



gert stulp, lars top, xiao xu, lisa sivak

“A complicated data-mining exercise, with much oversold results”

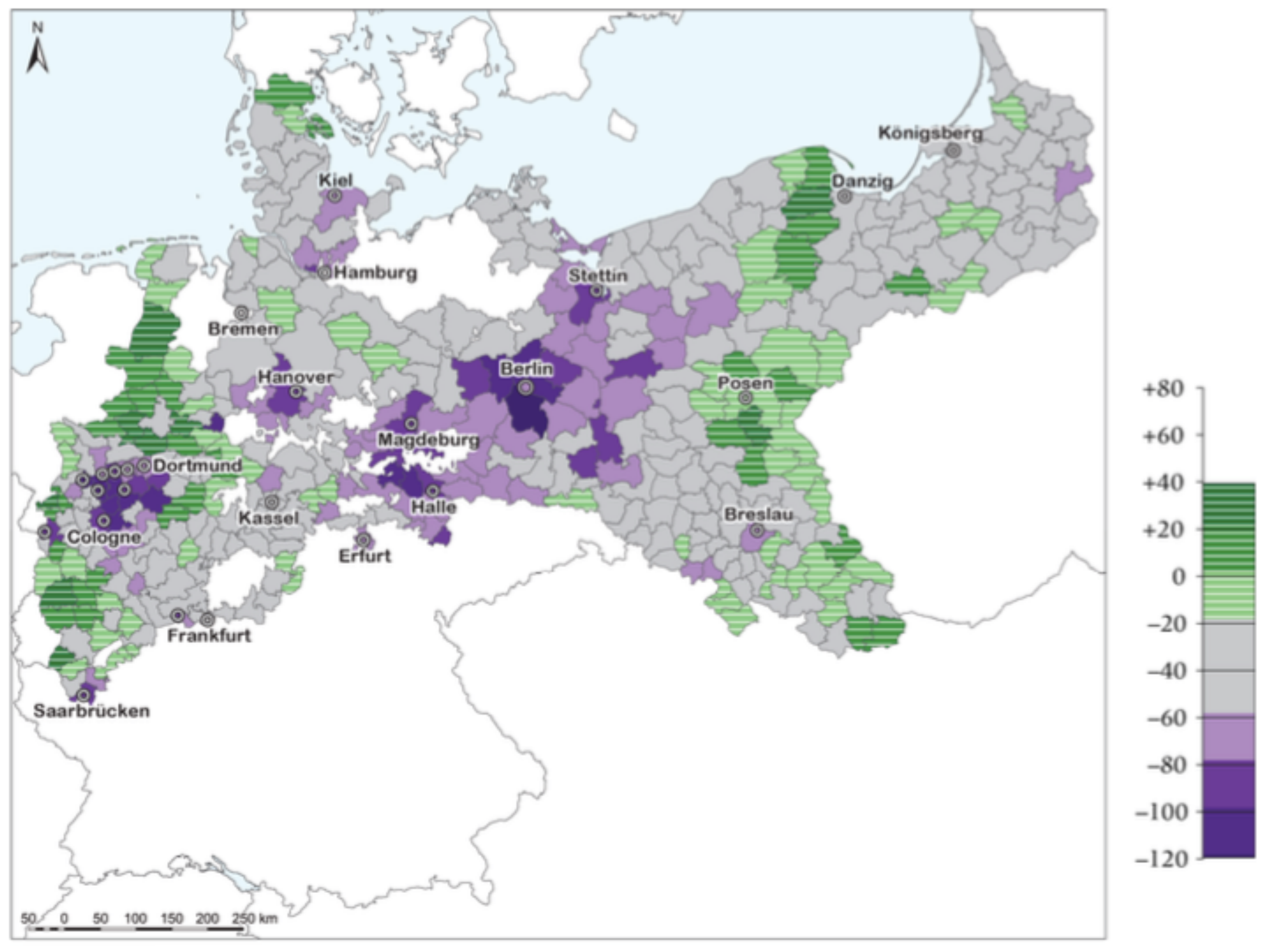



Total Fertility Rate
Year
0
1
2
3
4
5
6
7
8
1850
1900
1950
2000
Netherlands

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

WATKINS 1995

historical data



Spatial Analysis of the Causes of Fertility Decline in Prussia

Joshua R. Goldstein
Sebastian Klüsener

convenience samples

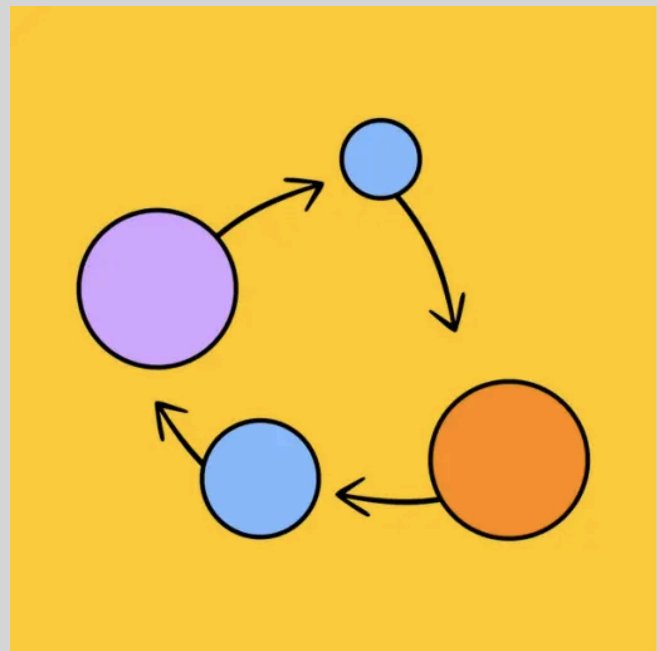
Does Fertility Behavior Spread among Friends?

Nicoletta Balbo[a] and Nicola Barban[b]

causal design

Family, Firms, and Fertility: A Study of Social Interaction Effects

Zafer Buyukkececi[1] · Thomas Leopold[2] · Ruben van Gaalen[3] · Henriette Engelhardt[4]



Channels of social influence on reproduction

LAURA BERNARDI

social learning
social contagion
social pressure
social support

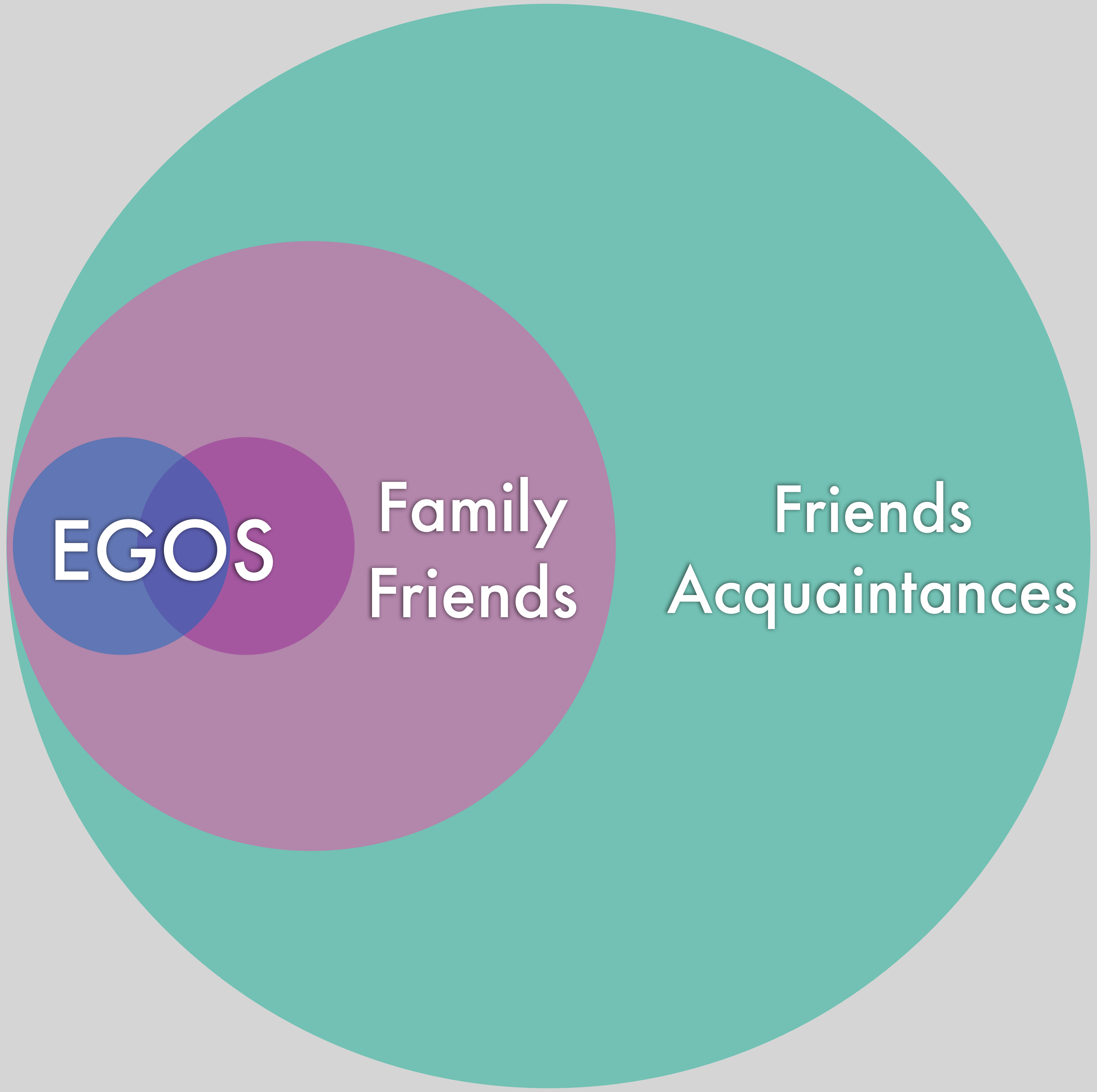
qualitative studies

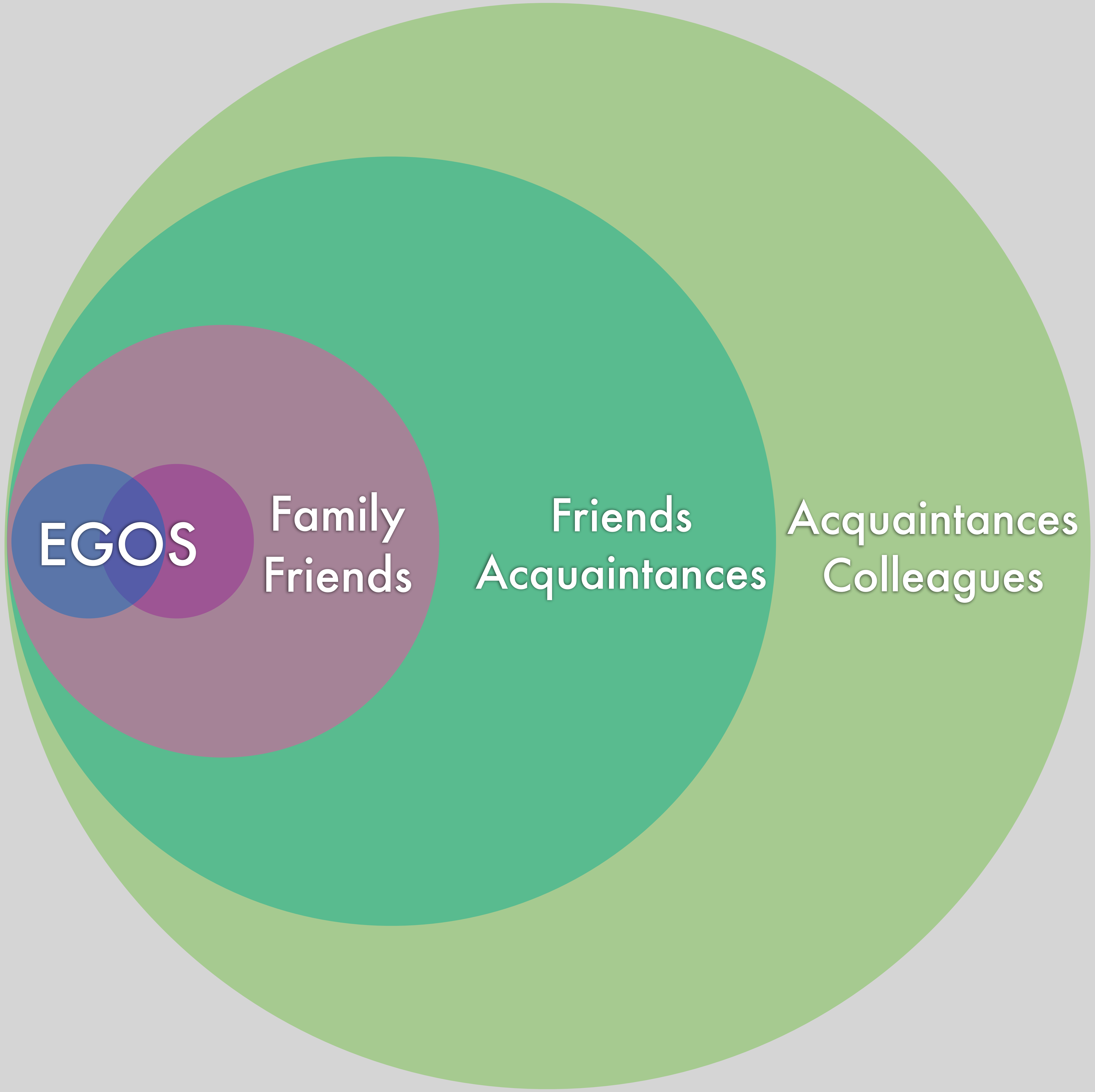
quantifying social influences
on fertility behaviour
using personal network data

