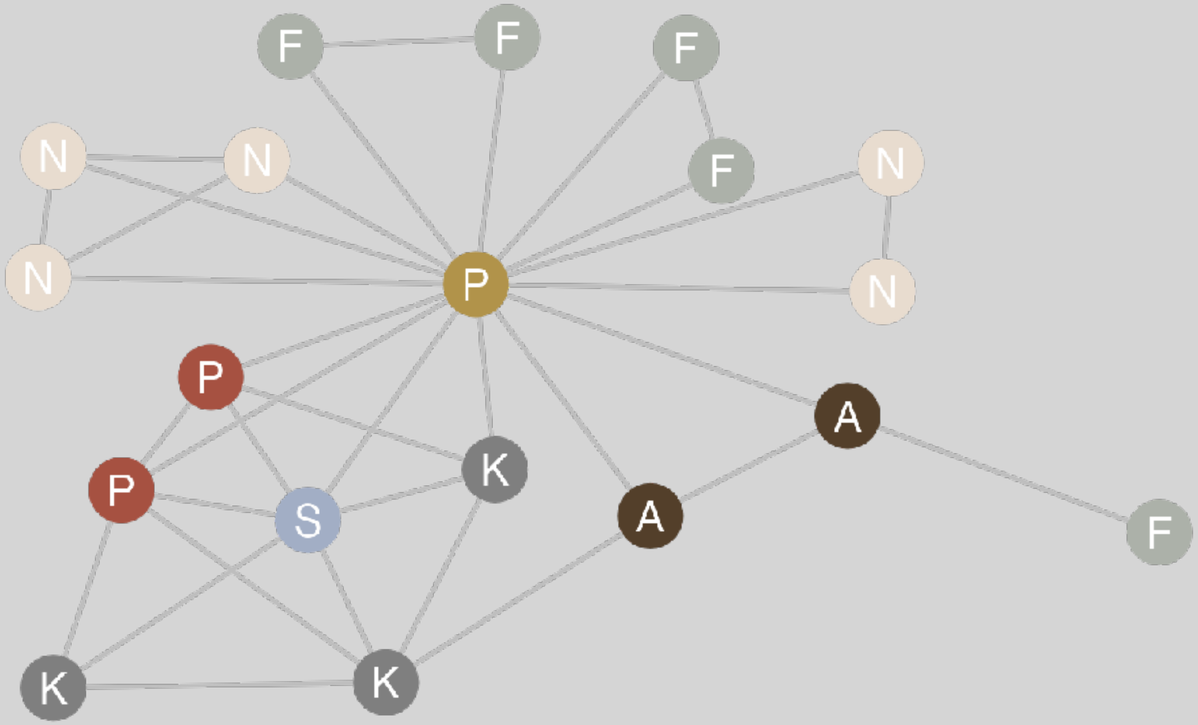
Predicting fertility outcomes with networks



gert stulp - pau vila soler - javier garcia-bernardo gertstulp.com

Graph Neural Networks

Pau
Vila Soler

Javier
Garcia-Bernardo

prediction of unseen cases

=

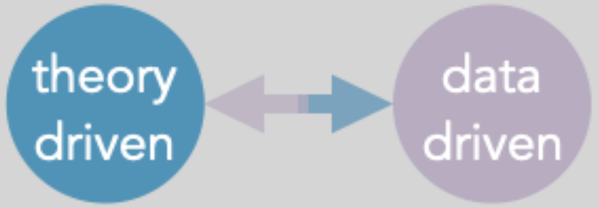
out-of-sample prediction

=

prediction on holdout data

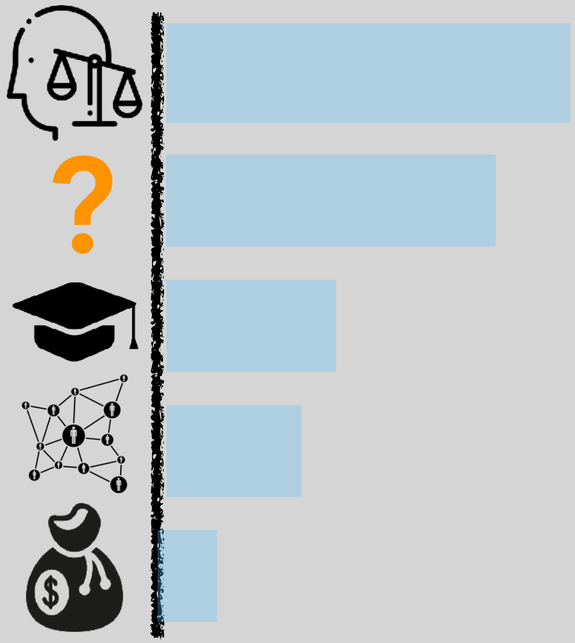
prediction

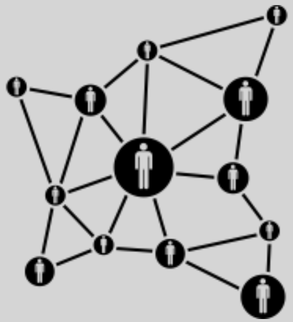
clear measure of effect size

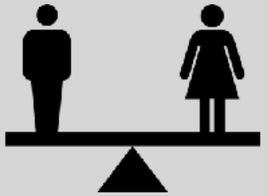


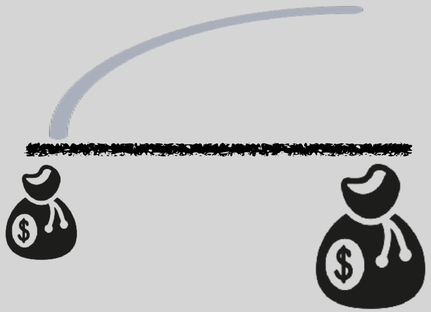
facilitates dialogue theory- and data-driven models

