Development of a persons network of the Netherlands

Jan van der Laan <dj.vanderlaan@cbs.nl>
Marjolijn Das
Edwin de Jonge

March 8, 2022

Why?

- Research and Development Department
  - Develop better ways of making statistics
  - Develop new statistics
- Network Science
  - Describe The Netherlands as a system

Person Network of The Netherlands

Date

October 2018

Data

Population Register
Income Tax Data
Business Register
School pupils
Students
...

→

Network

17.2M Vertices

Layers	#edges	$< k_{out} >$
Family	270M	15.7
Household	41M	2.4
Neighbours	353M	20.4
Colleagues/Work	566M	32.8
Classmates/School	234M	13.6

Family layer

The population register registers the parents of a person.

$$A \xrightarrow{\text{parent}} B \implies B \xrightarrow{\text{child}} A.$$

$$A \xrightarrow{\text{parent}} B \xrightarrow{\text{parent}} C \implies A \xrightarrow{\text{grand parent}} C.$$

$$A \xrightarrow{\text{parent}} B \wedge C \xrightarrow{\text{parent}} B \wedge A \neq C \implies A \xrightarrow{\text{sibling}} C.$$

$$A \xrightarrow{\text{parent}} B \xrightarrow{\text{sibling}} C \implies A \xrightarrow{\text{aunt/uncle}} C.$$

School layer

Table: Variables used to define classes.

Type of school	Variables defining a class
Primary education (Basisschool)	School id, location id, year
Secondary education (Voortgezet onderwijs)	School id, location id, type of education, year
Secondary Special education (Speciaal voortgezet onderwijs)	School id, location id, number of years followed
Vocational (MBO)	School id, type of education, number of years followed
Higher education (HBO, University)	School id, location id, type of education, number of years followed

Other layers

Household layer
Households are derived at CBS from the population register, address information, and additional sources.

Neighbour layer
The 10 closest (geograpic distance) households within 50 metres.
In case of ties households are randomly selected.
This network is therefore not symmetric.
Institutional households >4: each person seperate household.

Work layer
Persons working at the same company are colleagues.
When > 100 colleagues: select 100 closest (geographic) colleagues.

Using the network

The data - Triplets

rinpersoon_src		rinpersoon_dst		relation
R	111111111	R	222222222	Parent
R	111111111	R	333333333	Parent
R	111111111	R	123123123	Class-mate
R	111111111	R	234234234	Class-mate
R	111111111	⋮	⋮	⋮
R	333333333	R	111111111	Child
R	333333333	R	222222222	Co-parent
R	333333333	⋮	⋮	⋮
R	222222222	R	111111111	Child
R	222222222	R	333333333	Co-parent
R	222222222	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

Simple Analysis

Are persons more likely to have solar cells installed when their social circle had solar cells installed?

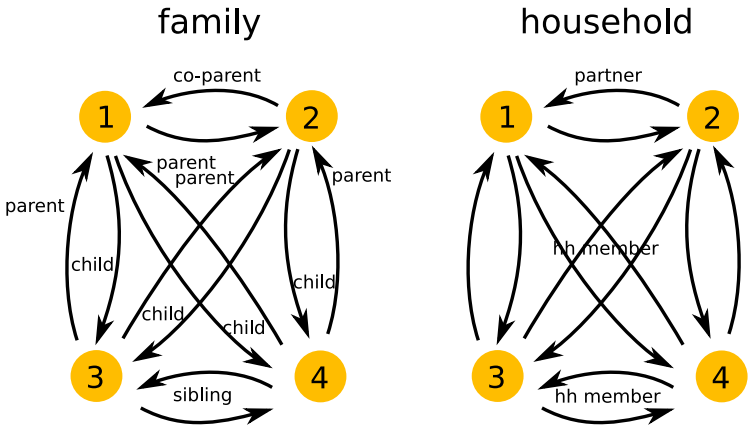
rinpersoon_src		rinpersoon_dst		has_cells_dst
R	111111111	R	222222222	1
R	111111111	R	333333333	0
R	111111111	R	123123123	0
R	111111111	R	234234234	1
R	333333333	R	111111111	1
R	333333333	R	222222222	0
R	222222222	R	111111111	0
R	222222222	R	333333333	0
⋮	⋮	⋮	⋮	⋮

Simple Analysis

Only look at direct relations.

rinpersoon_src		fcells_dst	age	homeowner	cells
R	111111111	0.5	20	0	0
R	333333333	0.5	35	1	1
R	222222222	0.0	48	1	1
⋮	⋮	⋮	⋮	⋮	

Multiple links



Our plans — Segregation on education attainment

Segregation

Is a network problem

— Who has access to whom?

— Who receives information/ideas from whom?