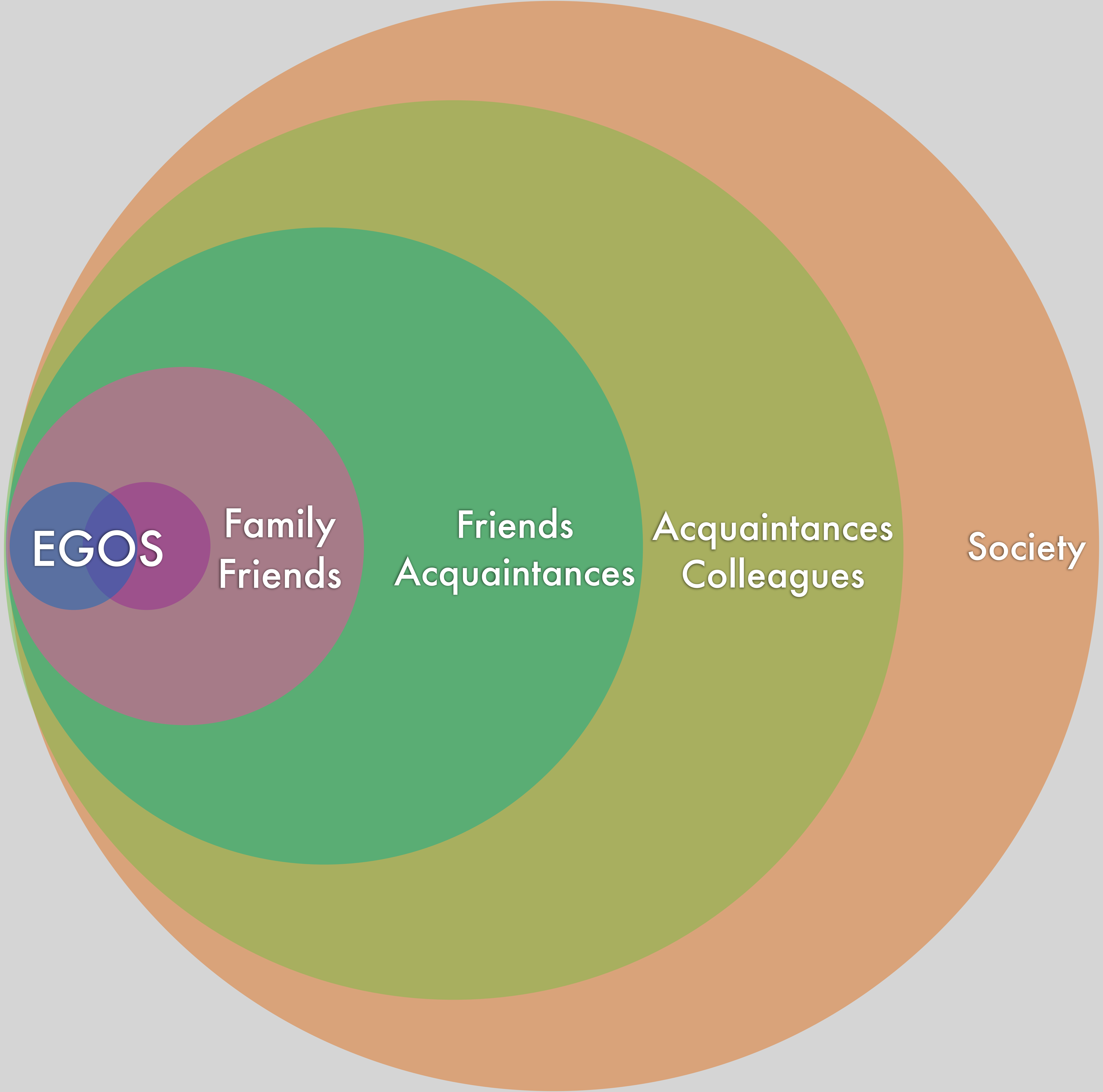
EGO

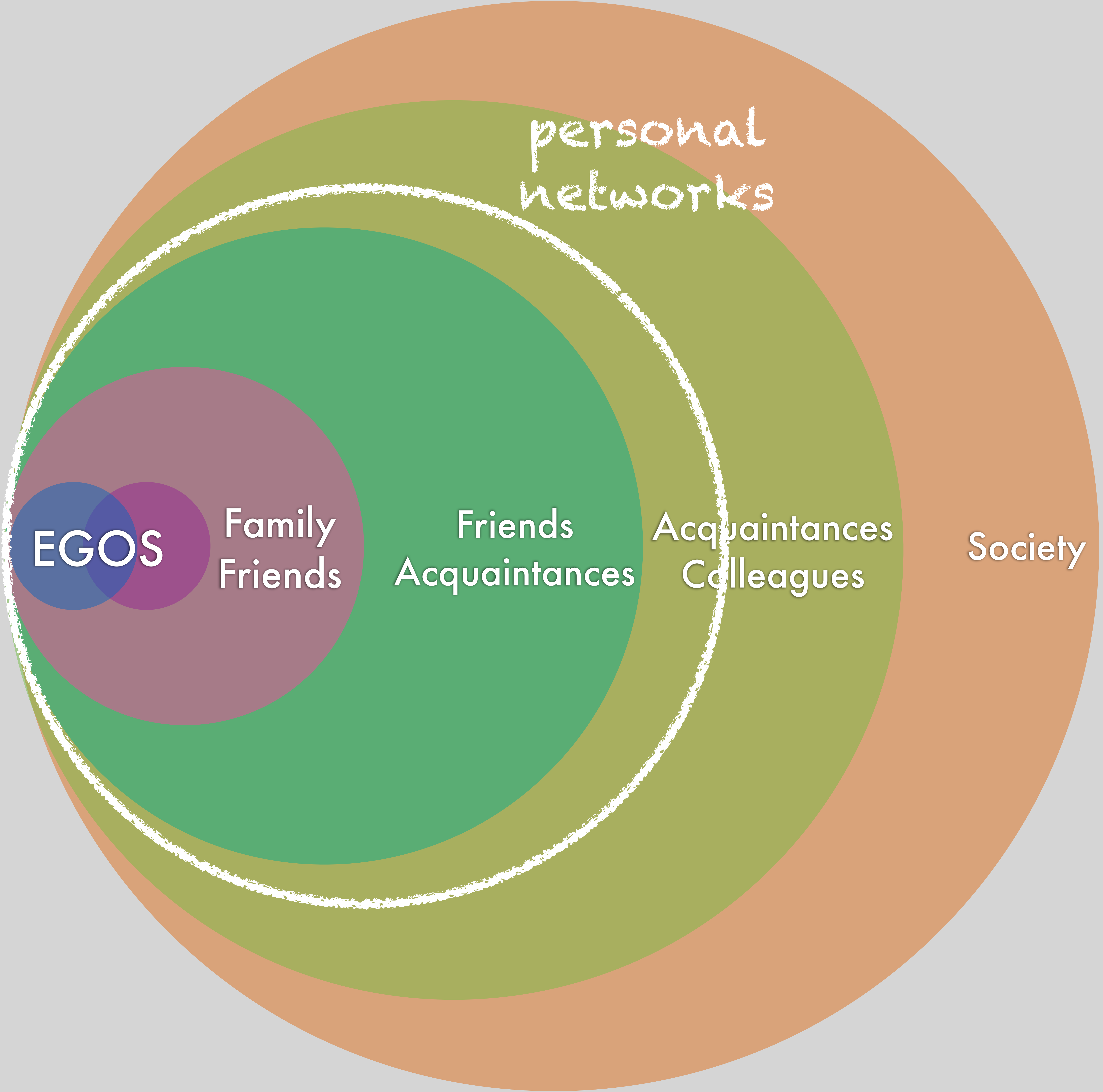
EGOS


EGOS
Family
Friends


EGOS
Family
Friends
Friends
Acquaintances

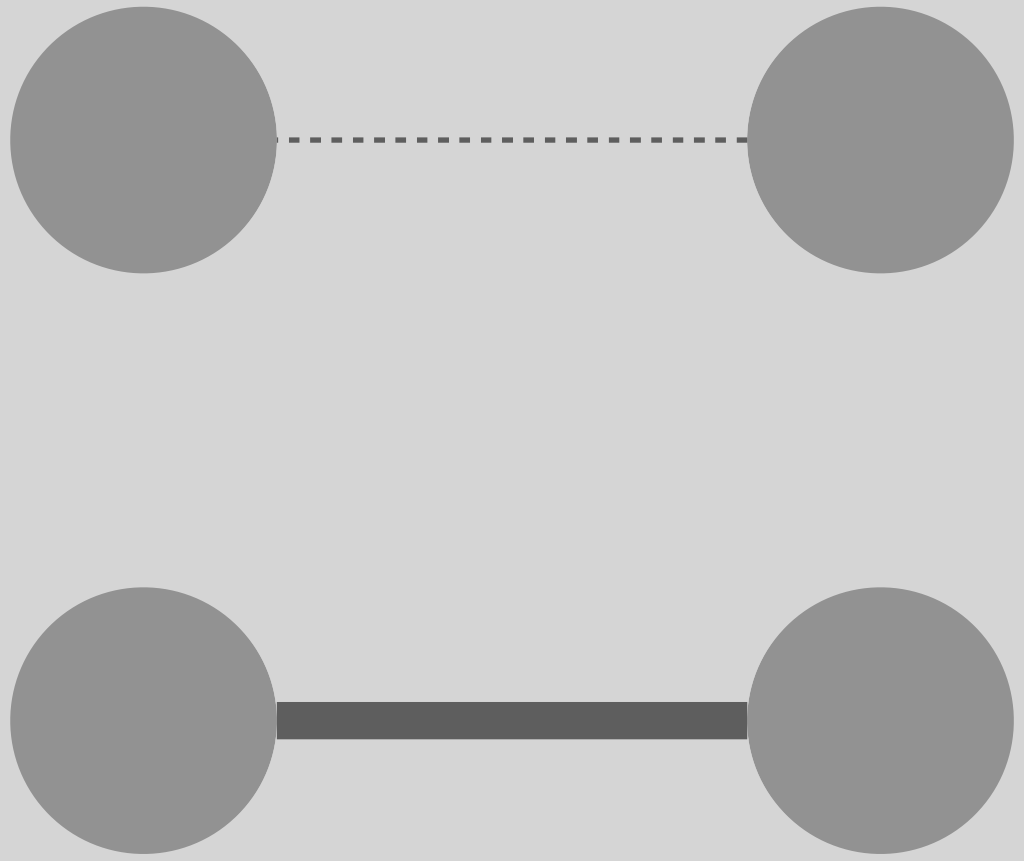

EGOS
Family
Friends
Friends
Acquaintances
Acquaintances
Colleagues


EGOS
Family
Friends
Friends
Acquaintances
Acquaintances
Colleagues
Society


personal
networks
EGOS
Family
Friends
Friends
Acquaintances
Acquaintances
Colleagues
Society

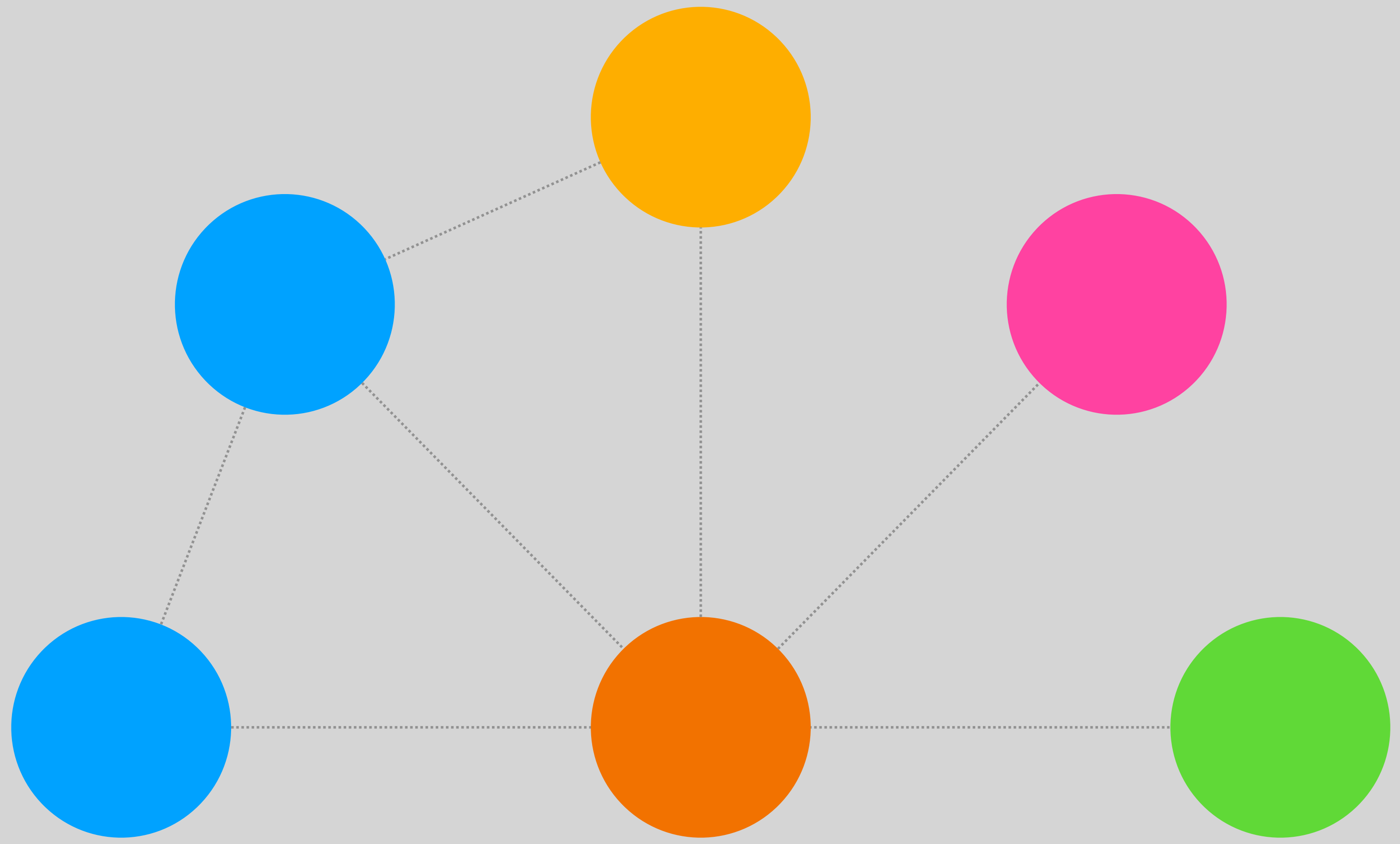
Personal Networks

tie (strength)



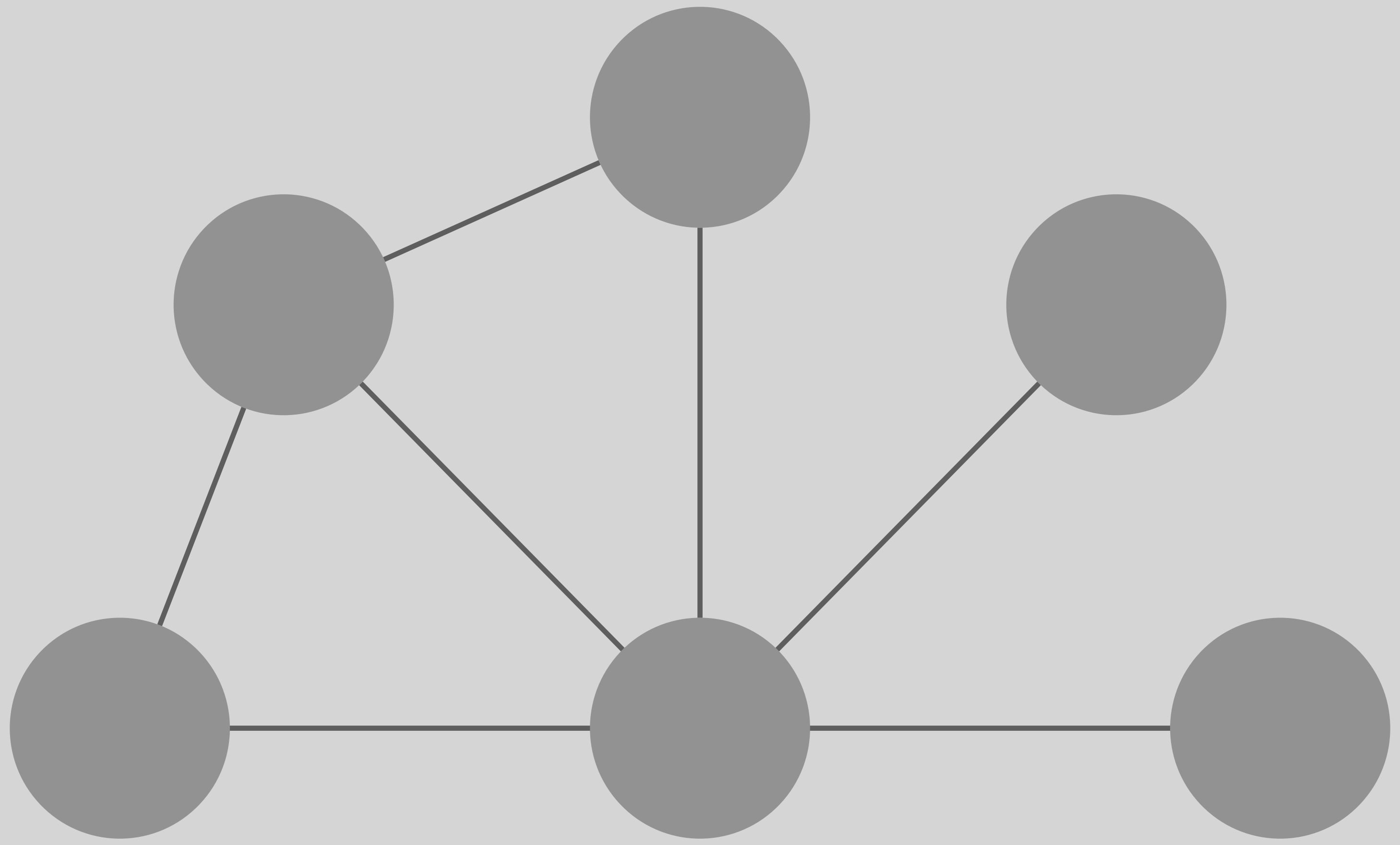
strong tie, more support/pressure
e.g., quality of relation with parent