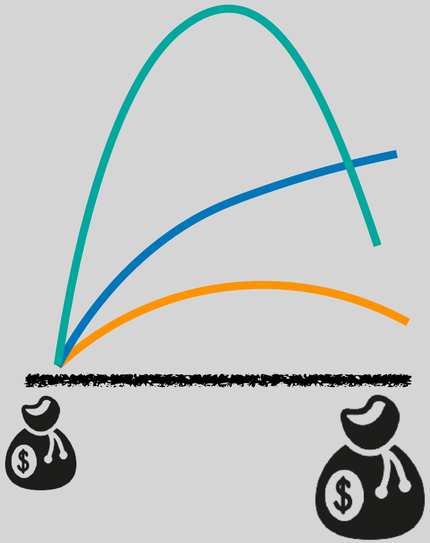
measure of distance theory and practice

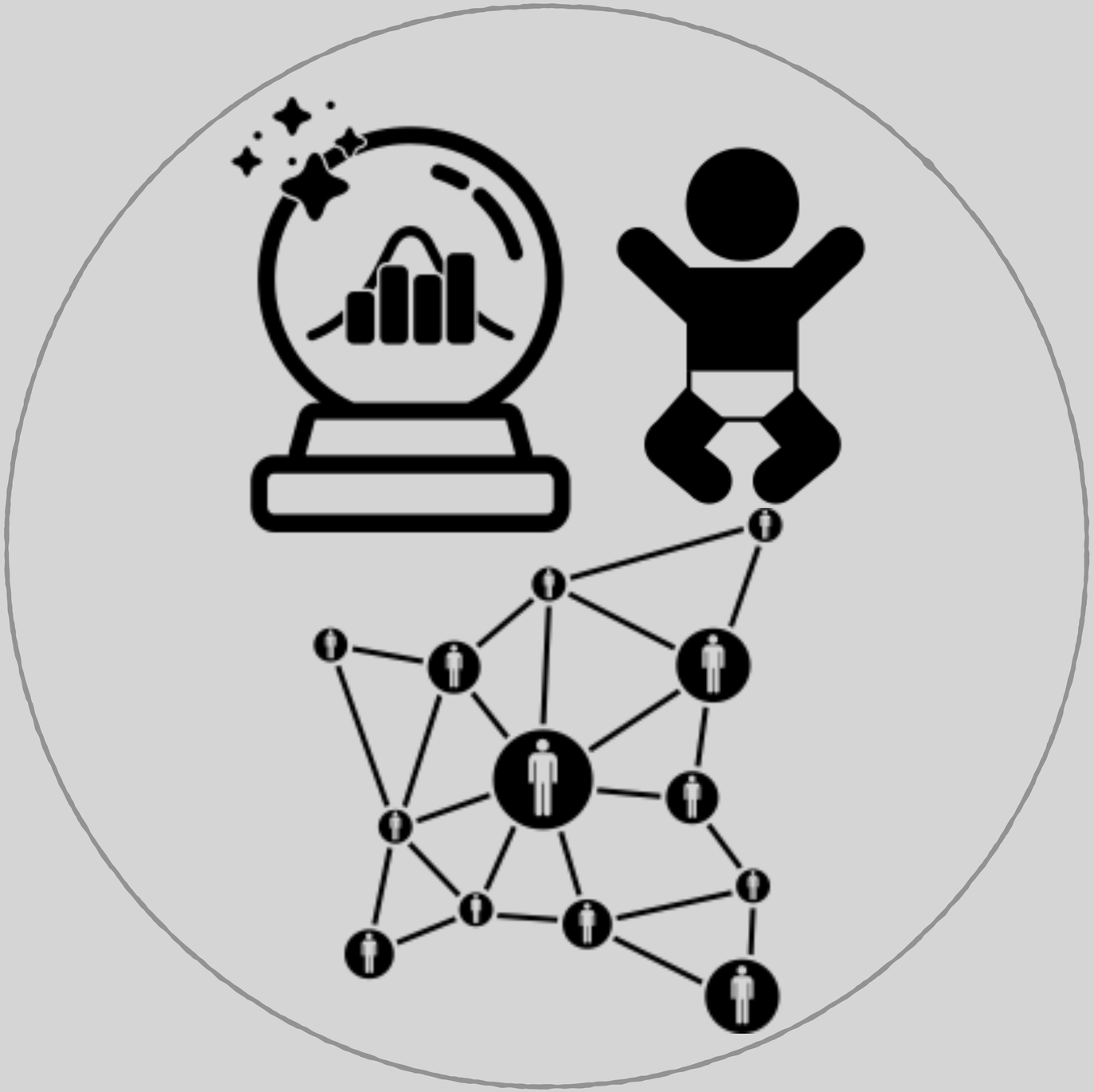












predicting fertility outcomes using personal network data

Methodology

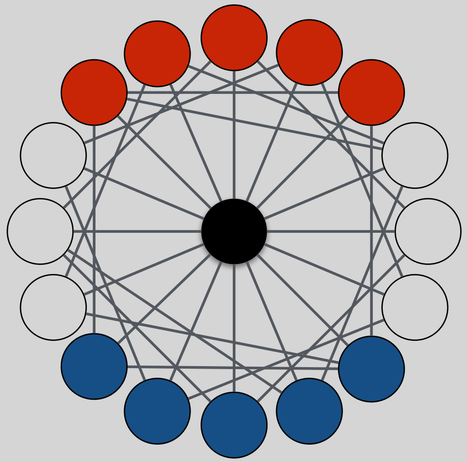
Longitudinal Internet Studies for the Social sciences

~750 women
age: 18 - 40

Ego

Age
Education
Income
Partnership status
Children
Detailed fertility preferences

Alters (25)

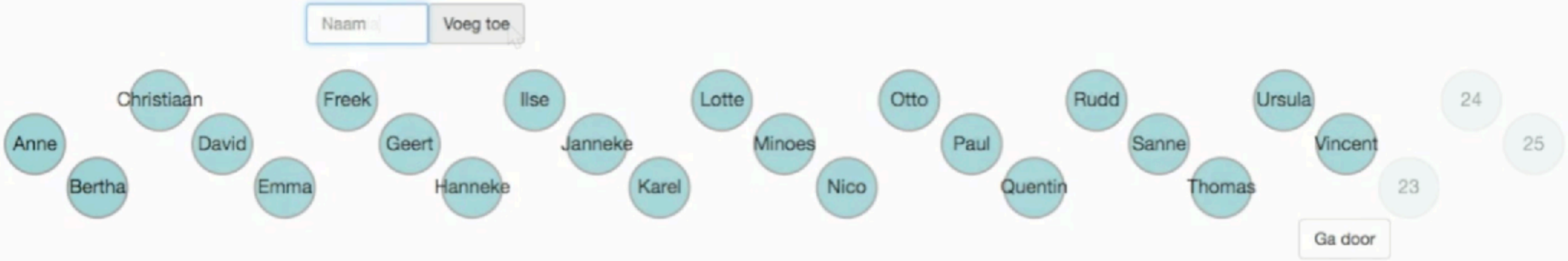


Sex
Age
Education
Relationship type
Closeness
Frequency of contact F2F
Frequency of other contact

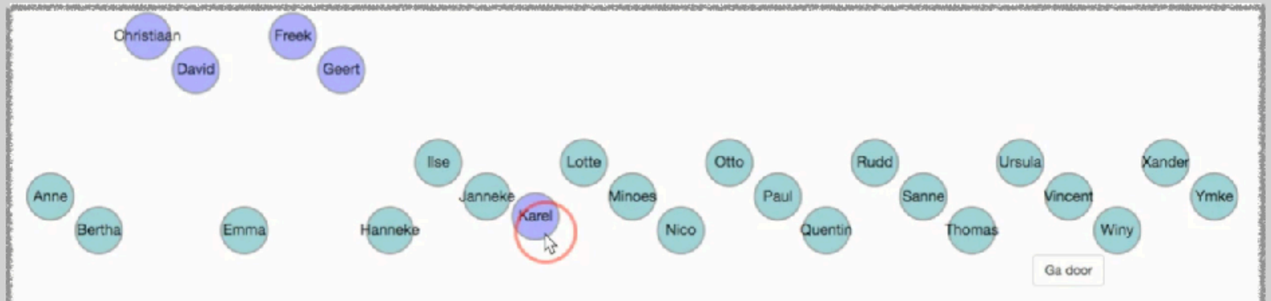
Number and age of children
Friend
Wants children
Does not want children
Help with children
Talk about children
Relationship with other alters

Methodology

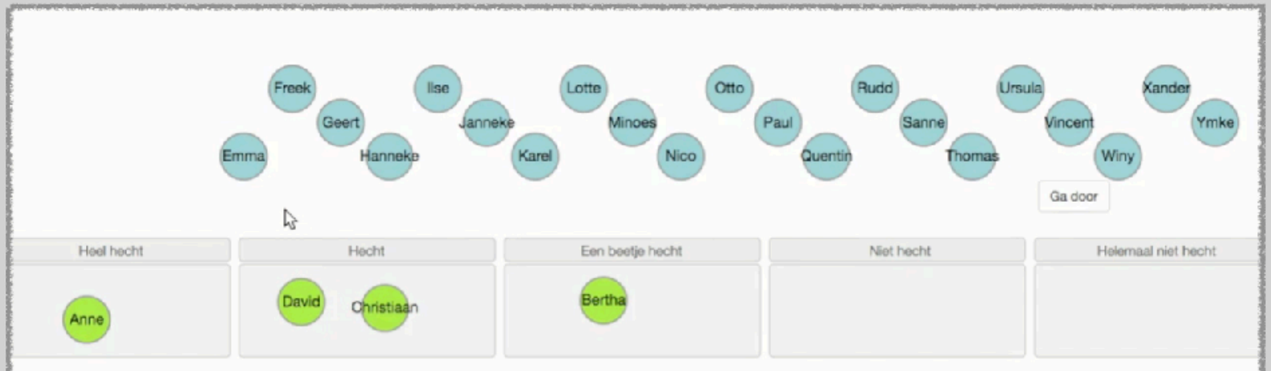
Please list 25 names of individuals 18 years or older with whom you have had contact in the last year. This can be face-to-face contact, but also contact via phone, internet, or email. You know these people and these people also know you from your name or face (think of friends, family, acquaintances, et cetera). You could reach out to these people if you would have to. Please name your partner in case you have one.



