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







clear measure of effect size



facilitates dialogue theoryand data-driven models



measure of distance theory and practice











...%?











predicting COULE COMACS personal dala



sciences







Age Income



- Education
- Partnership status
- # Children
- Detailed fertility preferences

- Education
- Relationship type
- Closeness
- Frequency of contact F2F Talk about children

Number and age of children Friend Wants children Does not want children Help with children Frequency of other contact 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.







Which of these 25 individuals could you ask for help



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.





sciences







Age Income



OUTCOME

- Education
- Partnership status
- # Children
- Detailed fertility preferences

- Education
- Relationship type
- Closeness
- Frequency of contact F2F Talk about children

Number and age of children Friend Wants children Does not want children Help with children Frequency of other contact Relationship with other alters





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

50

100



200 150 Frequency









Studies for the Social sciences



PANEL





Alters (25)



Age Income

Sex

Age

EGO VARIABLES

- Education
- Partnership status
- # Children
- Detailed fertility preferences

NETWORK VARIABLES

- Education
- Relationship type
- Closeness
- Frequency of contact F2F Talk about children
- Number and age of children Friend Wants children Does not want children Help with children
- Frequency of other contact Relationship with other alters









Personal Melworks

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

• • •

24 variables 13 variables 20 variables





structure

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

density among family density among friends density among childfree

• • •

• • •





OVETVIEW

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



N ~ 700 5 ego vars 57 network vars

out-of-sample R²

holdout [25%] 39



training evaluating

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







XGBoost Cubist Rules Support Vector Machines

COMPLEXITY

Graph Neural Networks









MACHINE LEARNING



FEATURE EXTRACTION

DEEP LEARNING



NOT CAR FEATURE EXTRACTION + CLASSIFICATION OUTPUT

INPUT



LASSO XGBoost SVM {62 vars}

GNN {19 vars}

XGB		
CUBIST		
GNN		
LASSO		
ELASTIC		
RIDGE		
SVM		
LM		
	30	35

children likely?

XGBOOSE winner (on average)

40 Out-of-Sample R²







XGB		
CUBIST		
GNN		
LASSO		
ELASTIC		
RIDGE		
SVM		
LM		
	30	35

children likely?







XGB		
CUBIST		
GNN		
LASSO		
ELASTIC		
RIDGE		
SVM		
LM		
	30	35

children likely?

CONN pretty good 40 Out-of-Sample R² 45





Take home messages

Tree-based methods better than 'linear' methods non-linearities?

(X) 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











Non-Linear effects child_num no_wants_child mean_closeness_want_kid_cor mean_closeness_wants_n age density_wantschildren_cor density_friends_cor no_child_total mean_nonf2f_want_ki mean_closeness_help_cor density_children_cor mean_f2f_wants_no_kid_cor density_kin_cor mean_closeness between_centr mean_f2f_want_kid_cor mean_closeness_has_kid_cor mean_f2f_has_kid_cor comm_1or2 mean_nonf2f_wants_no_ mean_nonf2f_kin_cor mean_f2f_kin_cor mean_nonf2f_has_kid_cor avg_betweenness degree_centr mean_nonf2f_help_cor comm_3orhigher avg_eigenv density mean_closeness_friends_cor no_child_less_happy components

o_kid_cor	no_kin	net_income	mean_f2f	no_high_edu	no_talk
d_cor	no_wants_no_child	no_older	density_help_cor	no_child_u5	mean_closeness_t
	no_help	mean_nonf2f_talk_cor	no_friends	mean_nonf2f_friends_cor	cliques
_kid_cor	educ_bin	mean_f2f_friends_cor	density_talk_cor	no_has_child	modularity
3	mean_closeness_kin_cor	mean_f2f_help_cor	partner_num	mean_f2f_talk_cor	no_women
	avg closeness	comp largest	mean_nonf2f	diameter	density_childfree

ego
composition
structure

alk cor





alters that want children

average closeness to alters who want children

net income

mean f2f contact





Interactions

age child_num no_wants_child mean_f2f no_kin educ_bin net_income no_talk no_high_edu no_wants_no_child no_help no_older partner_num no_has_child no_friends no_child_u5 mean_closeness no_child_total no_women no_child_less_happy





Interactions

age child_num no_wants_child mean_f2f no_kin educ_bin net_income no_talk no_high_edu no_wants_no_child no_help no_older partner_num no_has_child no_friends no_child_u5 mean_closeness no_child_total no_women no_child_less_happy

25





- Some variables have strong non-linear effects age, number of previous children
- (X) interactions account for little variation constraints of sample size?
- Tools developed to systematically assess non-linearities hardly done in 'orthodox' modelling



FertNet: Process Data from the Social Networks and Fertility Survey

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

Version:	0.1.1
Imports:	<u>haven</u> ($\geq 2.5.1$)
Suggests:	<u>testthat</u> (\geq 3.0.0), <u>tidygraph</u> (\geq 1.2.2)
Published:	2023-03-16
Author:	Stulp Gert 💿 [aut, cre]
Maintainer:	Stulp Gert <g.stulp at="" rug.nl=""></g.stulp>
License:	<u>CC BY 4.0</u>
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): <u>FertNet 0.1.1.tgz</u> , r-oldrel (arm64): <u>FertNet 0.1.1.tgz</u> , r-release <u>FertNet 0.1.1.tgz</u> , r-oldrel (x86_64): <u>FertNet 0.1.1.tgz</u>
Linking:	
Please use the cano	nical form <u>https://CRAN.R-project.org/package=FertNet</u> to link to this page.



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

© 2023 Gert Stulp.

This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.

See https://creativecommons.org/licenses/by/3.0/de/legalcode.

64): FertNet 0.1.1.tgz, r-release (x86_64):





...%?









Am I data dredging!?



Am I data dredqual?

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 Check for updates

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 Jol

General Article

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



Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley



Psychological Science 22(11) 1359-1366 © The Author(s) 2011 Reprints and permission: sagepub.com/journalsPermissions.nav DOI: 10.1177/0956797611417632 http://pss.sagepub.com



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