composition



support network, diversity in ideas
e.g., # kin, # friends, # can help

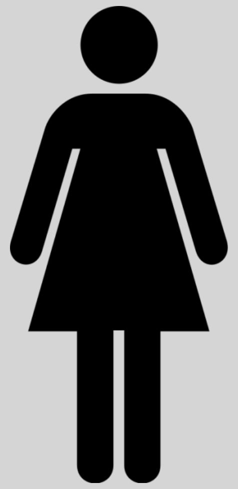
structure



reinforcing norms, flow information
e.g., density, # cliques

Methodology

Longitudinal Internet
Studies for the
Social sciences

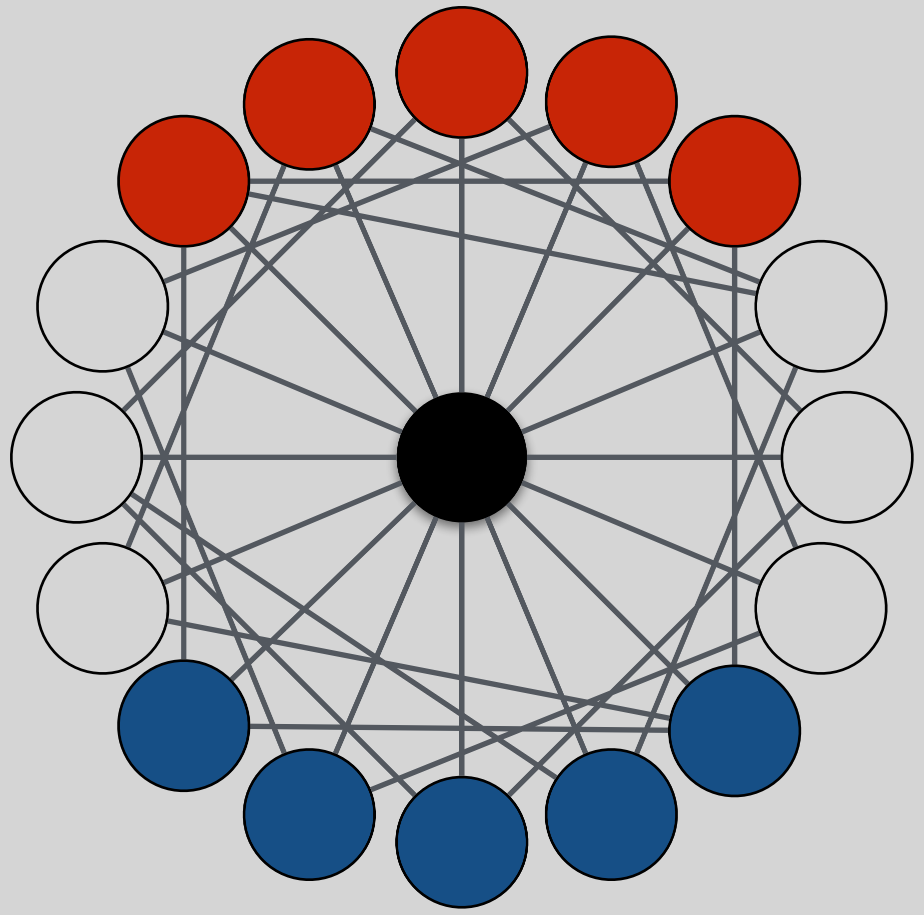


~750 women
age: 18 - 40

Ego

Age
Education
Income
Partnership status
Children
Detailed fertility preferences

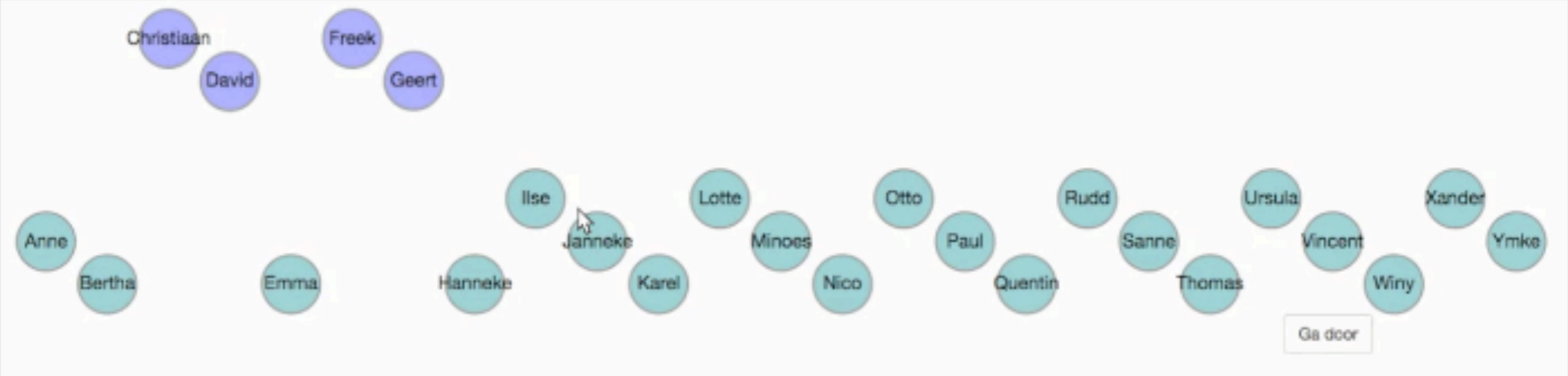
Alters (25)



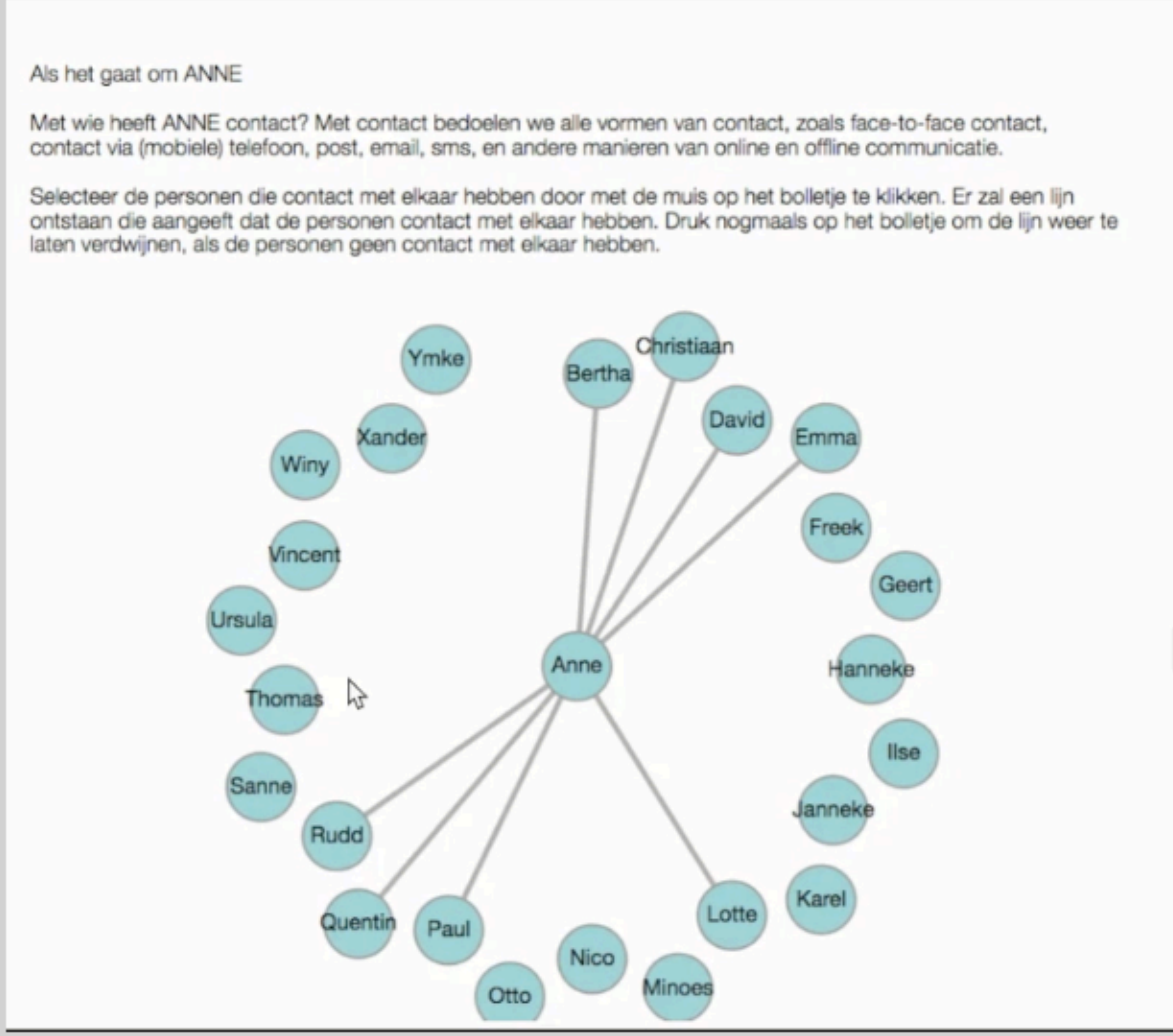
Sex
Age
Education
Relationship type
Closeness
Frequency of contact F2F
Frequency of other contact

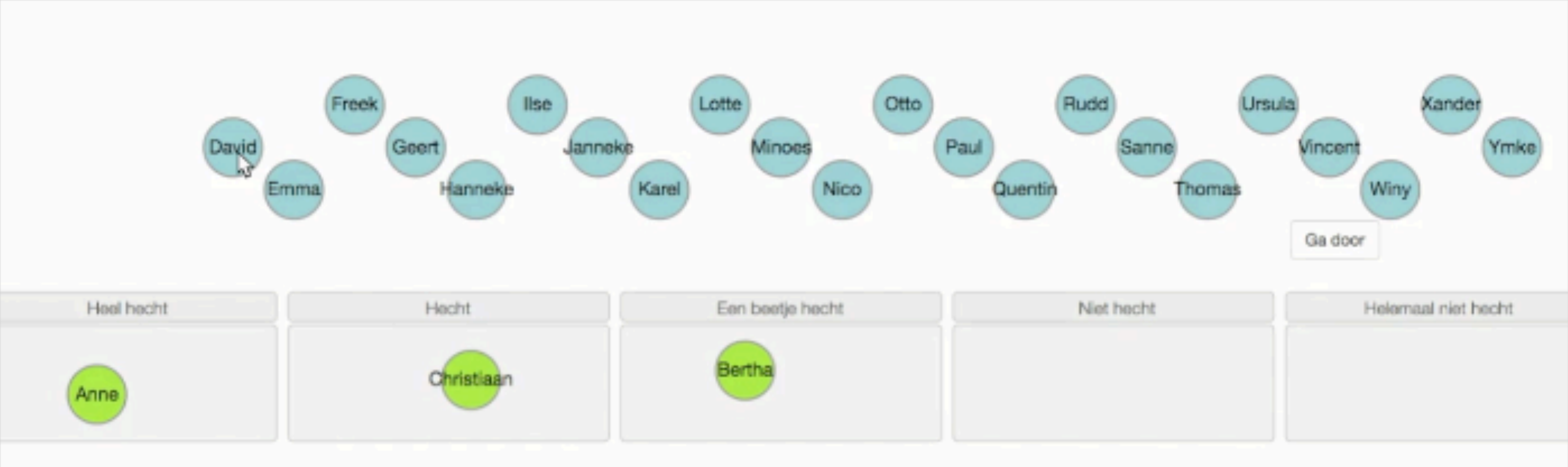
Number and age of children
Friend
Wants children
Does not want children
Help with children
Talk about children
Relationship with other alters

Methodology



Which of these 25 individuals could you ask for help with care for a child?





How close are you to these people?

Methodology

Longitudinal Internet
Studies for the
Social sciences

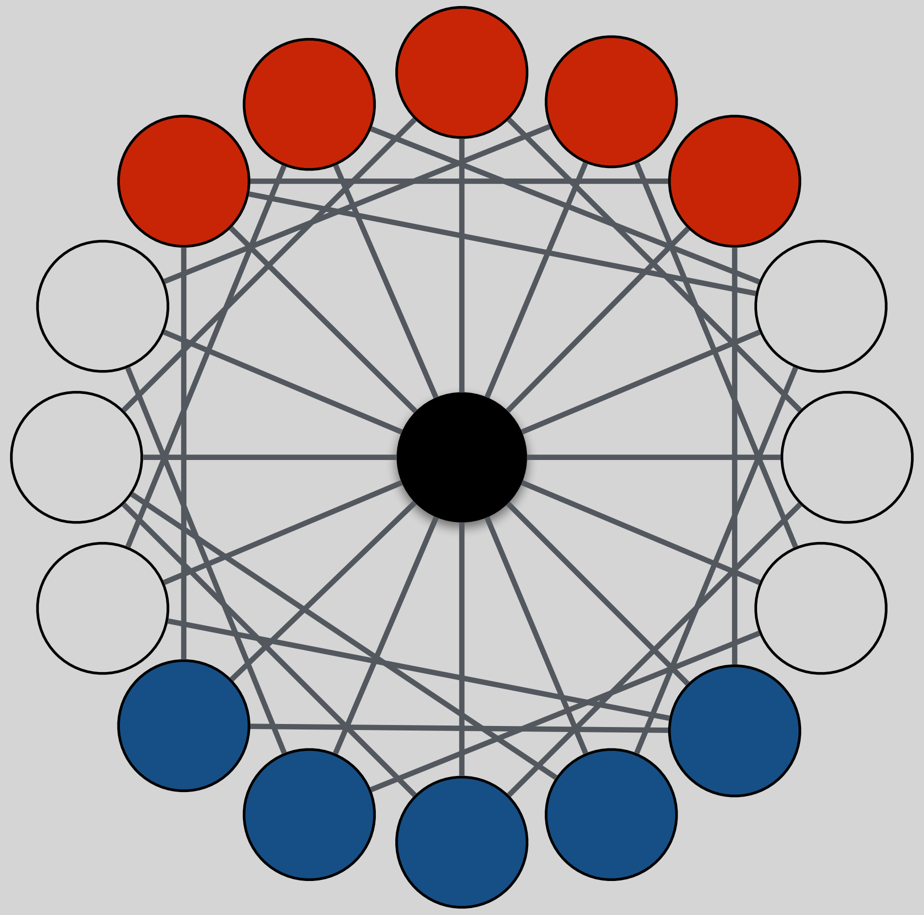
~750 women
age: 18 - 40

Ego

Age
Education
Income
Partnership status
Children
Detailed fertility preferences

OUTCOMES

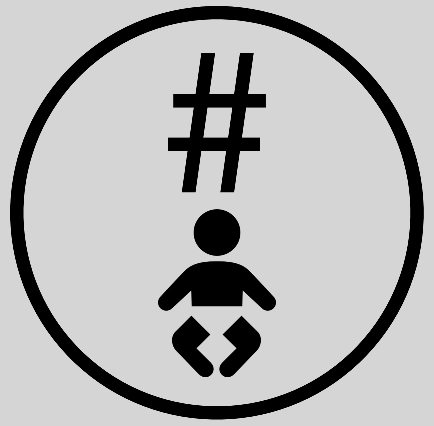
Alters (25)



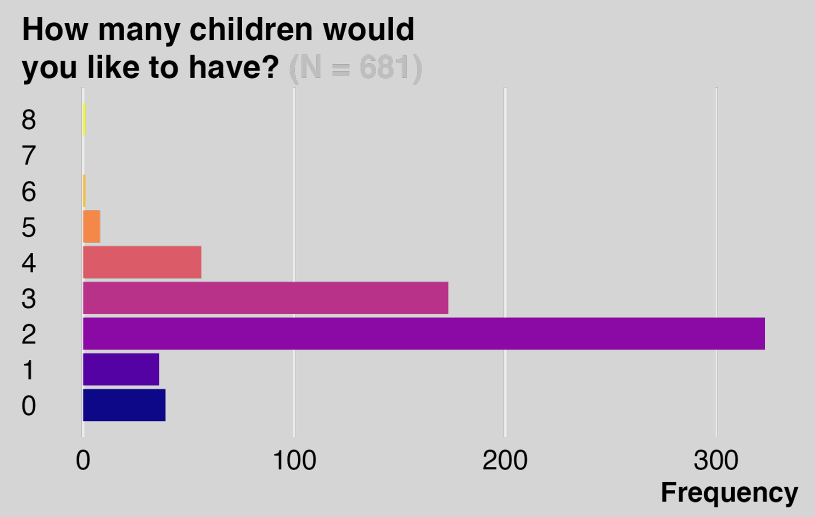
Sex
Age
Education
Relationship type
Closeness
Frequency of contact F2F
Frequency of other contact

Number and age of children
Friend
Wants children
Does not want children
Help with children
Talk about children
Relationship with other alters

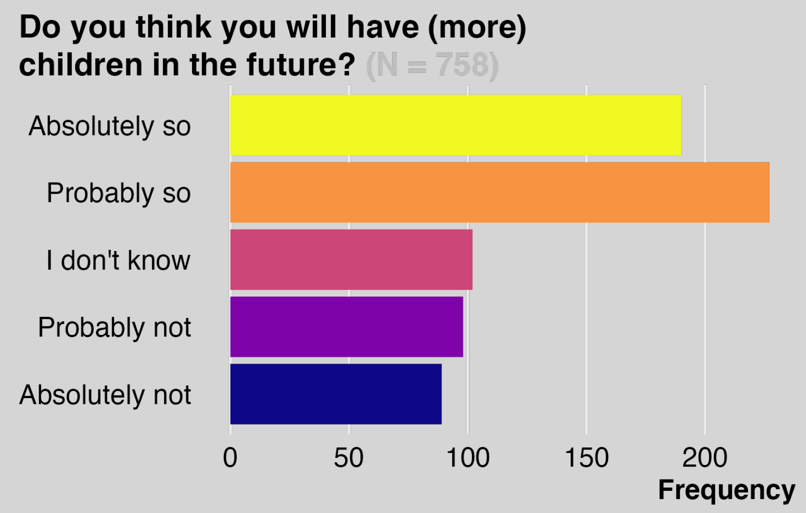
Outcomes



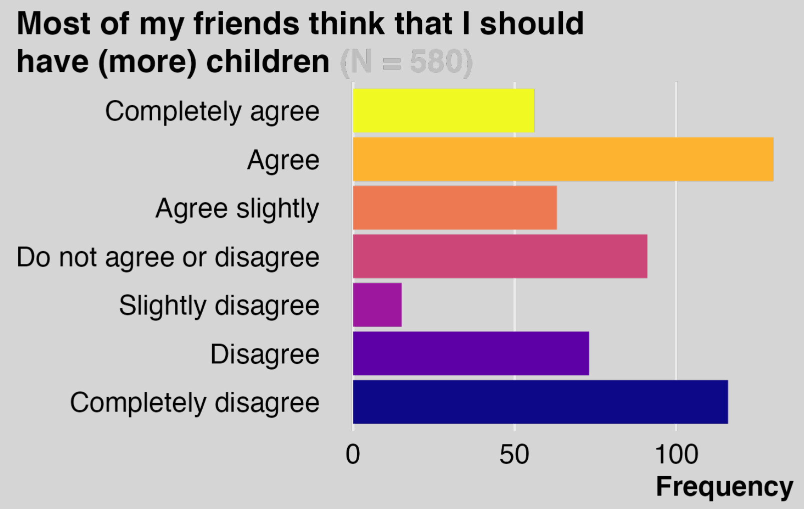
How many children would you like to have?