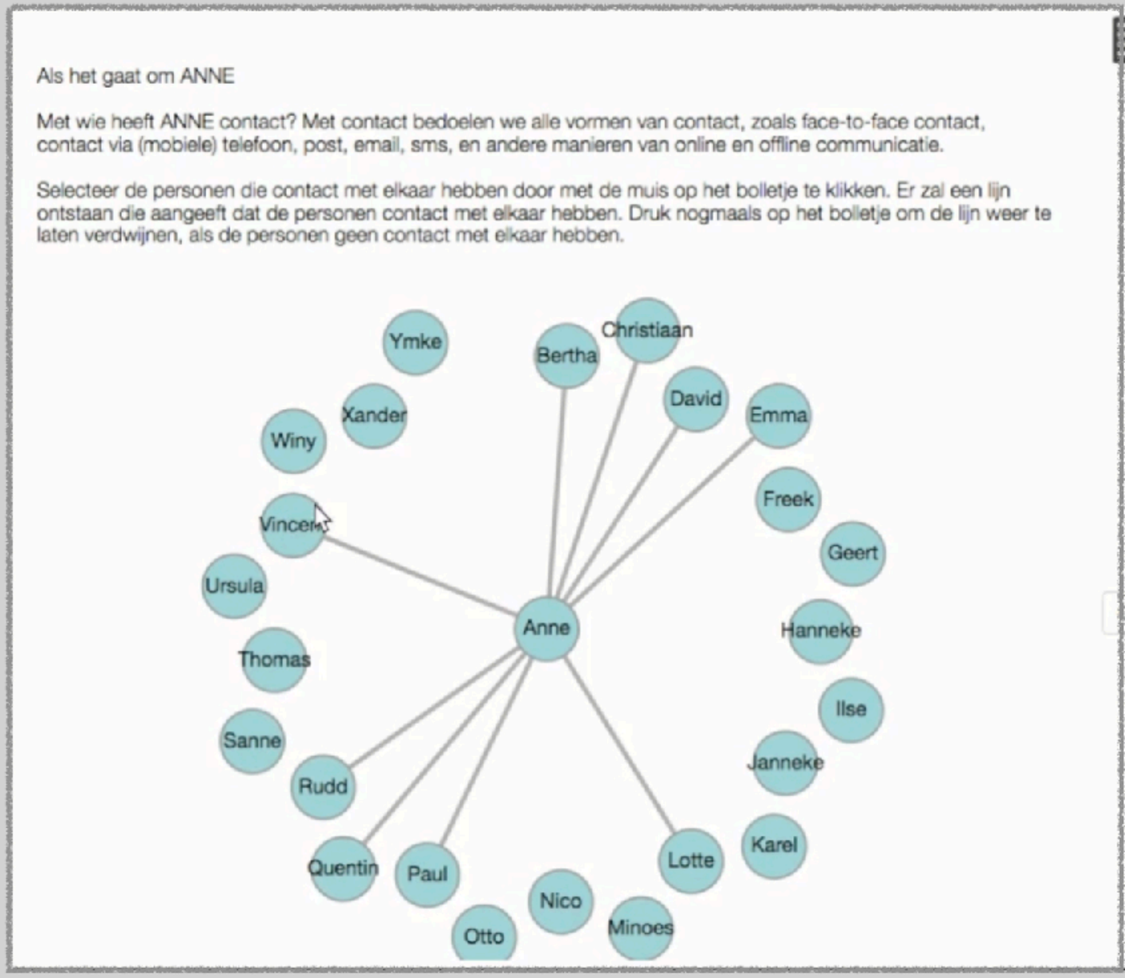
Methodology



Which of these 25 individuals could you ask for help



How close are you to these people?



Methodology

Longitudinal Internet Studies for the Social sciences

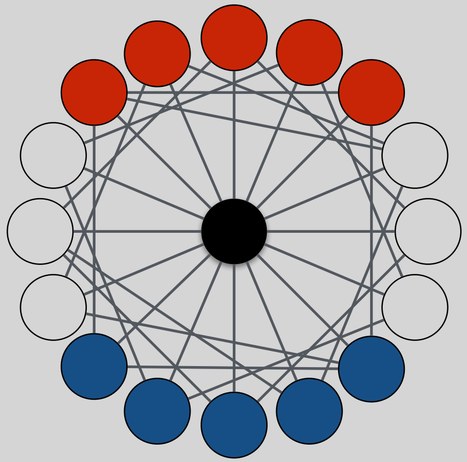
~750 women
age: 18 - 40

Ego

Age
Education
Income
Partnership status
Children
Detailed fertility preferences

OUTCOME

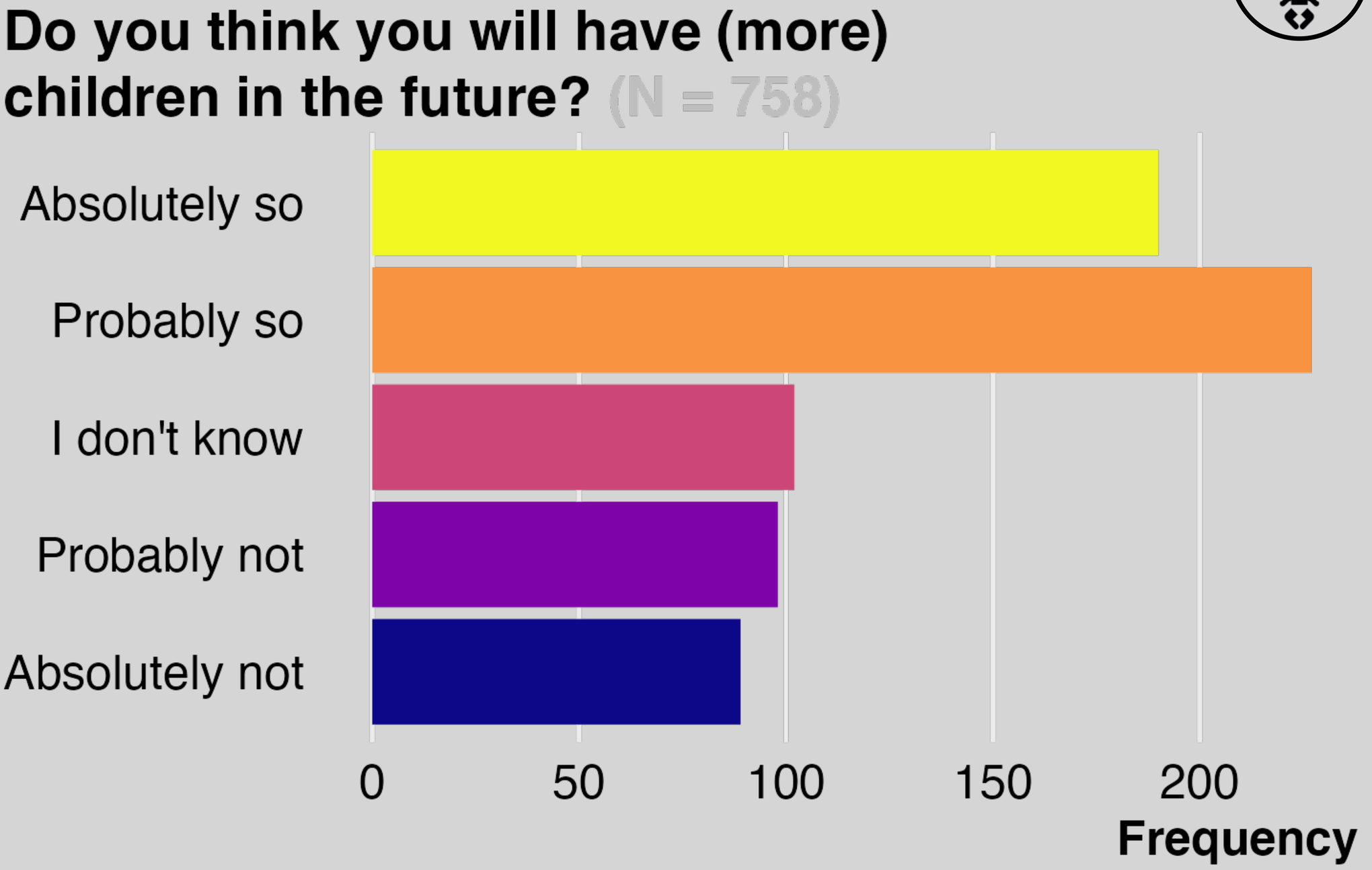
Alters (25)



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

Outcome


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

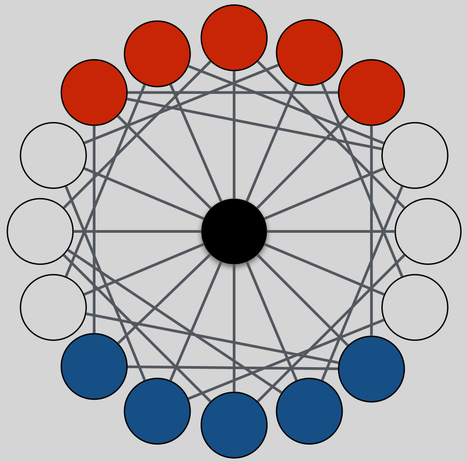
Methodology

Longitudinal Internet Studies for the Social sciences

~750 women
age: 18 - 40

Ego

Alters (25)