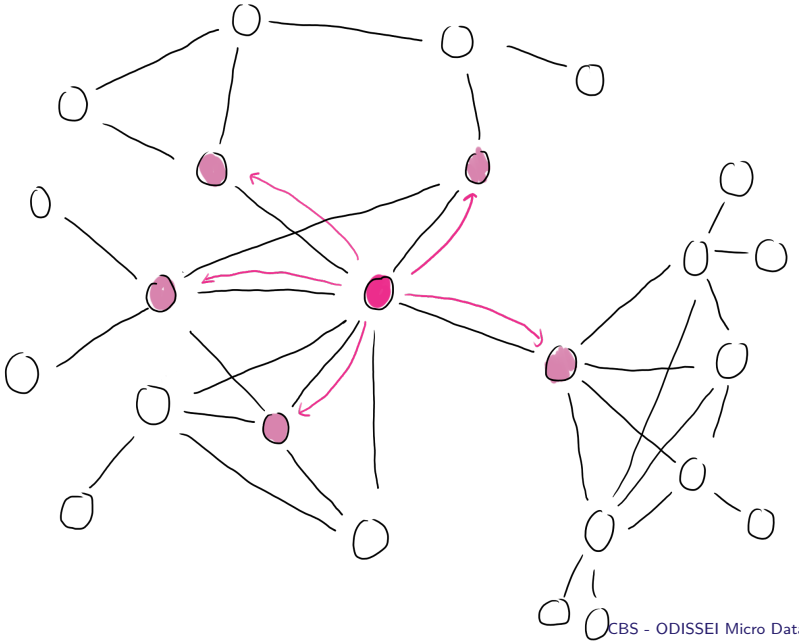
To what extent are persons exposed to persons from different groups than themselves.

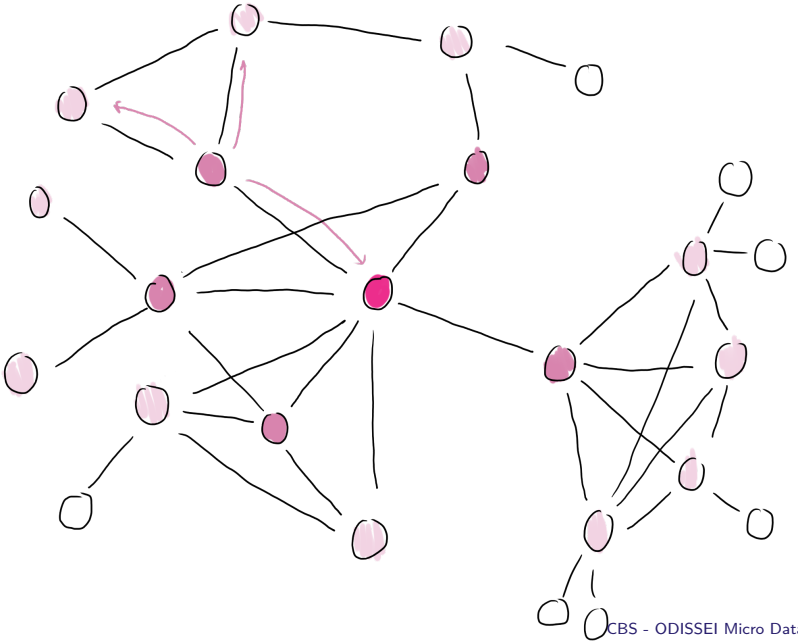
— Migration background

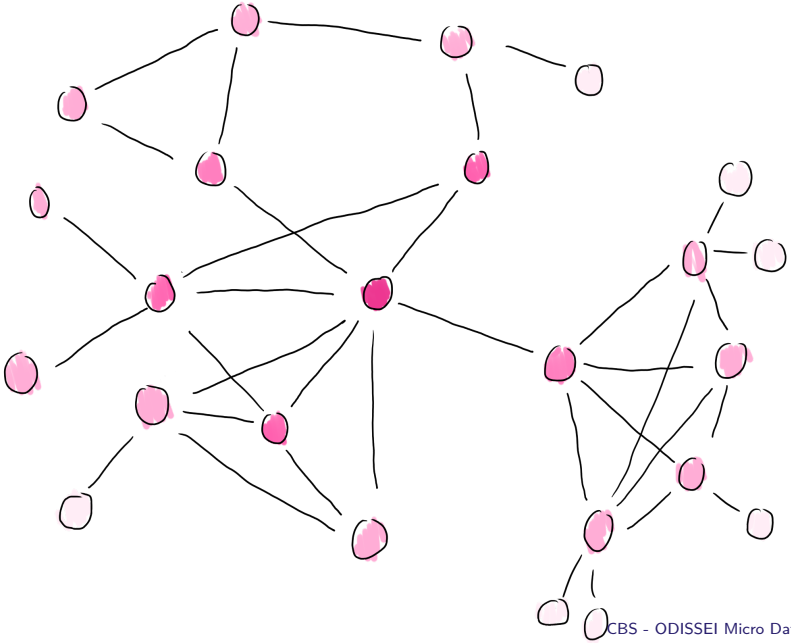
— Education level

— ...

Indirect contacts are important







Conclusion

Working on time series of regional statistics on segregation on educational attainment.

Further development and fine tuning of network derivation.

Collaboration with POPNET and ODISSEI on making the data easier to use by researchers.