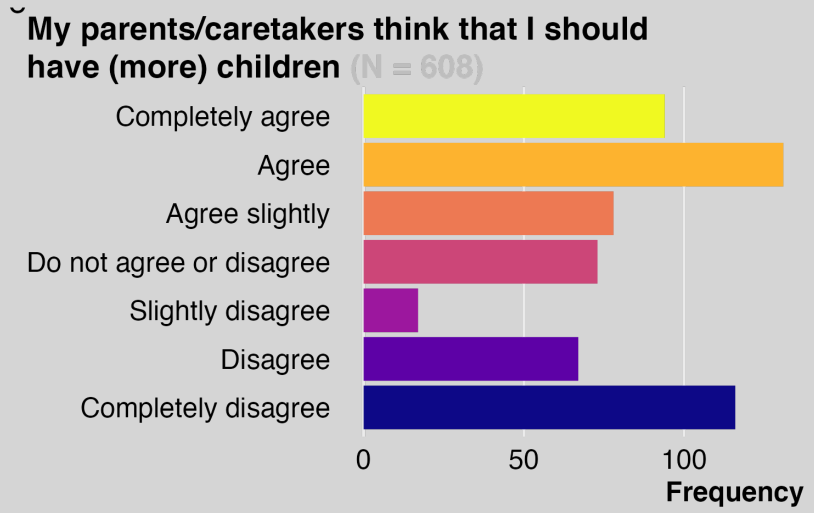
Do you think you will have (more) children in the future?



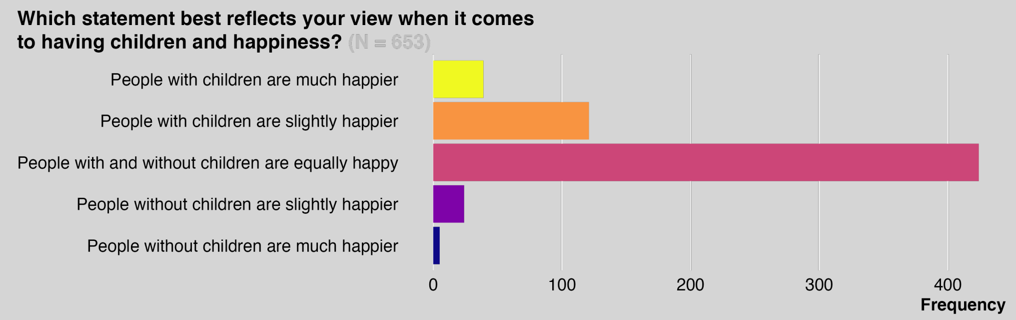
Perceived pressure to have children from friends



Perceived pressure to have children from parents/caretakers

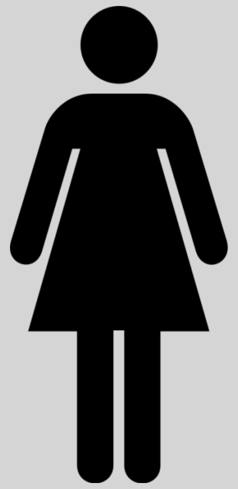


Do you think people with or without children are happier?



Methodology

Longitudinal Internet
Studies for the
Social sciences



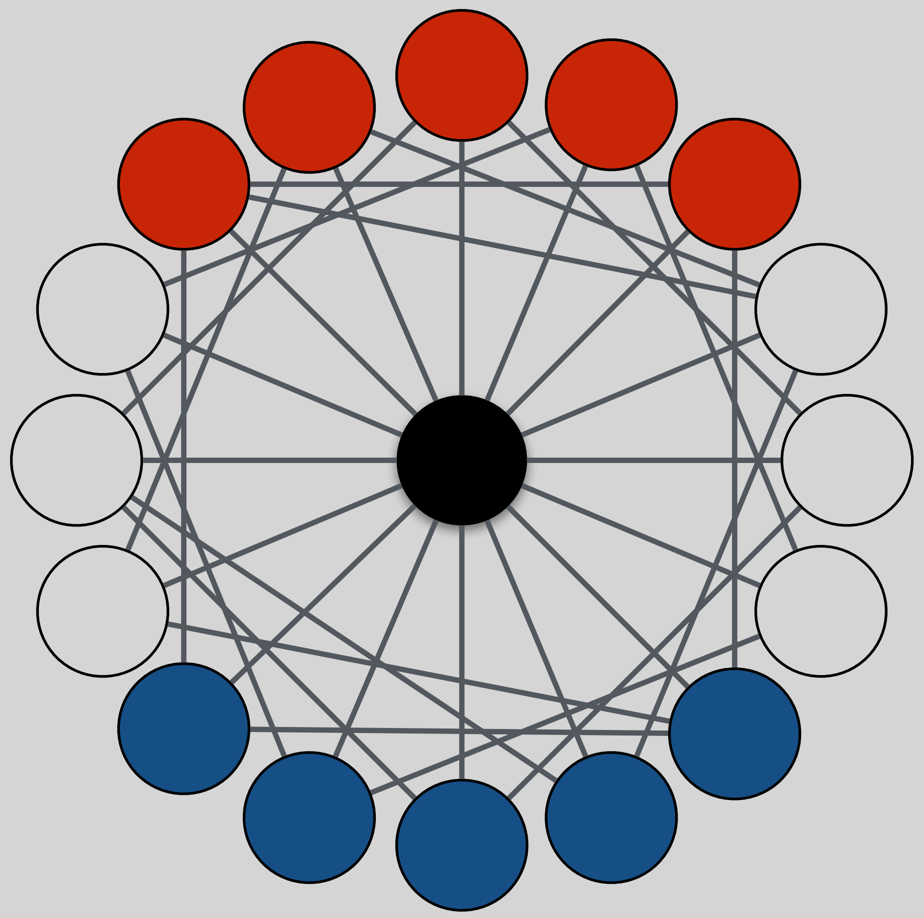
~750 women
age: 18 - 40

Ego

EGO VARIABLES

Age
Education
Income
Partnership status
Children

Alters (25)



NETWORK VARIABLES

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

Personal Networks

tie (strength)

average closeness
average f2f contact
average other contact

average closeness family
average closeness friends
average closeness childfree
…

24 variables

composition

% family
% friends
% childfree
% with children
% who want children
% childfree
% highly educated
% women
% can provide childcare
% can talk to about children

…

13 variables

structure

density
cliques
isolates and duos
communities
modularity
degree centralisation
betweenness centralisation
…

density among family
density among friends
density among childfree
…

20 variables

Personal Networks

tie (strength)

average closeness
average f2f contact
average other cont

average closeness f
average closeness f
average closeness c
…

24 variables

composition

% family
% friends

% can talk to about children

…

13 variables

structure

density
cliques
and duos
nities
y
ntralisation
ess centralisation

density among family
density among friends
density among childfree
…

20 variables

HOW TO CHOOSE WHICH VARIABLES TO FOCUS ON?

Animal Behaviour 168 (2020) 109–120

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

Is less more? A commentary on the practice of ‘metric hacking’ in animal social network analysis

Quinn M. R. Webber [a, *], David C. Schneider [a, b, c], Eric Vander Wal [a, c]

[a] Cognitive and Behavioural Ecology Interdisciplinary Program, Memorial University of Newfoundland, St John's, NL, Canada
[b] Department of Ocean Sciences, Ocean Sciences Centre, Memorial University of Newfoundland, St John's, NL, Canada
[c] Department of Biology, Memorial University of Newfoundland, St John's, NL, Canada

Animal Behaviour 168 (2020) 109–120

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Commentary

Is less more? A commentary on the practice of 'metric hacking' in animal social network analysis

Quinn M. R. Webber [a, *], David C.

[a] Cognitive and Behavioural Ecology Interdisciplinary Prog
[b] Department of Ocean Sciences, Ocean Sciences Centre, M
[c] Department of Biology, Memorial University of Newfoun

Check for updates

General Article

Psychological Science
22(11) 1359–1366

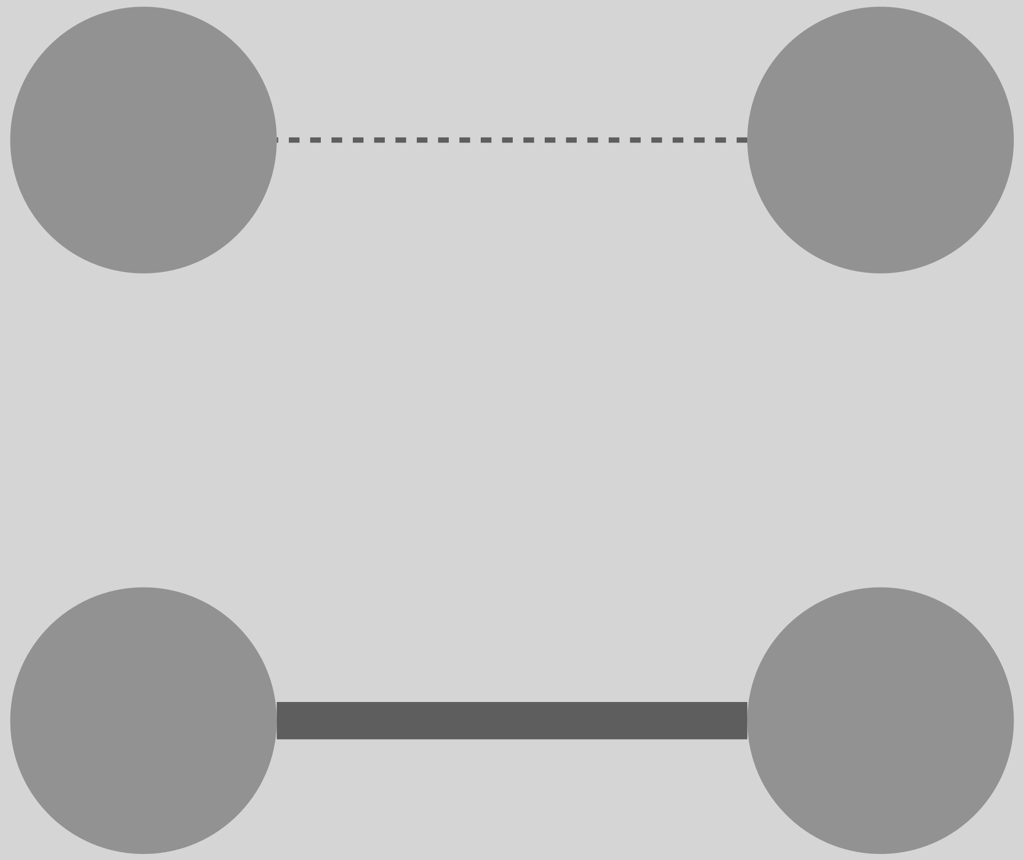
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Personal Networks

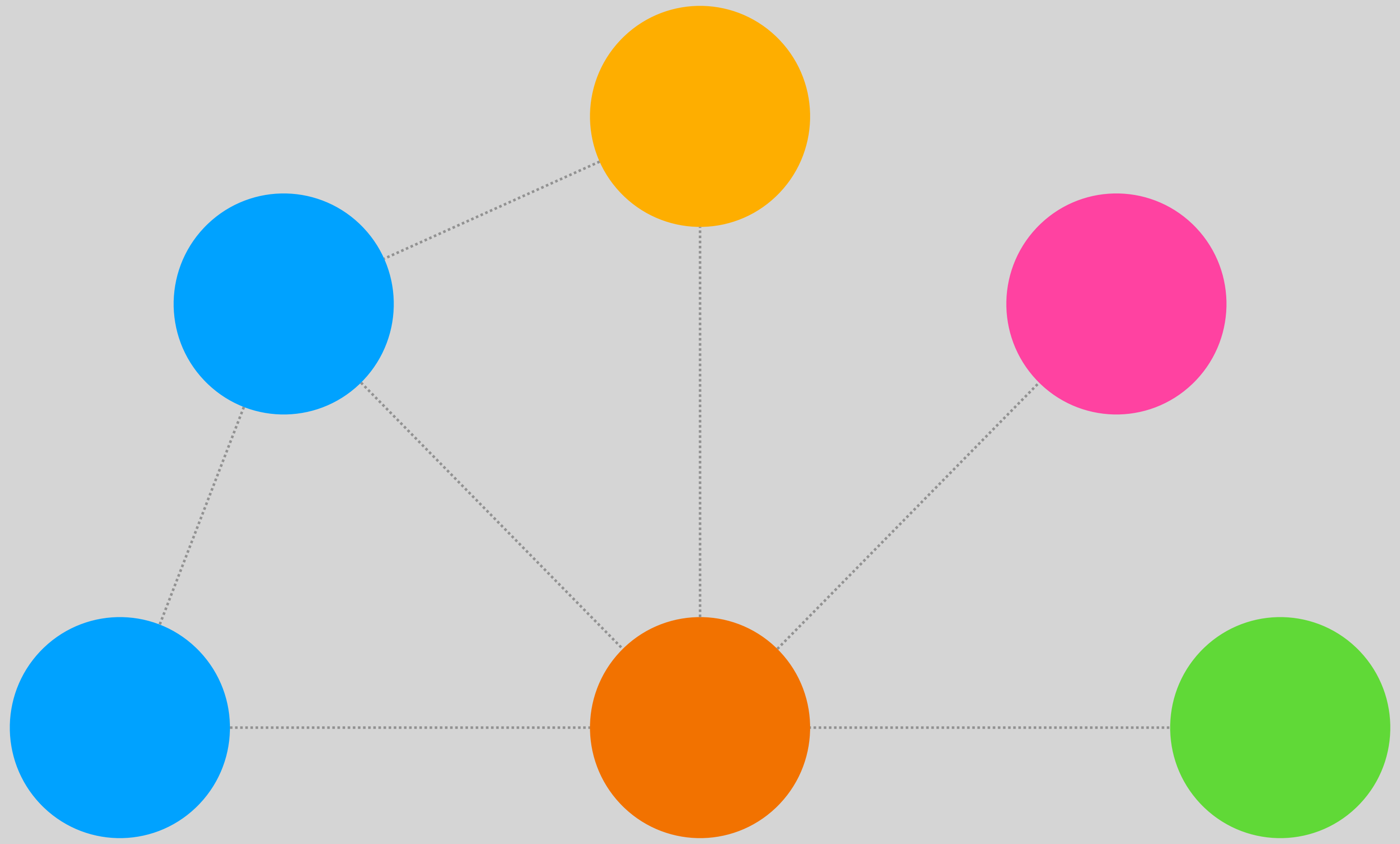
tie (strength)