EGO VARIABLES

Age
Education
Income
Partnership status
Children
Detailed fertility preferences

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
…

composition

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

…

structure

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

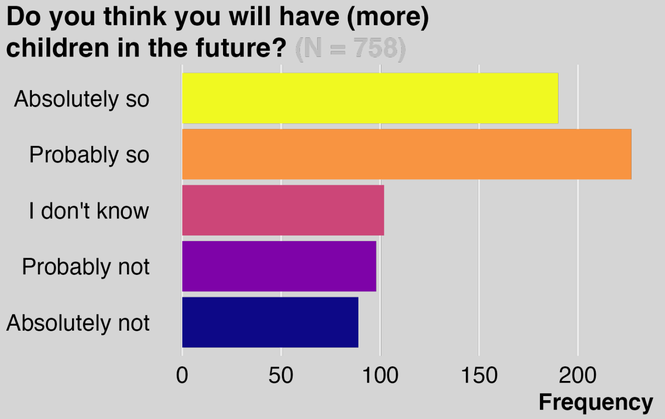
density among family
density among friends
density among childfree
…

24 variables

13 variables

20 variables

Overview



N ~ 700
5 ego vars
57 network vars

train [75%]

age	partner?	# kin	density	outcome
29	no	2	10%	0
41	yes	0	76%	1
34	yes	24	30%	1
33	yes	16	5%	0

training

8 learning algorithms

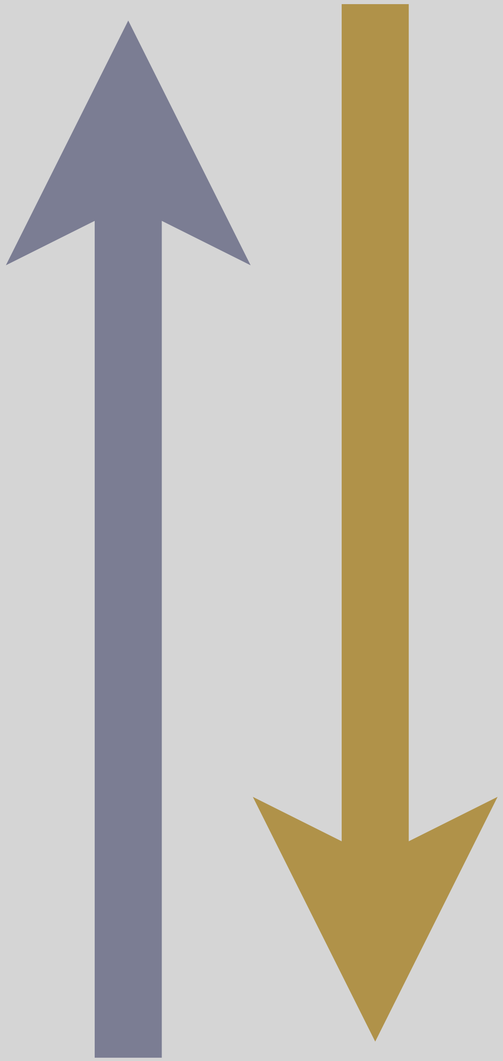
linear/LASSO/RIDGE regression
Elastic net
XGBoost
Cubist rules
Support Vector Machines
Graph Neural Networks

evaluating
out-of-sample R^2

holdout [25%]

age	partner?	# kin	density	outcome
36	yes	12	20%	0
39	yes	17	33%	1

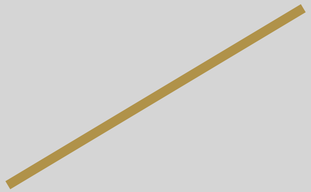
INTERPRETABILITY



Linear regression

LASSO/RIDGE regression

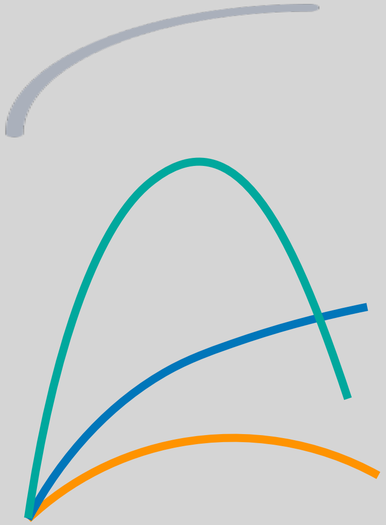
Elastic Net



XGBoost

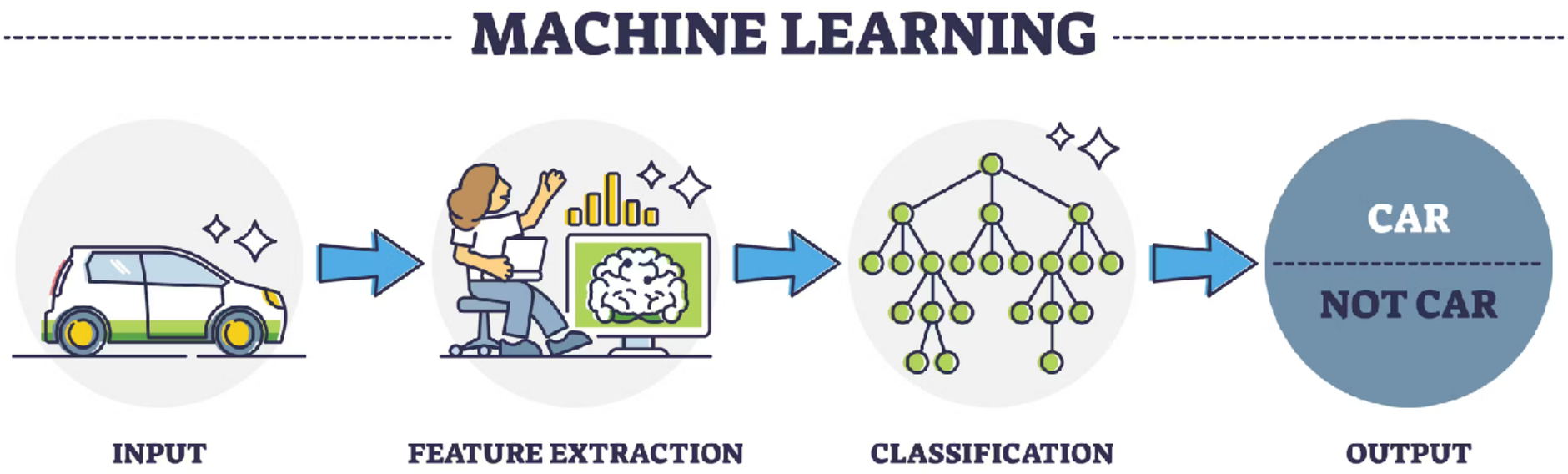
Cubist Rules

Support Vector Machines

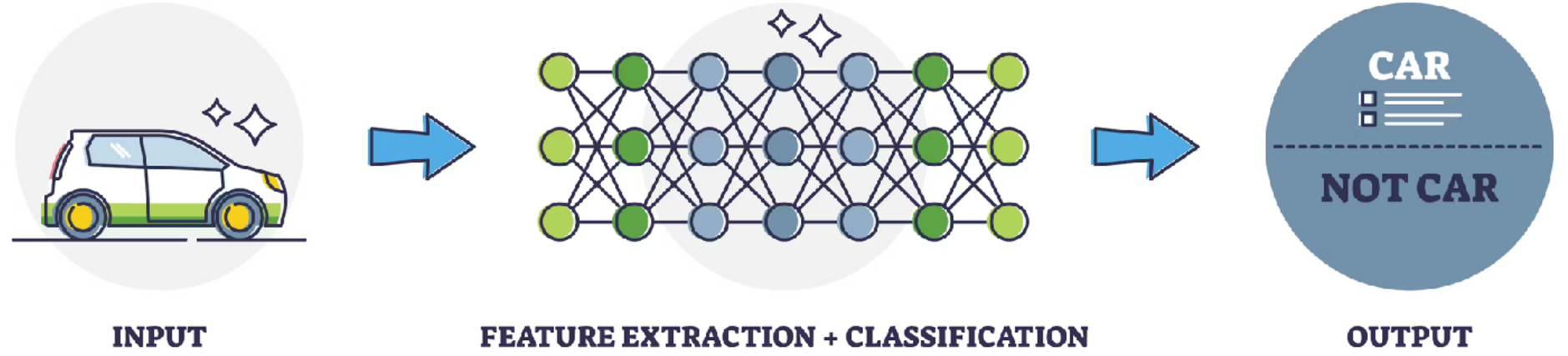


Graph Neural Networks

COMPLEXITY

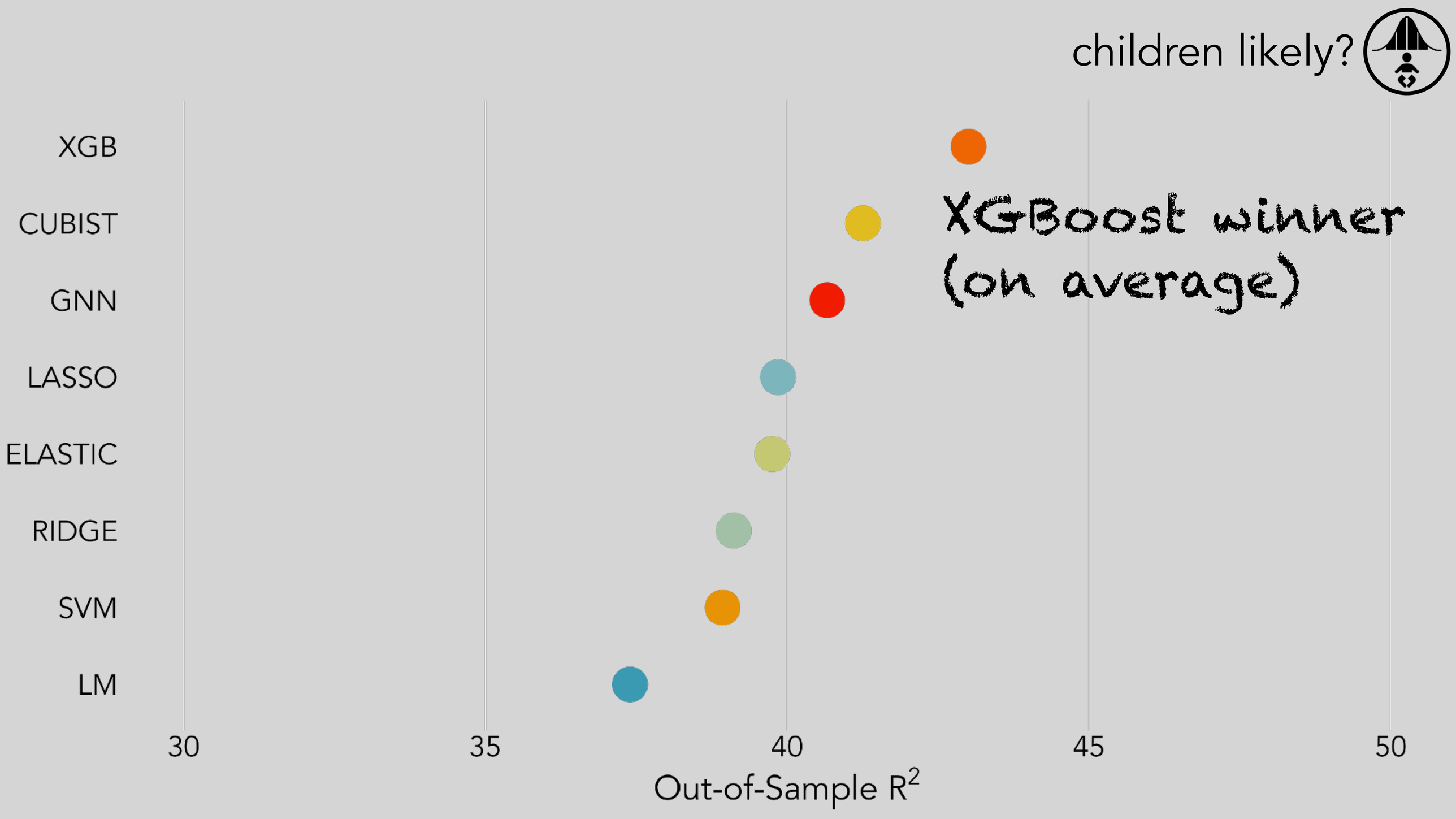


LASSO
XGBoost
SVM
{62 vars}

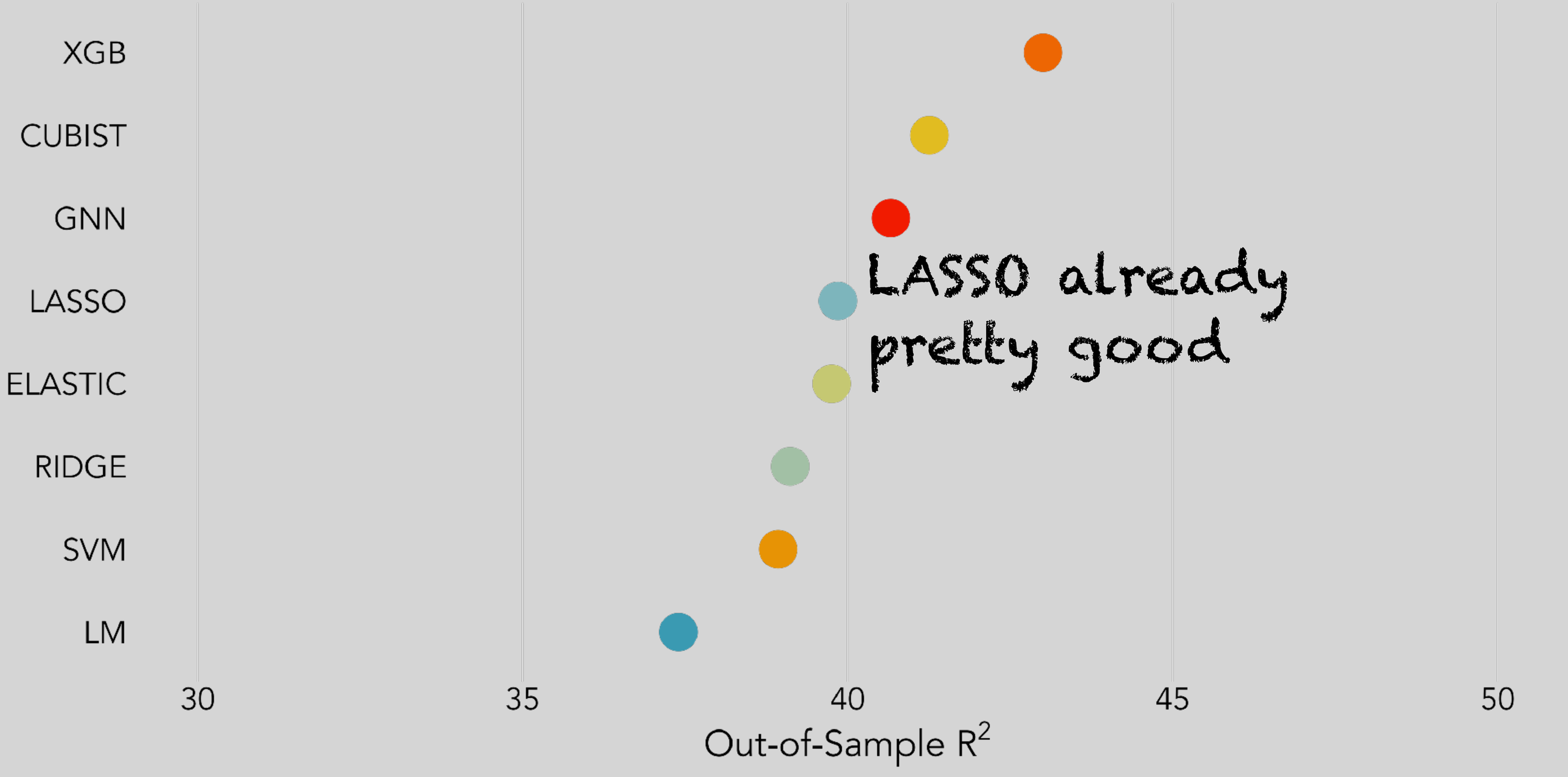