strong tie, more support/pressure

e.g., quality of relation with parent

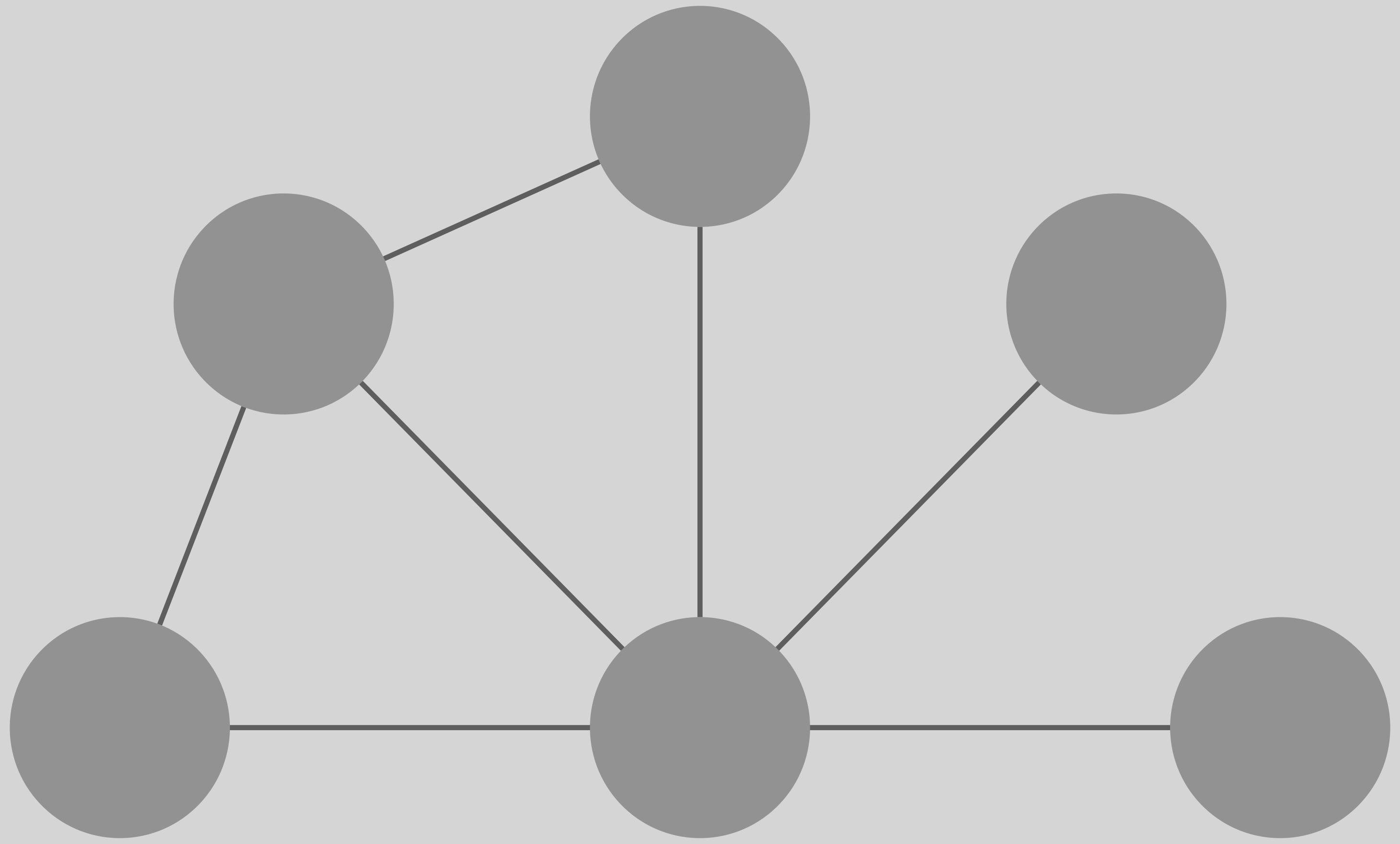
composition



support network, diversity in ideas

e.g., # kin, # friends, # can help

structure



reinforcing norms, flow information

e.g., density, # cliques

Lasso Regression

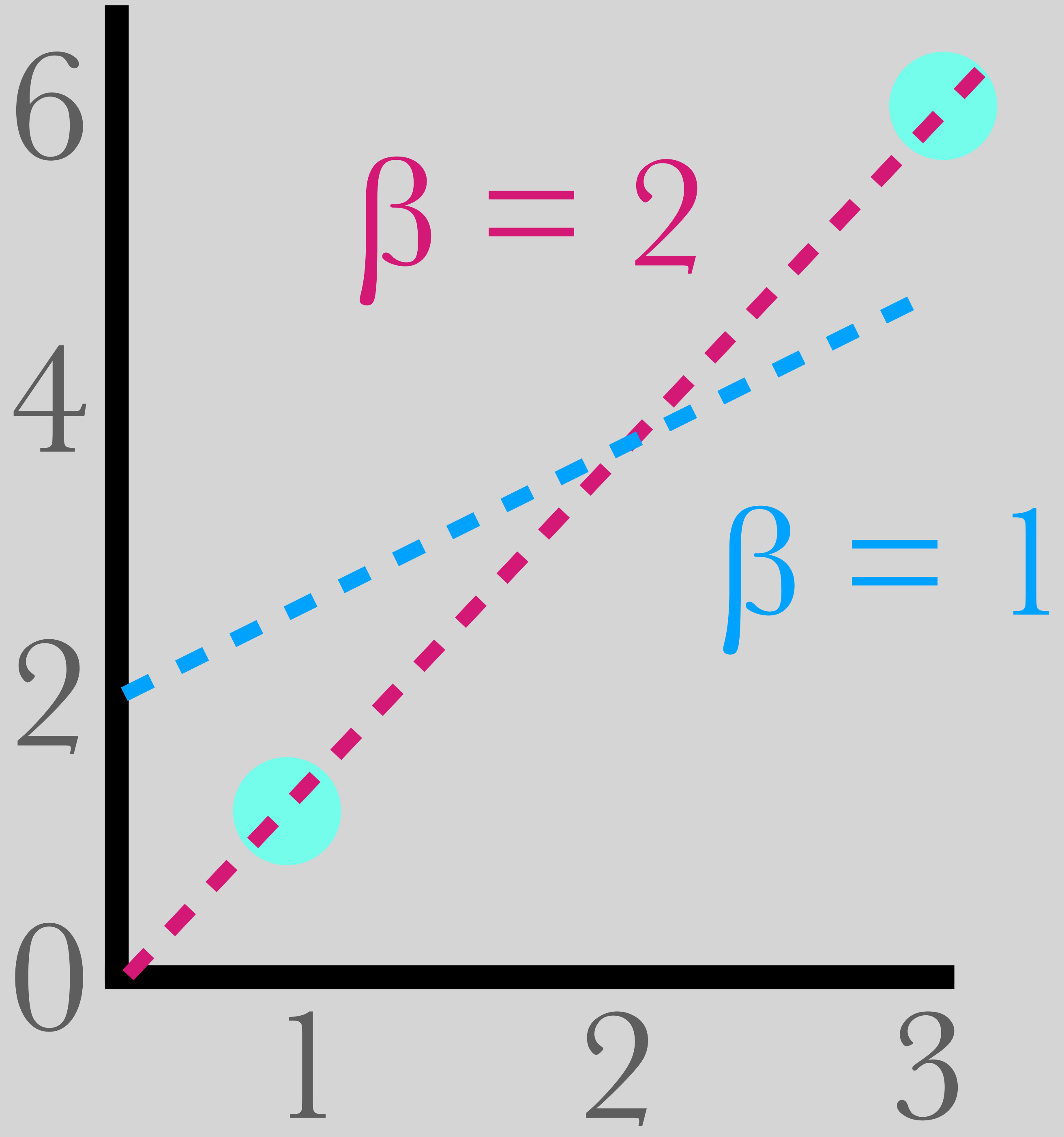
$$\underbrace{\sum_{i=0}^{n}(y_i - \hat{y_i})^2}_{\text{linear regression}} + \underbrace{\lambda \sum_{i=1}^{p}|\beta_j|}_{\text{penalty term}}$$

- ✅ can handle many, correlated variables
- ✅ leads to sparse, predictive, interpretable models
- ❌ reduced variance through increased bias

Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

assume $\lambda = 6$



Linear regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$

LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|2| = 0 + 12 = 12$$

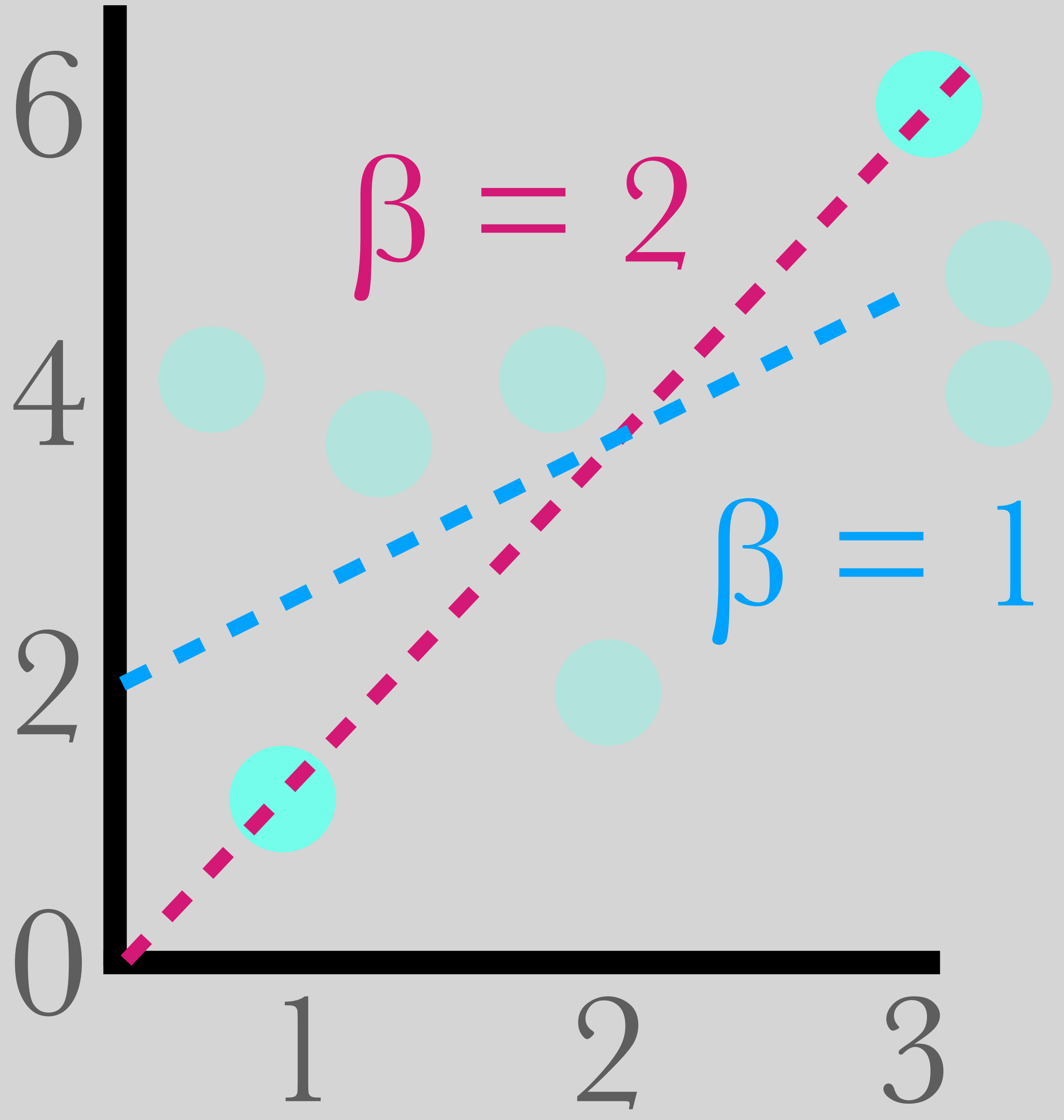
LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|1| = 2^2 + 1^2 + 6 = 11$$

Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

assume $\lambda = 6$



Linear regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$

LASSO regression

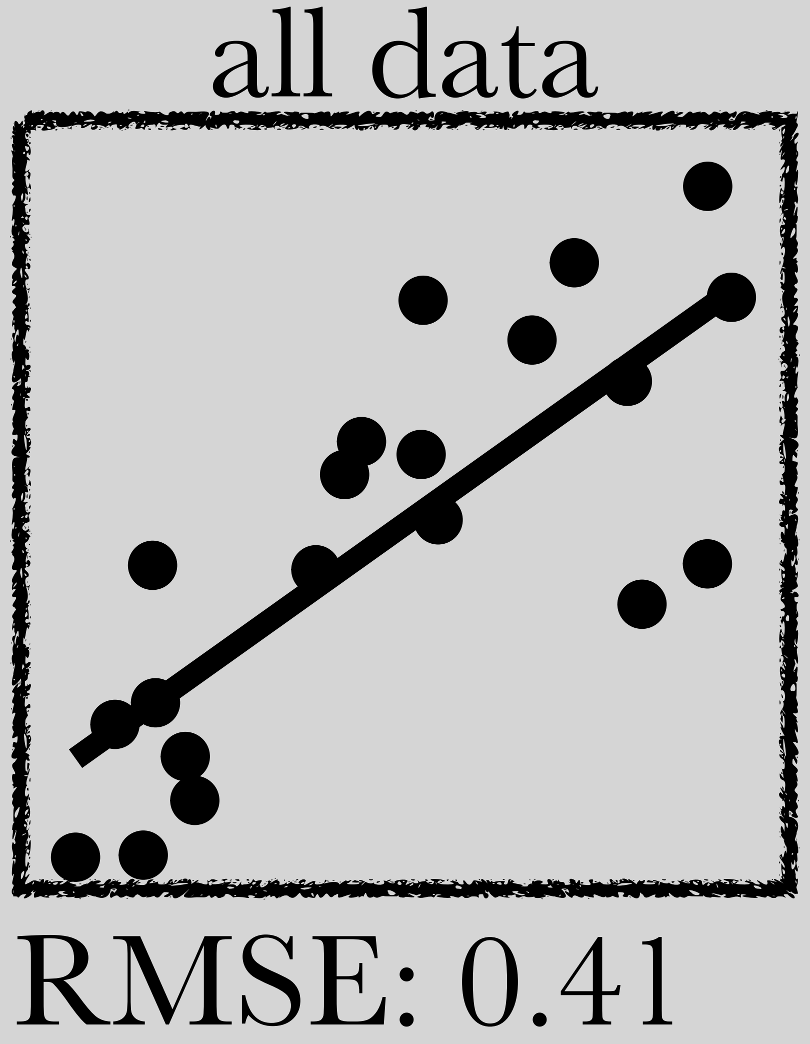
$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|2| = 0 + 12 = 12$$

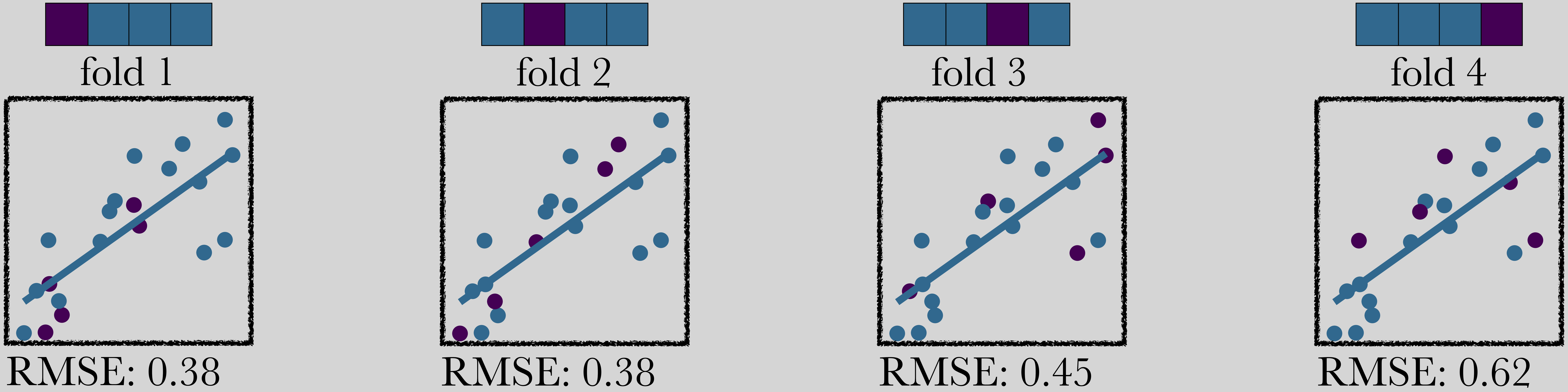
LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|1| = 2^2 + 1^2 + 6 = 11$$