GNN
{19 vars}

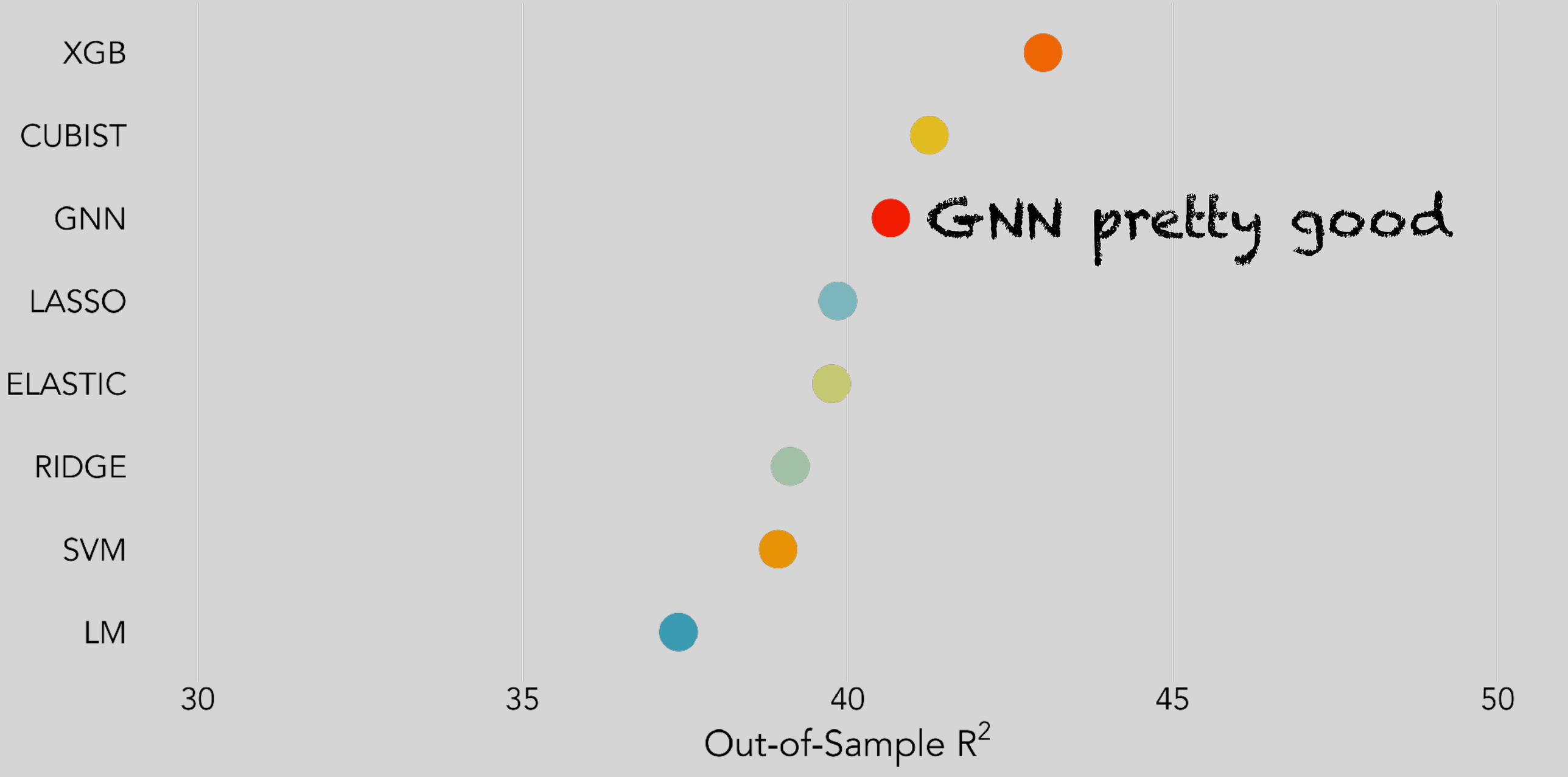
children likely?



children likely?



children likely?



Take-home messages

✅ tree-based methods better than ‘linear’ methods

non-linearities?

❌ improvement over ‘linear’ methods not impressive

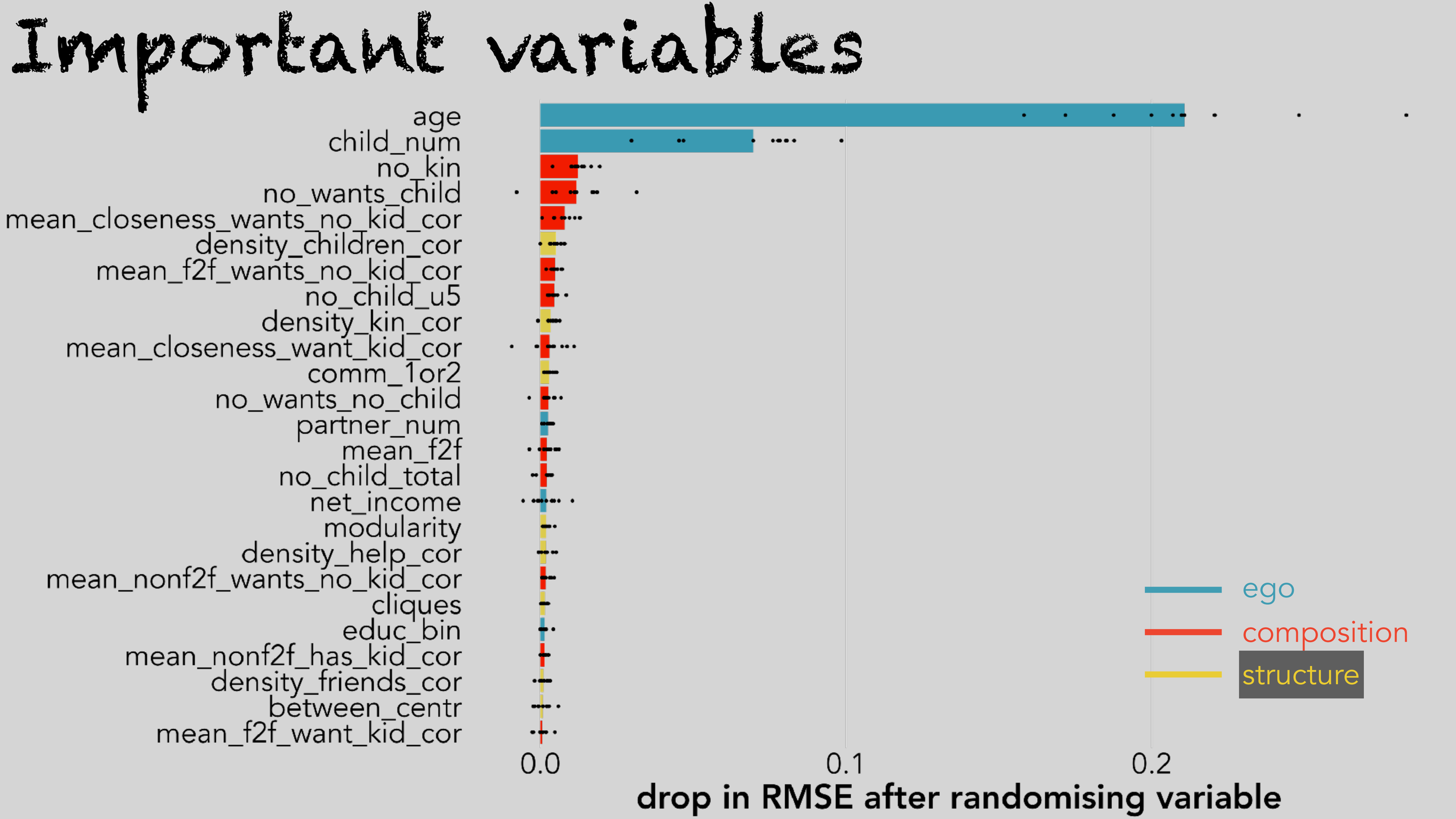
is lack of interpretability and dozens of hours compute worth it?

✅ GNN performed well!

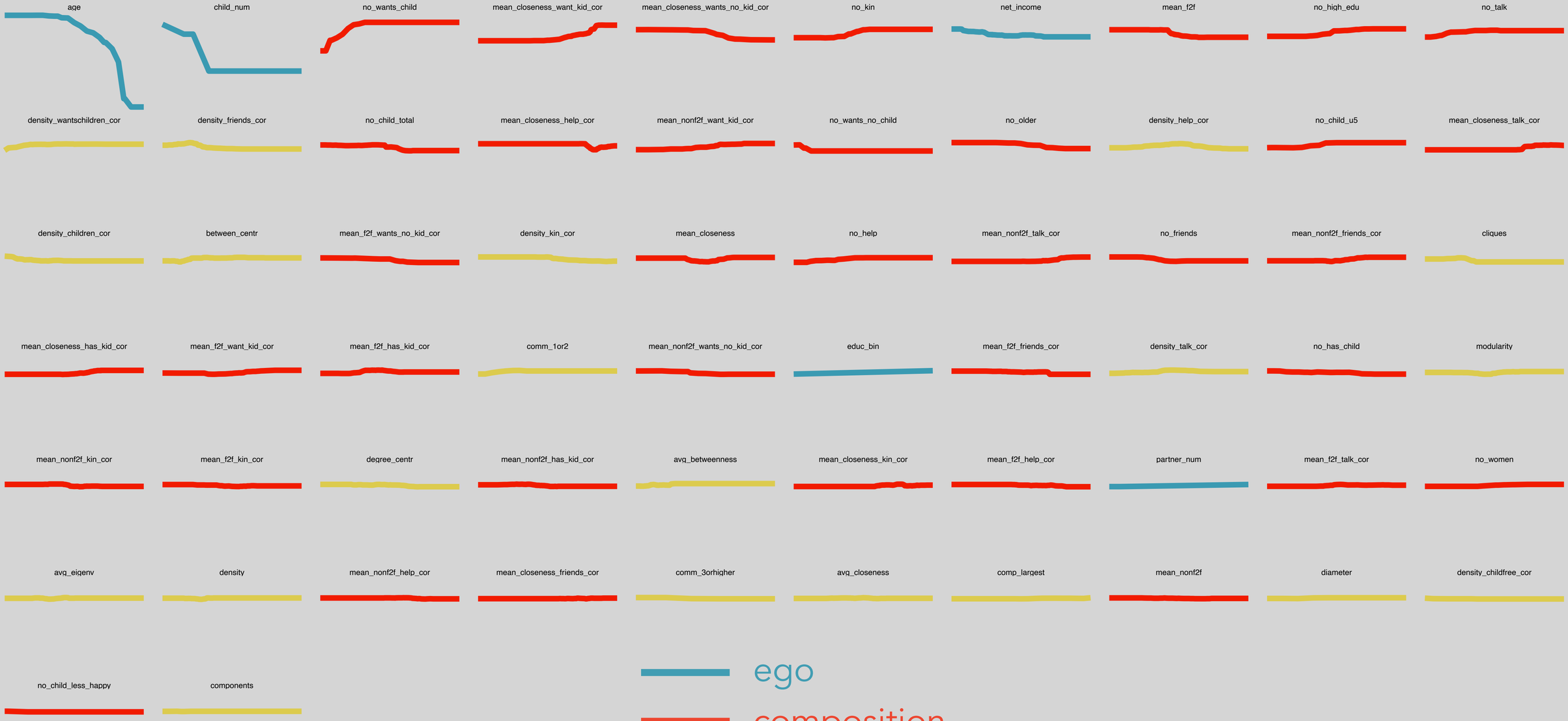
requires fewer decisions, capitalises on network structure


XGBoost

Important variables



Non-linear effects



ego

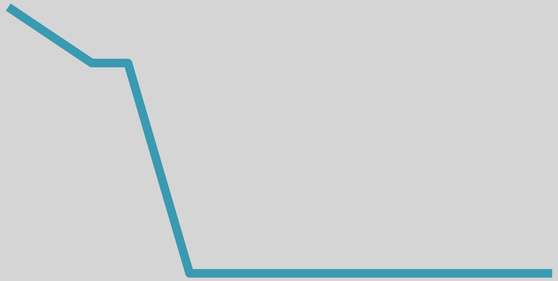
composition

structure

Non-linear effects

age

children



alters that want children

average closeness to alters who want children

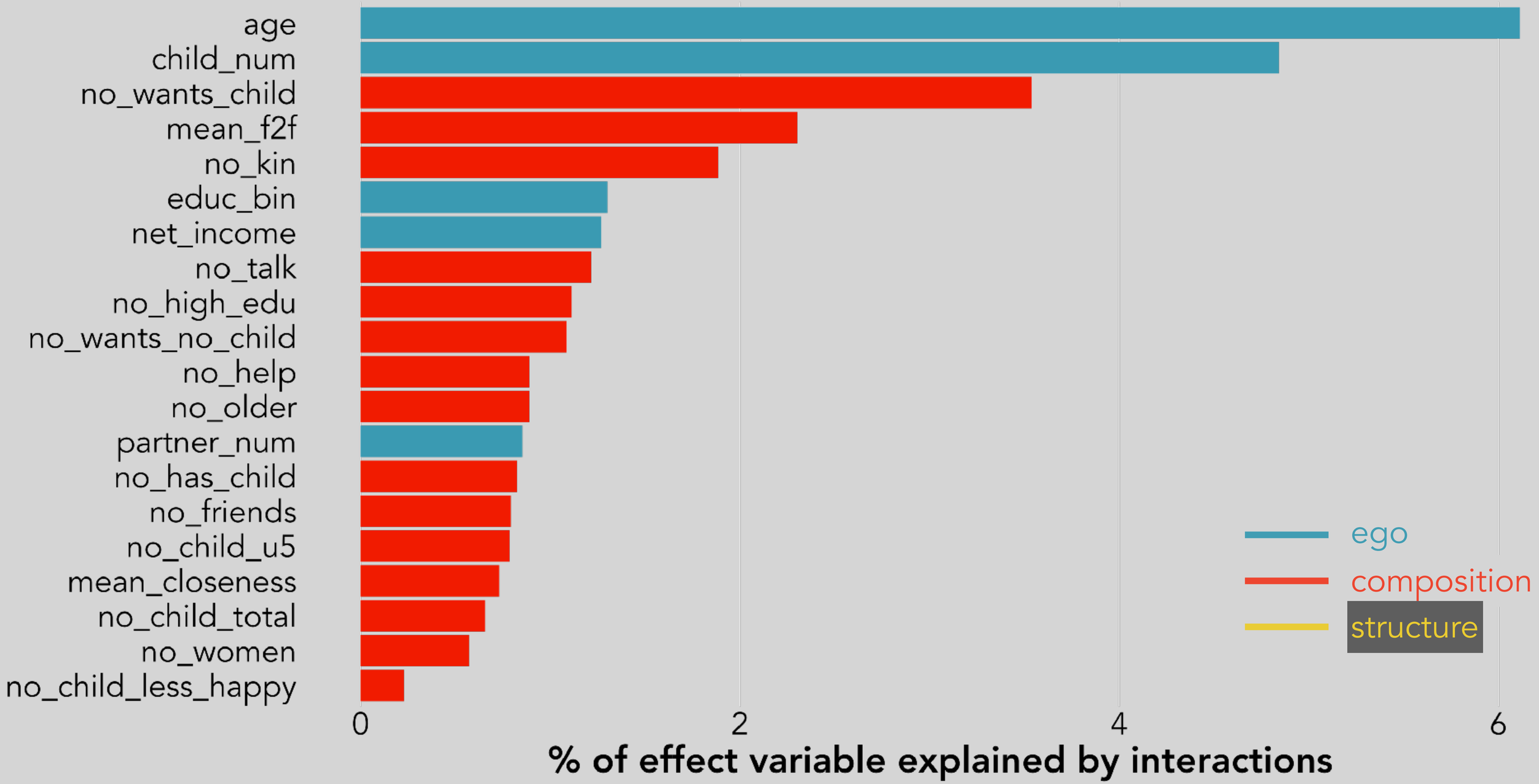
average closeness to alters who do no want children