Cross-Validation

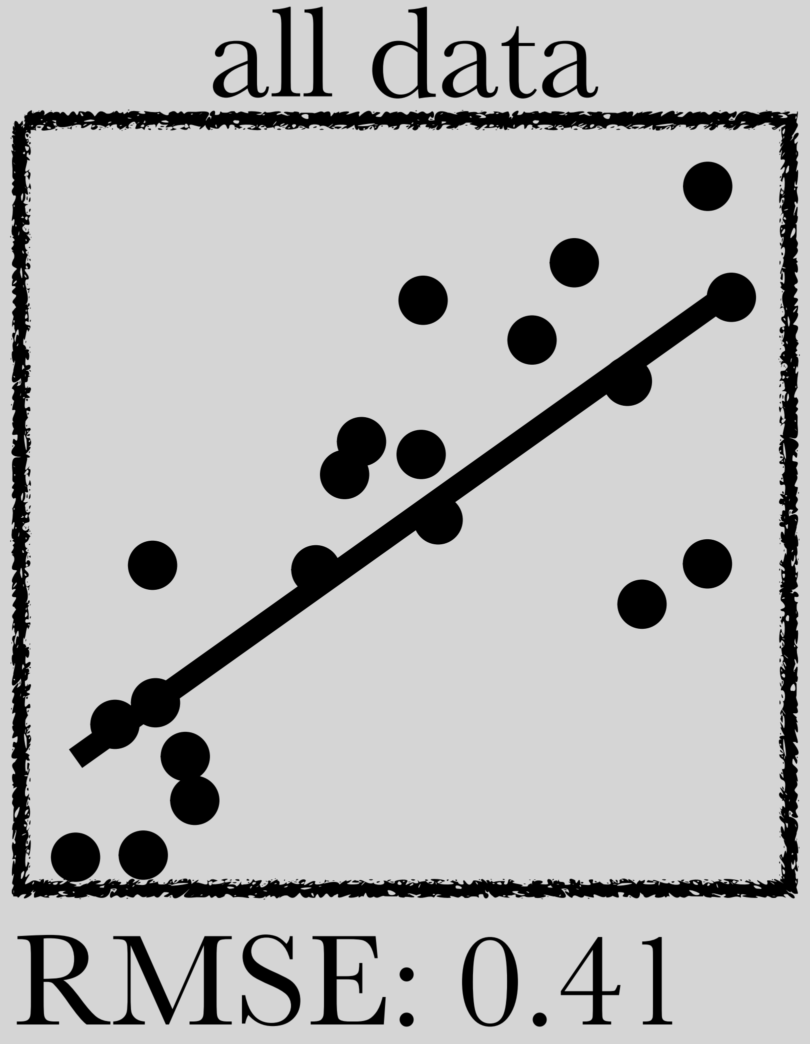
λ is determined through cross-validation and **out-of-sample predictive ability**

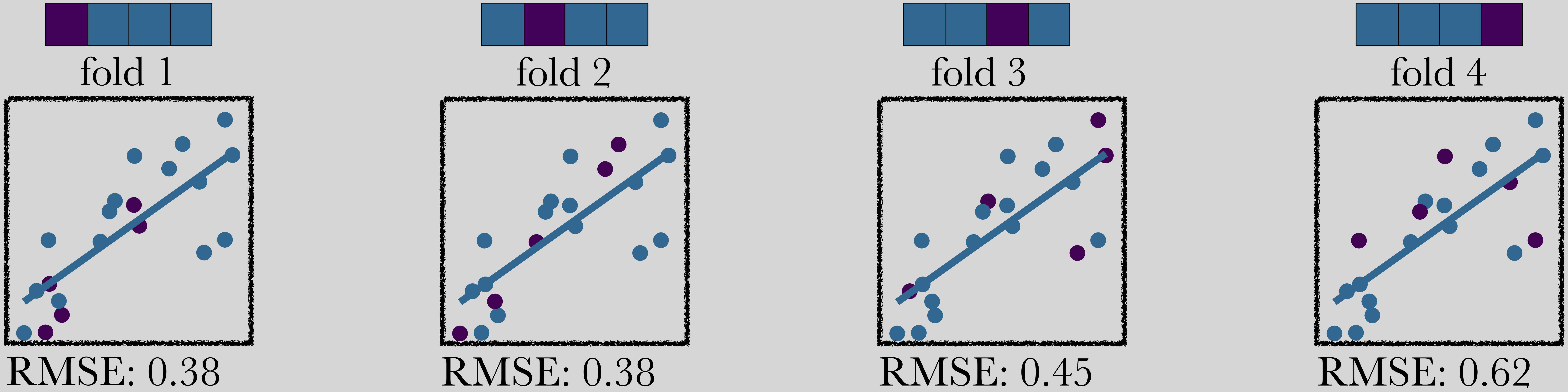




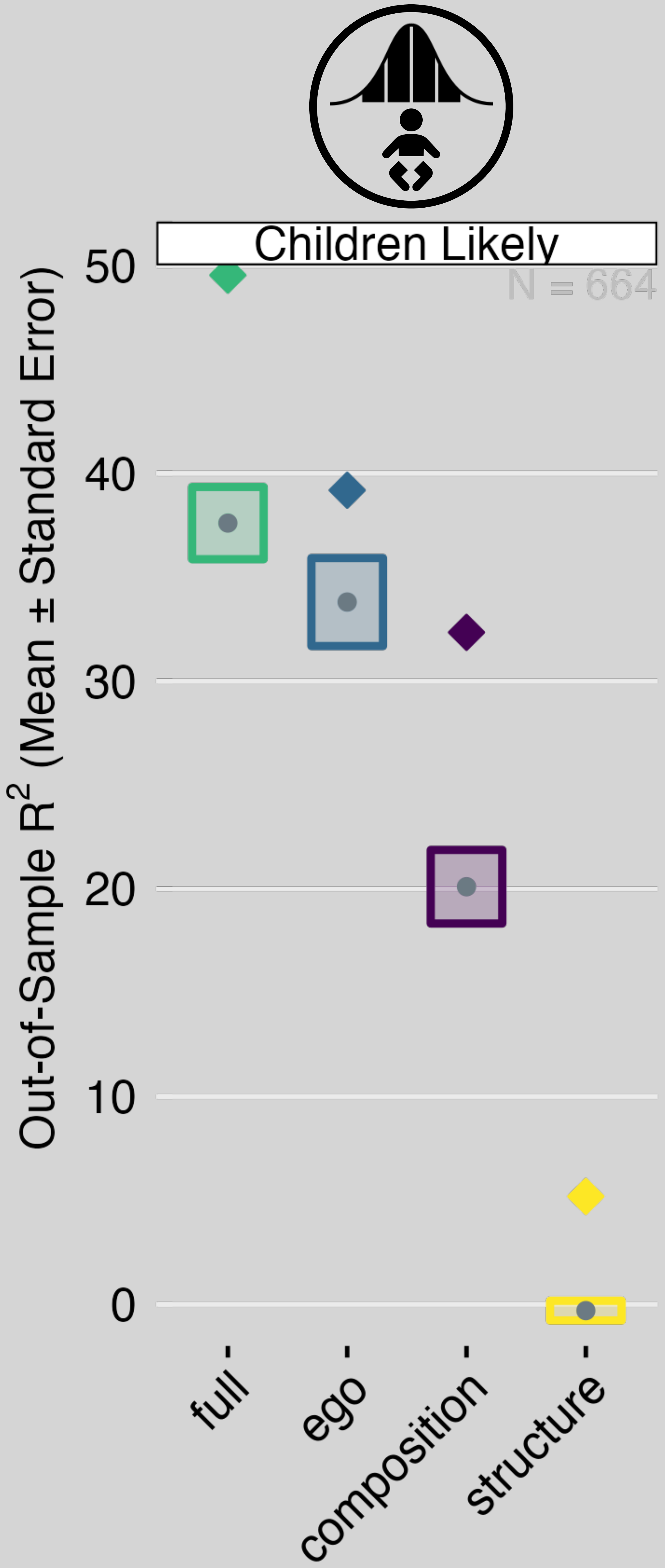
Cross-Validation

strength of model determined through cross-validation and **quantified by out-of-sample predictive ability**





Results


Children Likely
N = 664
Out-of-Sample R2 (Mean ± Standard Error)
50
40
30
20
10
0
full
ego
composition
structure

Happiness
N = 581
Pressure Friends
N = 515
Pressure Parents
N = 546
Ideal # Children
N = 600
Children Likely
N = 664
Out-of-Sample R2 (Mean ± Standard Error)
0
10
20
30
40
50
full
ego
composition
structure

Take-Home Messages

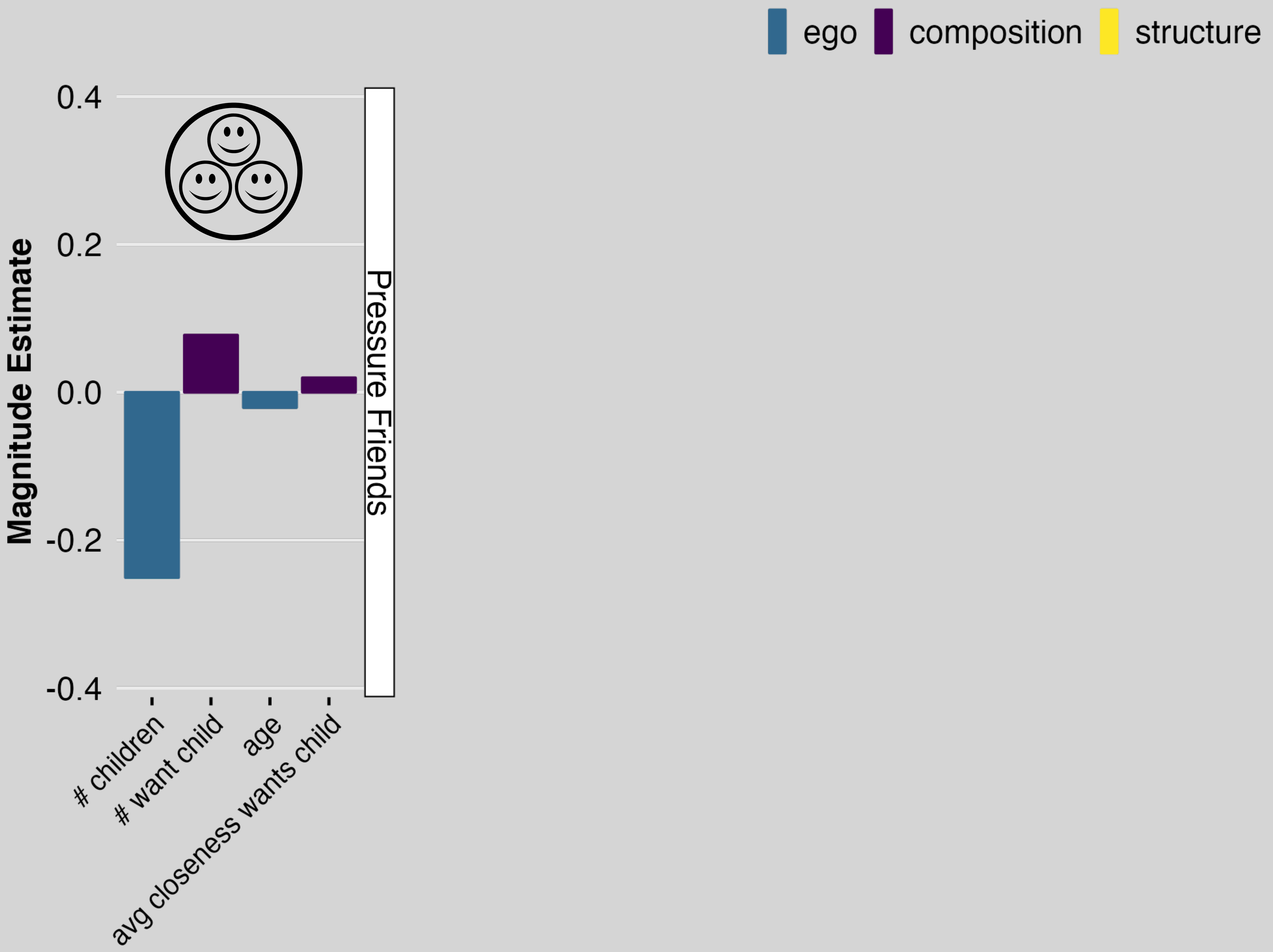
✅ predicting pretty well!

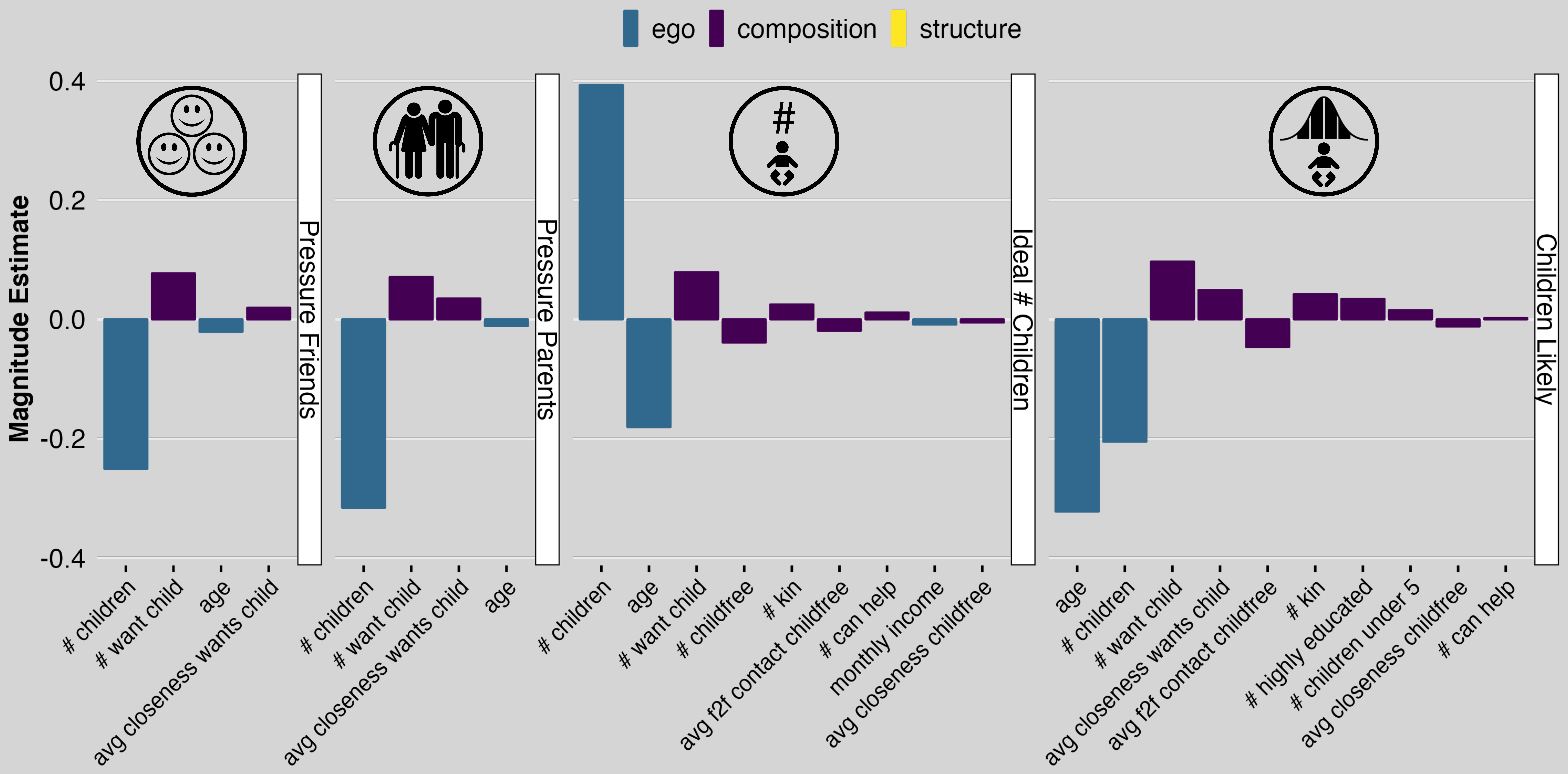
Happiness
N = 581
Pressure Friends
N = 515
Pressure Parents
N = 546
Ideal # Children
N = 600
Children Likely
N = 664
#
Out-of-Sample R2 (Mean ± Standard Error)
50
40
30
20
10
0
full
ego
composition
structure

Take-Home Messages

✓ predicting pretty well!