kin

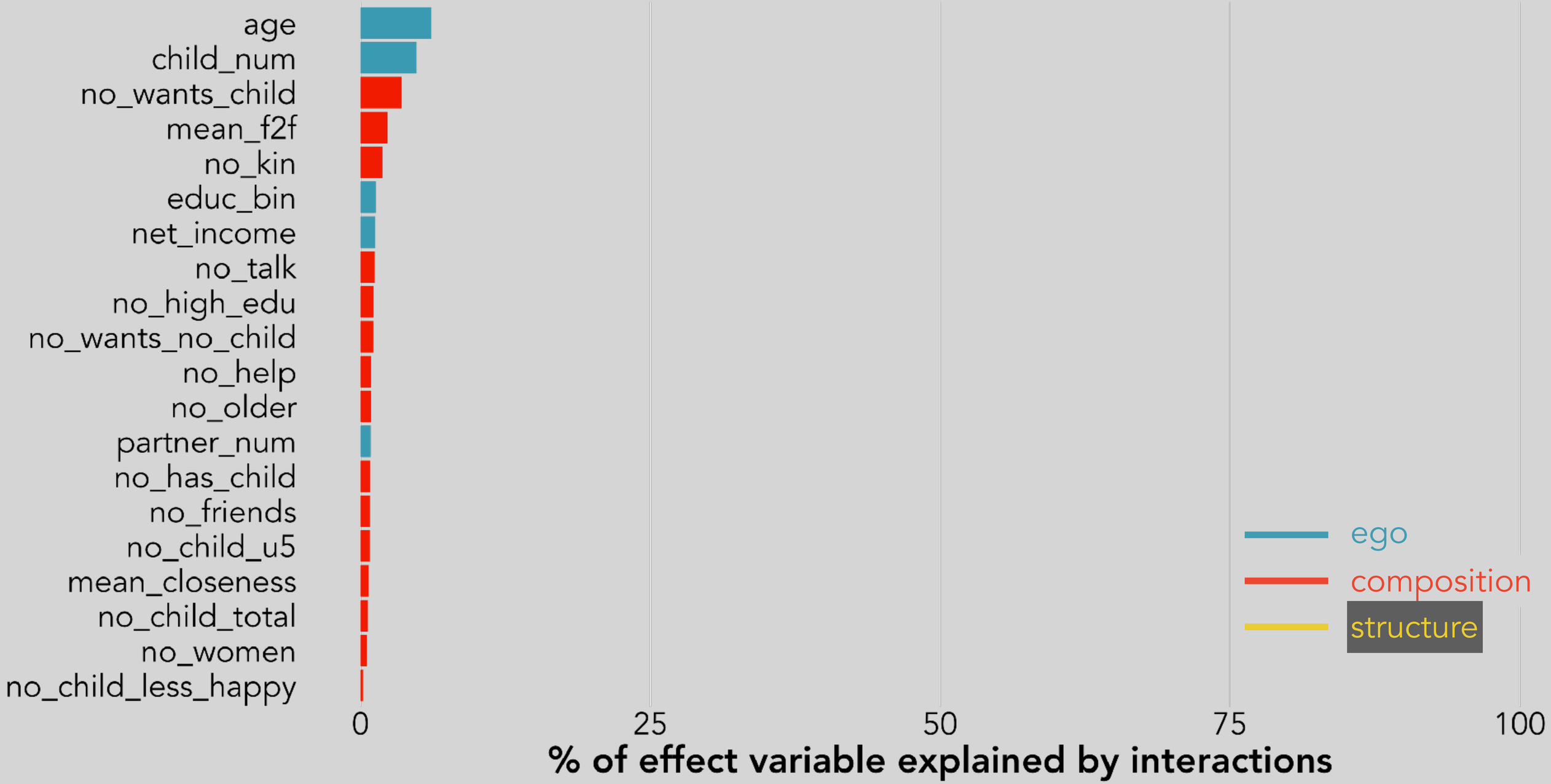
net income

mean f2f contact

Interactions



Interactions



Why fancy ML slighty better?

✅ some variables have strong non-linear effects

age, number of previous children

❌ interactions account for little variation

constraints of sample size?

✅ tools developed to systematically assess non-linearities

hardly done in ‘orthodox’ modelling

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

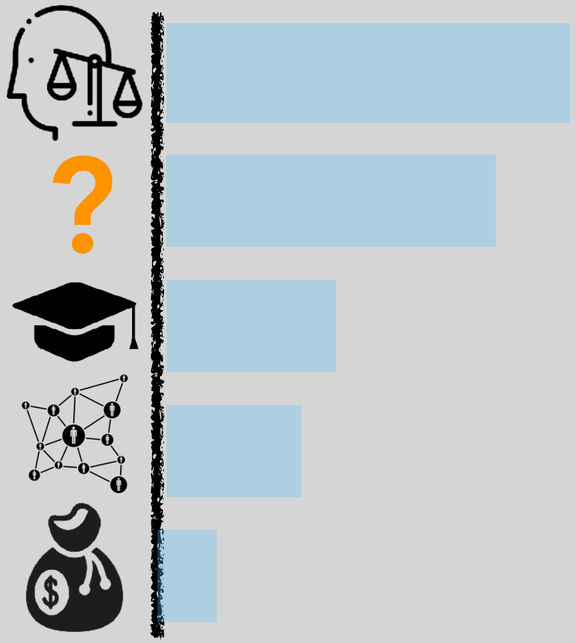
DEMOGRAPHIC RESEARCH

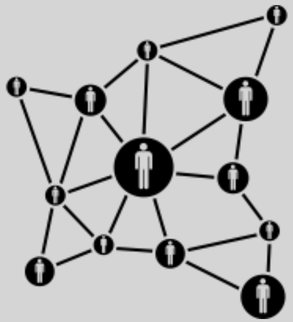
VOLUME 49, ARTICLE 19, PAGES 493–512
PUBLISHED 8 SEPTEMBER 2023
https://www.demographic-research.org/Volumes/Vol49/19/
DOI: 10.4054/DemRes.2023.49.19

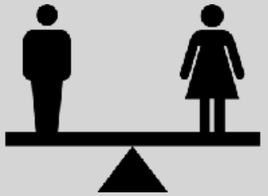
Data Description

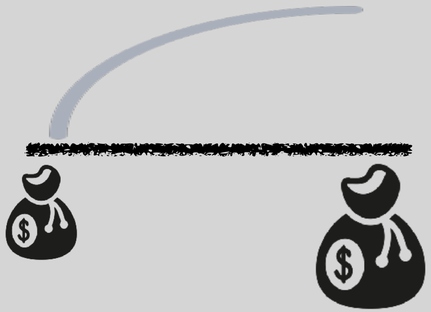
Describing the Dutch Social Networks and Fertility Study and how to process it

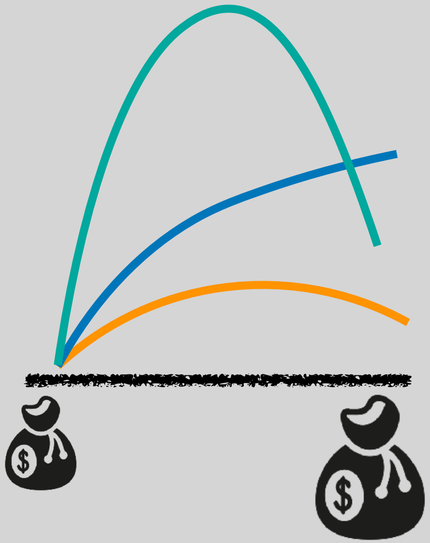
Gert Stulp











Am I data dredging!?

Am I data dredging!?

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. Schnei

[a] Cognitive and Behavioural Ecology Interdisciplinary Program, Memo
[b] Department of Ocean Sciences, Ocean Sciences Centre, Memorial Un
[c] Department of Biology, Memorial University of Newfoundland, St Jo

Check for updates

General Article

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

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
22(11) 1359–1366

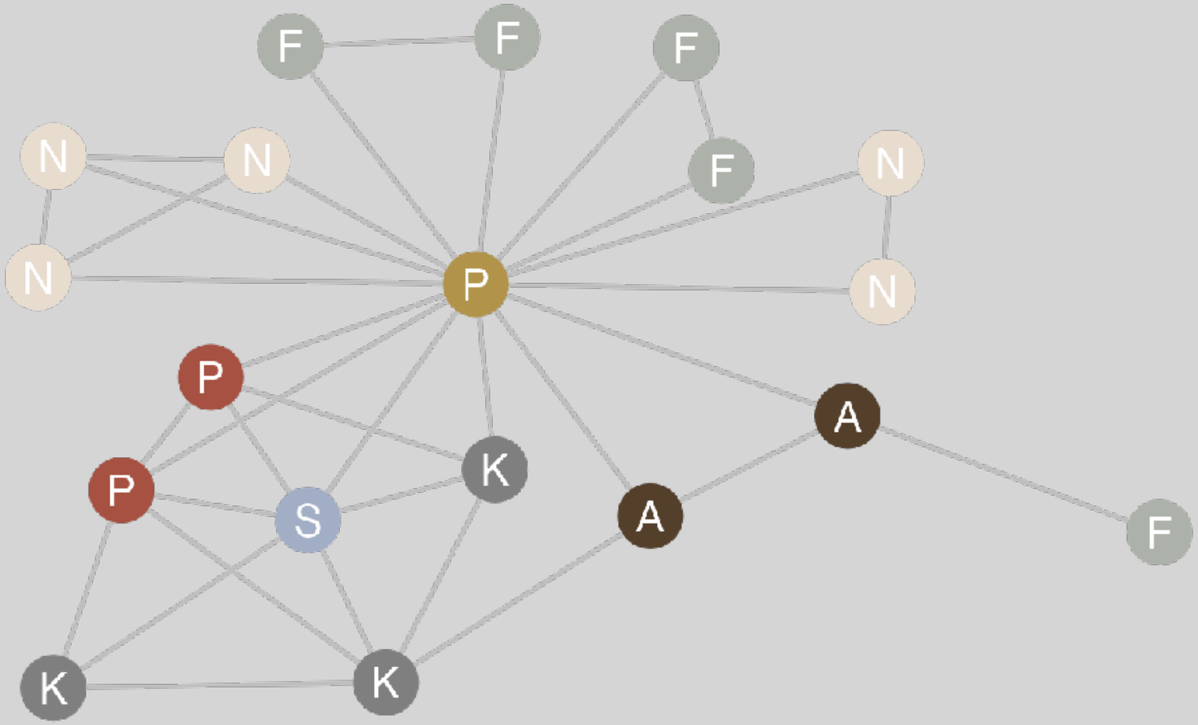
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

good prediction does not mean causal understanding

good prediction does not mean causal understanding

but what does (supposed) causal understanding mean without good predictions

Predicting fertility outcomes with networks



gert stulp - pau vila soler - javier garcia-bernardo gertstulp.com