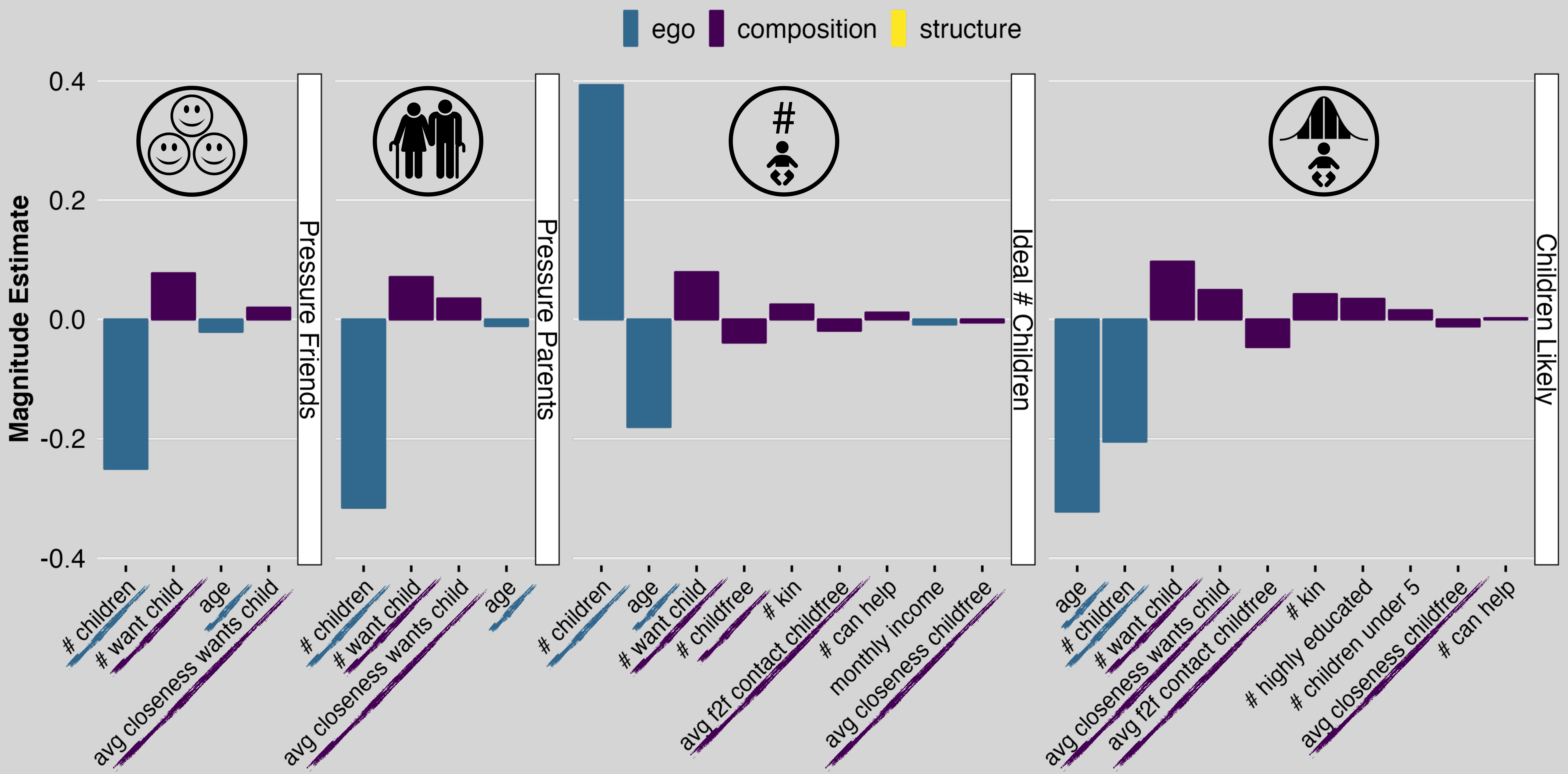
(✗) massive overfitting (~15 %-points)


ego
composition
structure
Magnitude Estimate
0.4
0.2
0.0
-0.2
-0.4
Pressure Friends
children
want child
age
avg closeness wants child


ego
composition
structure
Magnitude Estimate
0.4
0.2
0.0
-0.2
-0.4
Pressure Friends
Pressure Parents
Ideal # Children
Children Likely
#
children
want child
age
avg closeness wants child
children
want child
avg closeness wants child
age
children
age
want child
childfree
kin
avg f2f contact childfree
can help
monthly income
avg closeness childfree
age
children
want child
avg closeness wants child
avg f2f contact childfree
kin
highly educated
children under 5
avg closeness childfree
can help

Take-Home Messages

- ✓ predicting pretty well!
- (✗) massive overfitting (~15 %-points)
- ✓ personal variables important, composition so-so, structure not



ego
composition
structure
Magnitude Estimate
0.4
0.2
0.0
-0.2
-0.4
Pressure Friends
children
want child
age
avg closeness wants child
Pressure Parents
children
want child
avg closeness wants child
age
#
Ideal # Children
children
age
want child
childfree
kin
avg f2f contact childfree
can help
monthly income
avg closeness childfree
Children Likely
age
children
want child
avg closeness wants child
avg f2f contact childfree
kin
highly educated
children under 5
avg closeness childfree
can help

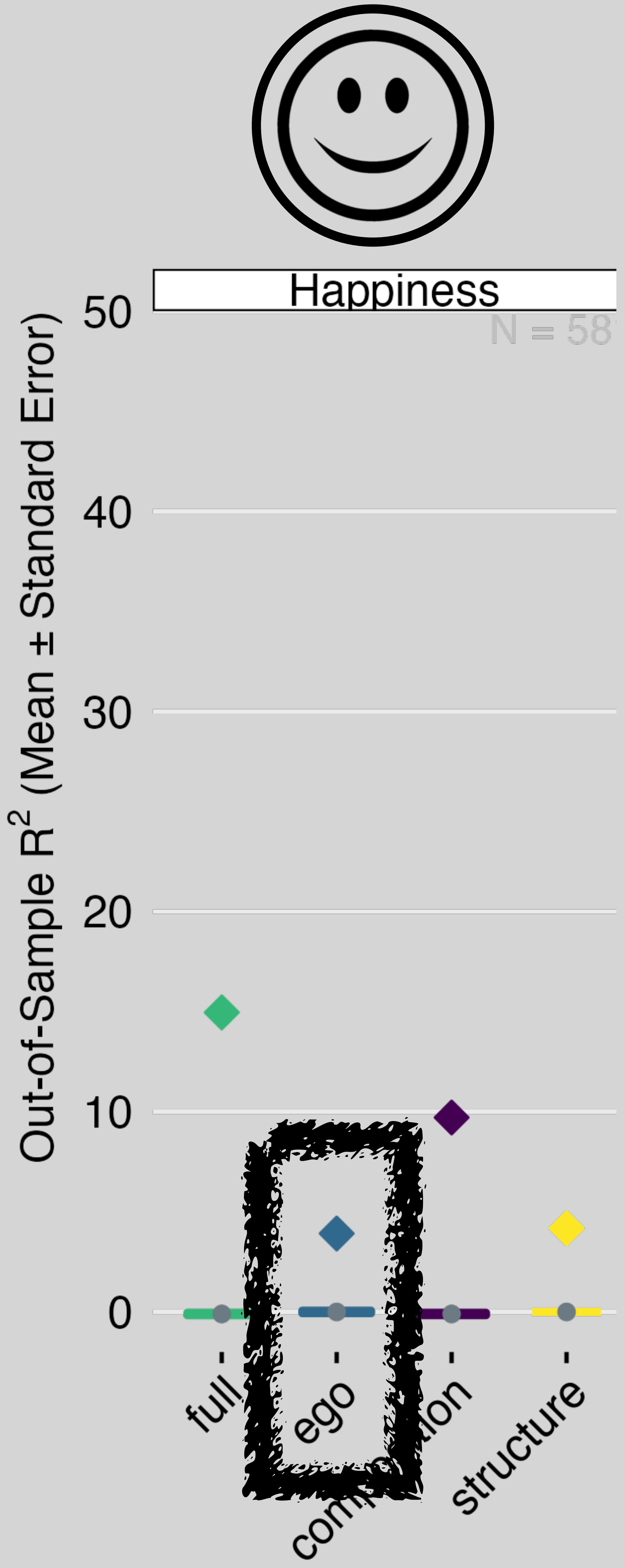
Important Variables

- age
- # children
- # people who **do** want children
- # people who **do not** want children
- strength of relationship to these people

Take-Home Messages

- ✓ predicting pretty well!
- (✗) massive overfitting (~15 %-points)
- ✓ personal variables important, composition so-so, structure not
- ✓ people who want children and who do not important

Take-Home Messages



- ✓ predicting pretty well!
  - difficult to assess how well
- ✗ massive overfitting (~15 %-points)
  - potentially misleading conclusions
- ✓ personal variables important, composition so-so, structure not
  - networks may not be unimportant, few ego variables
- ✓ people who want children and who do not important
  - understudied

R package FertNet

FertNet: Process Data from the Social Networks and Fertility Survey

Processes data from The Social Networks and Fertility Survey, downloaded from <https://dataarchive.lissdata.nl>, including correcting respondent errors and transforming network data into network objects to facilitate analyses and visualisation.

Version:	0.1.1
Imports:	haven (≥ 2.5.1)
Suggests:	testthat (≥ 3.0.0), tidygraph (≥ 1.2.2)
Published:	2023-03-16
Author:	Stulp Gert [aut, cre]
Maintainer:	Stulp Gert <g.stulp at rug.nl>
License:	CC BY 4.0
NeedsCompilation:	no
Materials:	README NEWS
CRAN checks:	FertNet results

Documentation:

Reference manual:	FertNet.pdf

Downloads:

Package source:	FertNet_0.1.1.tar.gz
Windows binaries:	r-devel: FertNet_0.1.1.zip, r-release: FertNet_0.1.1.zip, r-oldrel: FertNet_0.1.1.zip
macOS binaries:	r-release (arm64): FertNet_0.1.1.tgz, r-oldrel (arm64): FertNet_0.1.1.tgz, r-release (x86_64): FertNet_0.1.1.tgz, r-oldrel (x86_64): FertNet_0.1.1.tgz

Linking:

Please use the canonical form `https://CRAN.R-project.org/package=FertNet` to link to this page.

DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 49, ARTICLE 19, PAGES 493–512
PUBLISHED 8 SEPTEMBER 2023
https://www.demographic-research.org/Volumes/Vol49/19/
DOI: 10.4054/DemRes.2023.49.19

Data Description

Describing the Dutch Social Networks and Fertility Study and how to process it

Gert Stulp

"A complicated data-mining exercise, with much oversold results"

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

Check for updates

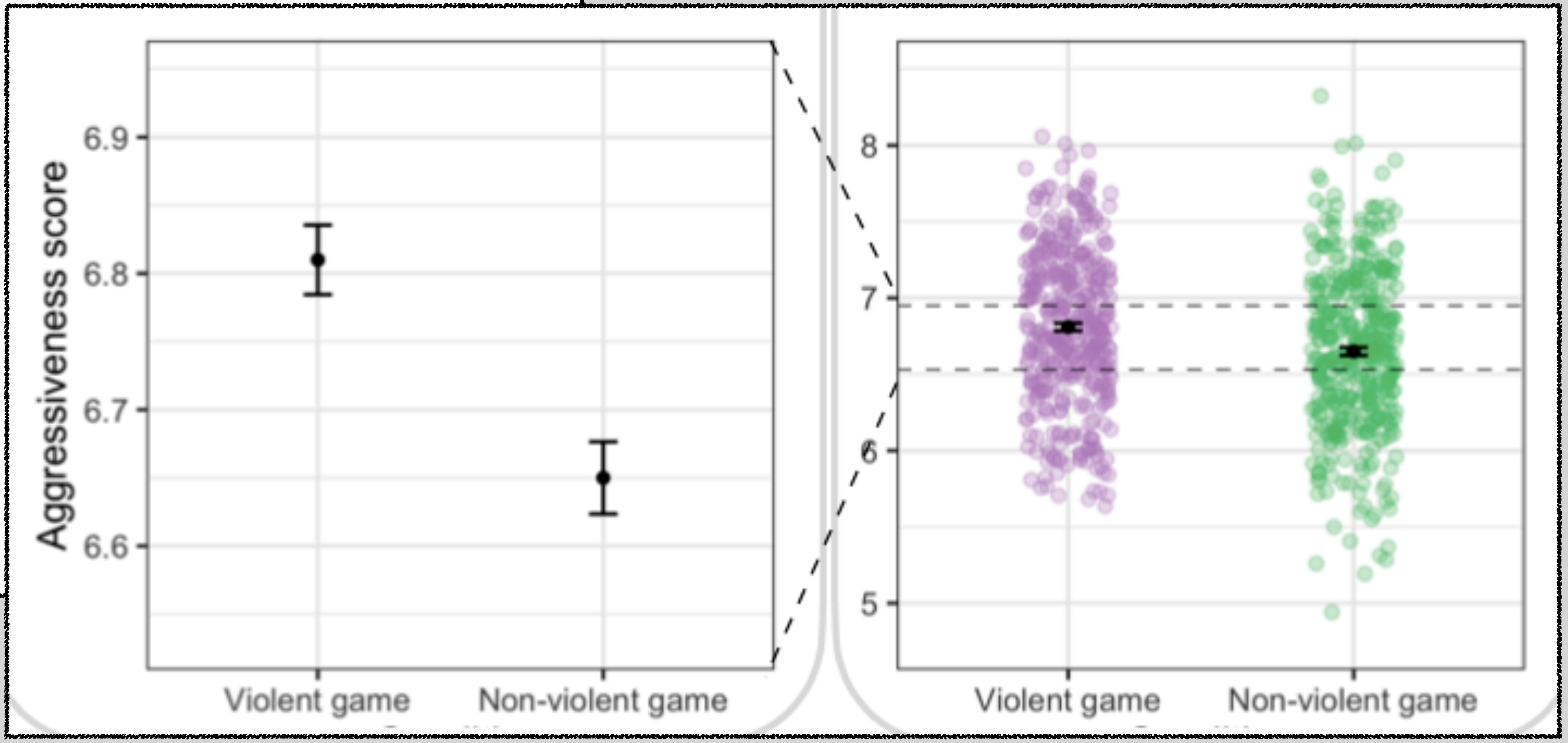
An illusion of predictability in scientific results: Even experts confuse inferential uncertainty and outcome variability

Sam Zhang[a,1], Patrick R. Heck[b], Michelle N. Meyer[c], Christopher F. Chabris[c], Daniel G. Goldstein[d], and Jake M. Hofman[d,1]

Edited by Elke Weber, Princeton University, Princeton, NJ; received February 22, 2023; accepted June 26, 2023

Traditionally, scientists have placed more emphasis on communicating inferential uncertainty (i.e., the precision of statistical estimates) compared to outcome variability (i.e., the predictability of individual outcomes). Here, we show that this can lead to sizable misperceptions about the implications of scientific results. Specifically, we present three preregistered, randomized experiments where participants saw the same scientific findings visualized as showing only inferential uncertainty, only outcome variability, or both and answered questions about the size and importance of findings they were shown. Our results, composed of responses from medical professionals, professional data scientists, and tenure-track faculty, show that the prevalent form of visualizing only inferential uncertainty can lead to significant overestimates of treatment effects, even among highly trained experts. In contrast, we find that depicting both inferential uncertainty and outcome variability leads to more accurate perceptions of results while appearing to leave other subjective impressions of the results unchanged, on average.

statistics | uncertainty | science communication | visualization | experiments



PreFer

Predicting Fertility data challenge

- Be a part of a unique data challenge
- Contribute to fertility research & computational social sciences
- Write a paper for special issue
- Work with amazing data:
  - LISS panel
  - Dutch population registries

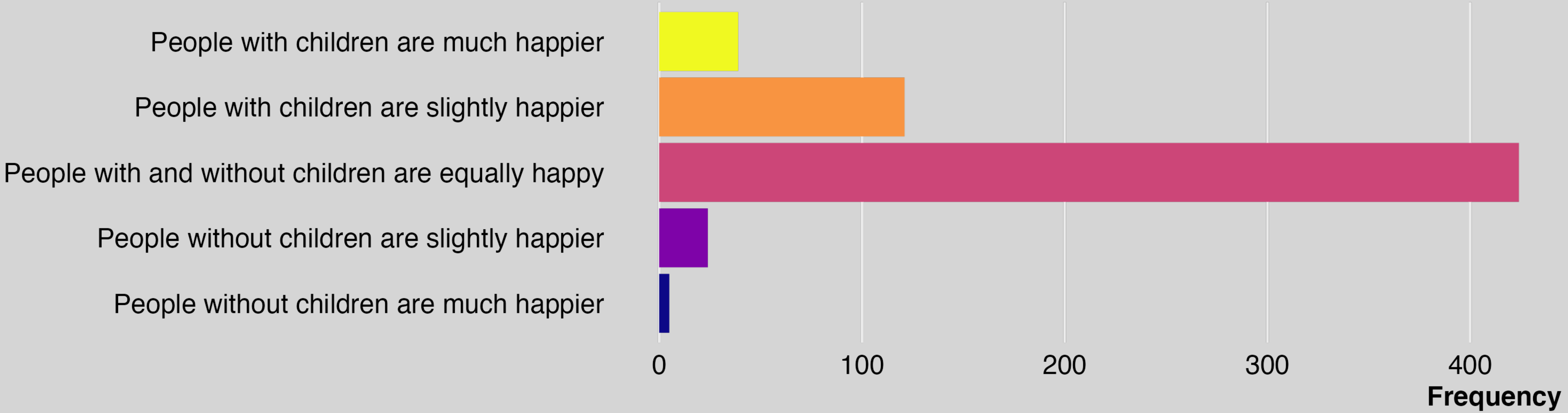
SIGN UP HERE!

the Future

- ✓ assessing non-linearities and interactions
  - more advanced machine learning techniques
- ✓ second wave of data collection
  - causality, although …


How many children would you like to have? (N = 681)
8
7
6
5
4
3
2
1
0
0
100
200
300
Frequency
Do you think you will have (more) children in the future? (N = 758)
Absolutely so
Probably so
I don't know
Probably not
Absolutely not
0
50
100
150
200
Frequency
Most of my friends think that I should have (more) children (N = 580)
Completely agree
Agree
Agree slightly
Do not agree or disagree
Slightly disagree
Disagree
Completely disagree
0
50
100
Frequency
My parents/caretakers think that I should have (more) children (N = 608)
Completely agree
Agree
Agree slightly
Do not agree or disagree
Slightly disagree
Disagree
Completely disagree
0
50
100
Frequency

Which statement best reflects your view when it comes to having children and happiness? (N = 653)

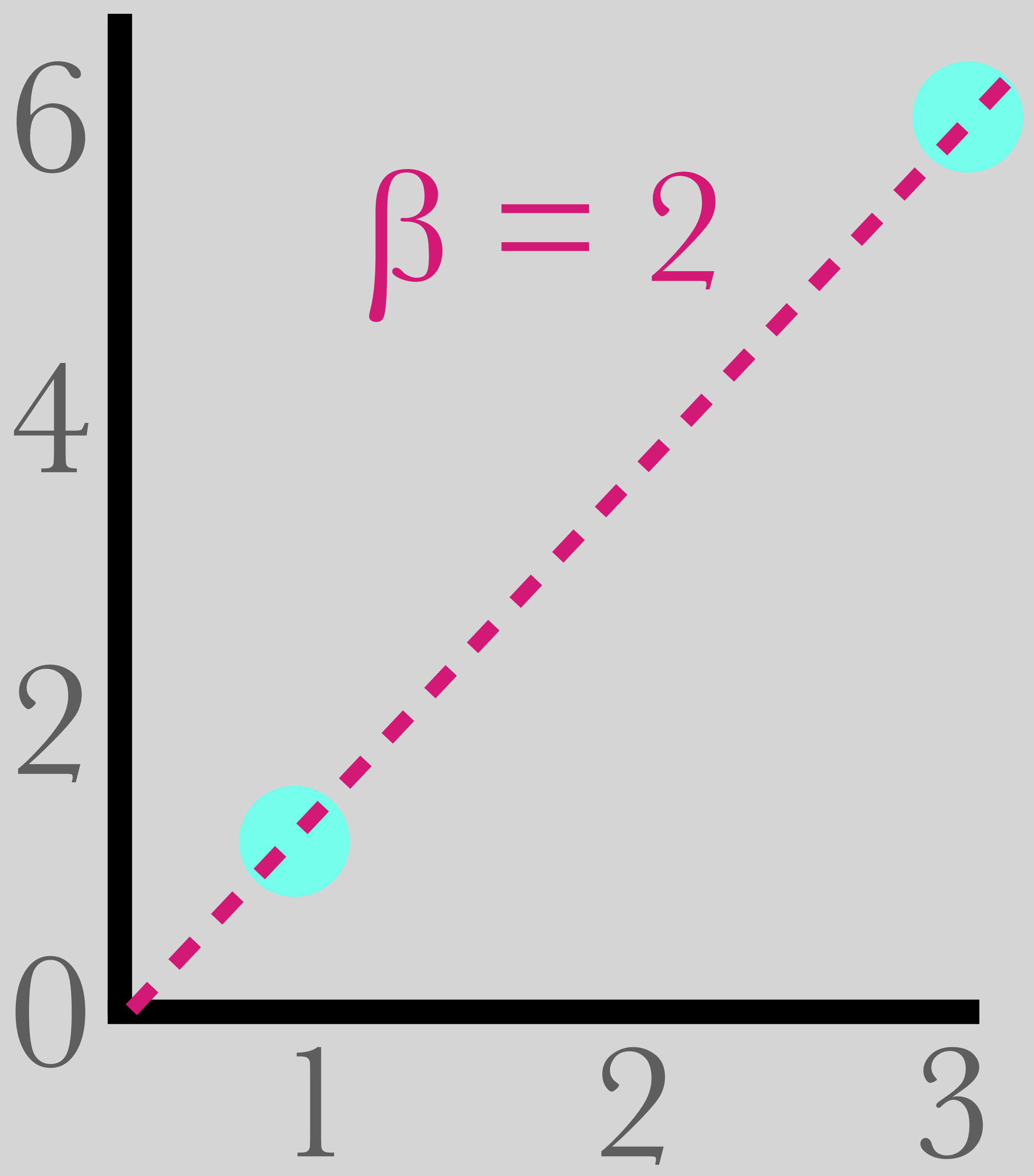


Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

Linear regression

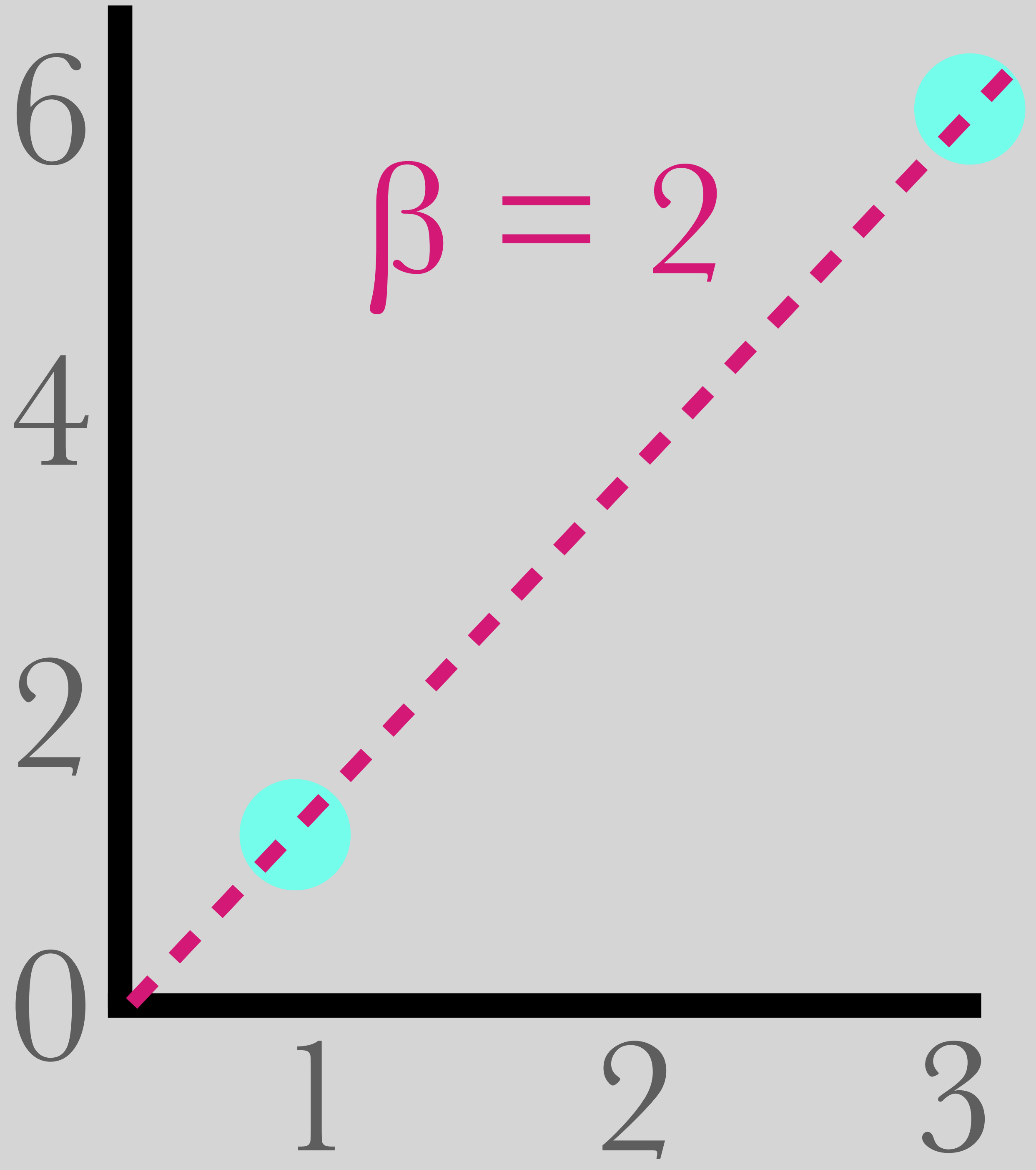
$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$



Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

assume $\lambda = 6$



Linear regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$

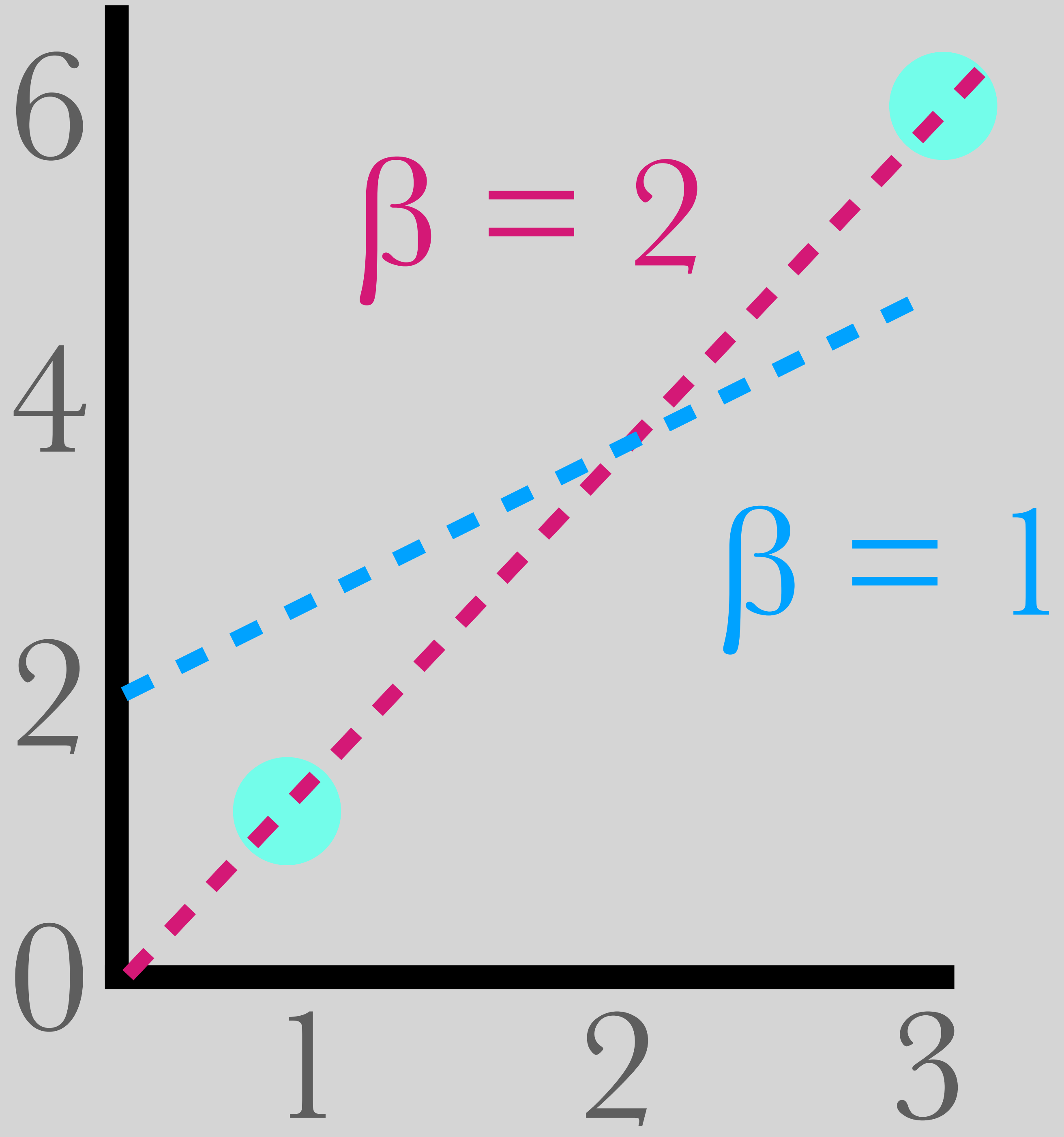
LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|2| = 0 + 12 = 12$$

Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

assume $\lambda = 6$



Linear regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$

LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|2| = 0 + 12 = 12$$

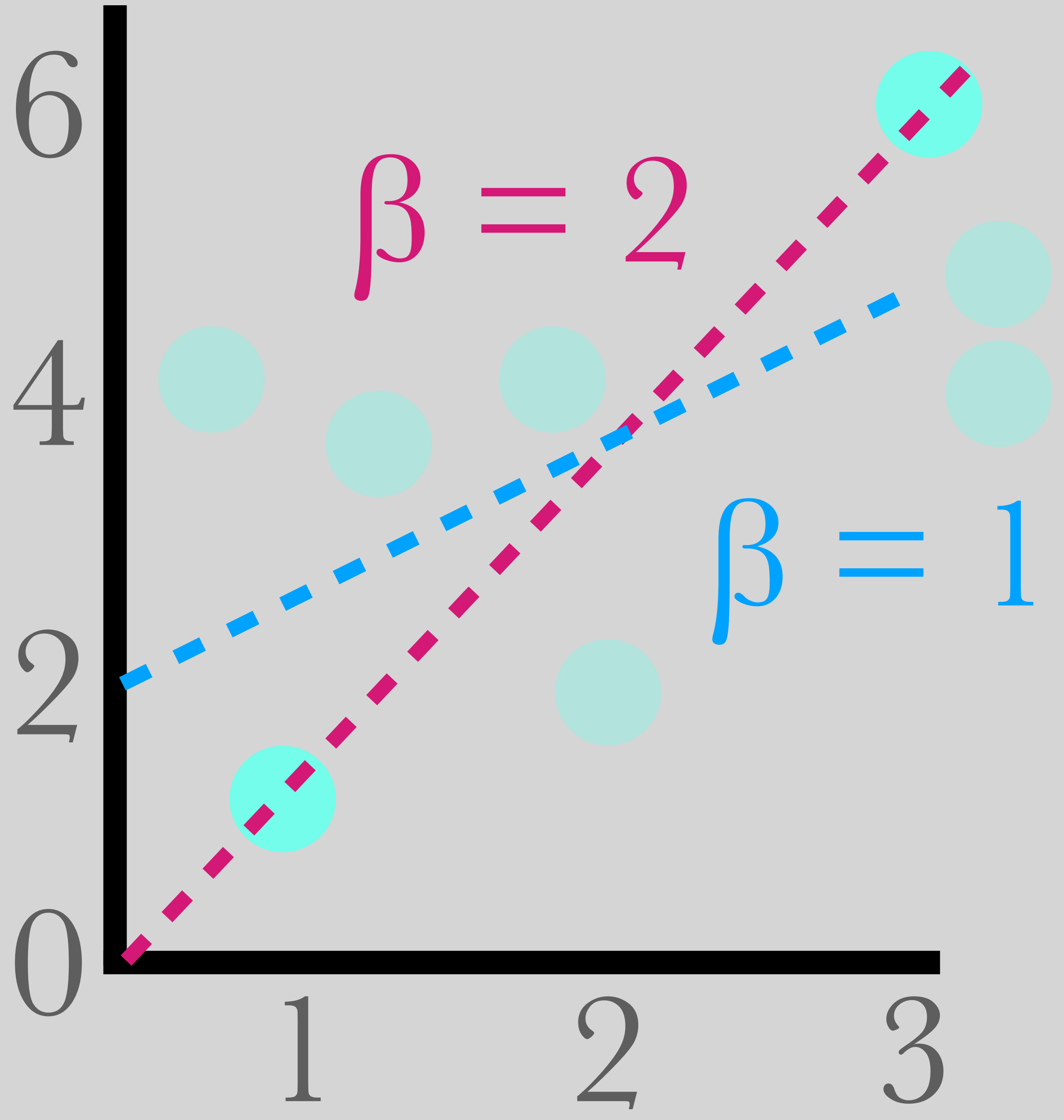
LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1}|1| = 2^2 + 1^2 + 6 = 11$$

Lasso Regression

$$\sum_{i=0}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{p} |\beta_j|$$

assume $\lambda = 6$



Linear regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 = (1-1)^2 + (3-3)^2 = 0$$

LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1} |2| = 0 + 12 = 12$$

LASSO regression

$$\sum_{i=0}^{2}(y_i - \hat{y}_i)^2 + 6\sum_{j=1}^{1} |1| = 2^2 + 1^2 + 6 = 11$$