



Analyse Sociaal Netwerk Nederland

Jan van der Laan <dj.vanderlaan@cbs.nl>

Edwin de Jonge <e.dejonge@cbs.nl>

Marjolein Das <m.das@cbs.nl>

2 oktober 2018

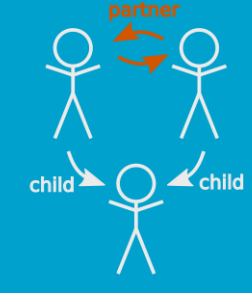
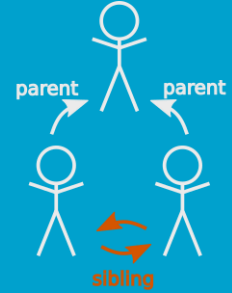
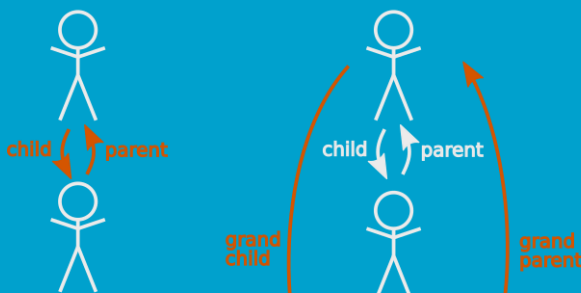
Onderzoeksthema complexiteit

Netwerkanalyse

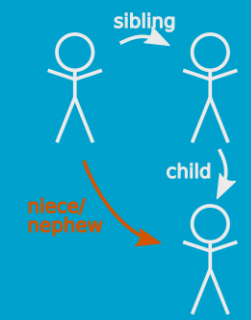
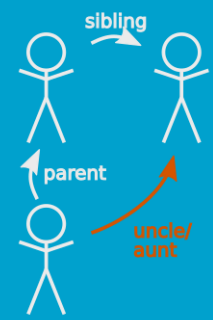
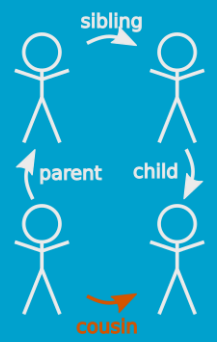
- Afleiden statistieken uit netwerk
- Persoonskenmerken afleiden uit netwerk
 - Achtergrondkenmerken in verdere analyses
 - Statistieken

Huidig onderzoek

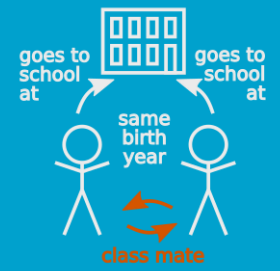
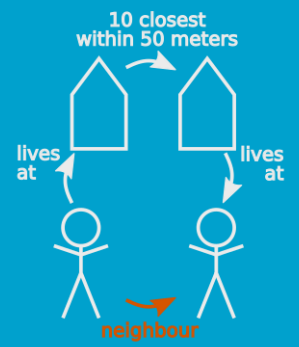
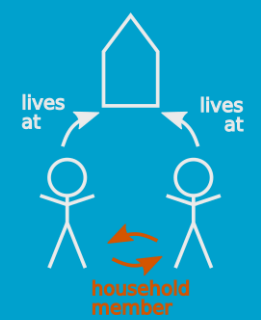
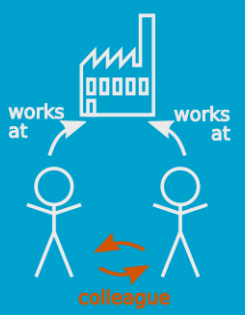
- Segregatie
- Sociaal Kapitaal



16.9 mln nodes



39 mljrd links



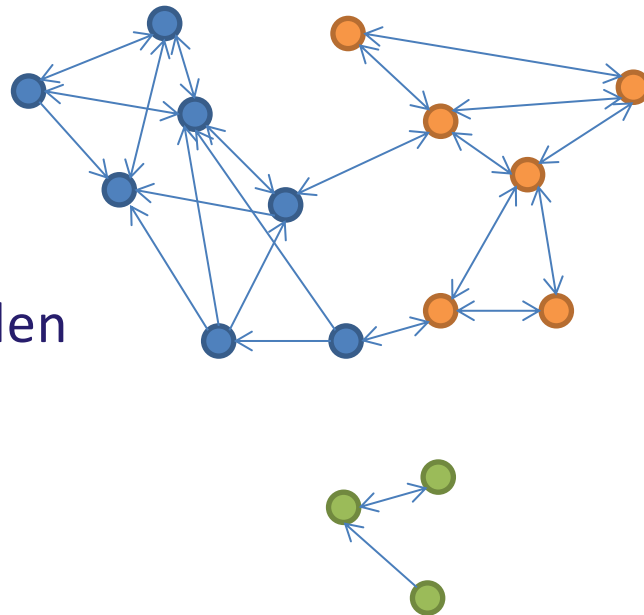
gem 1800 links pp



Community detection

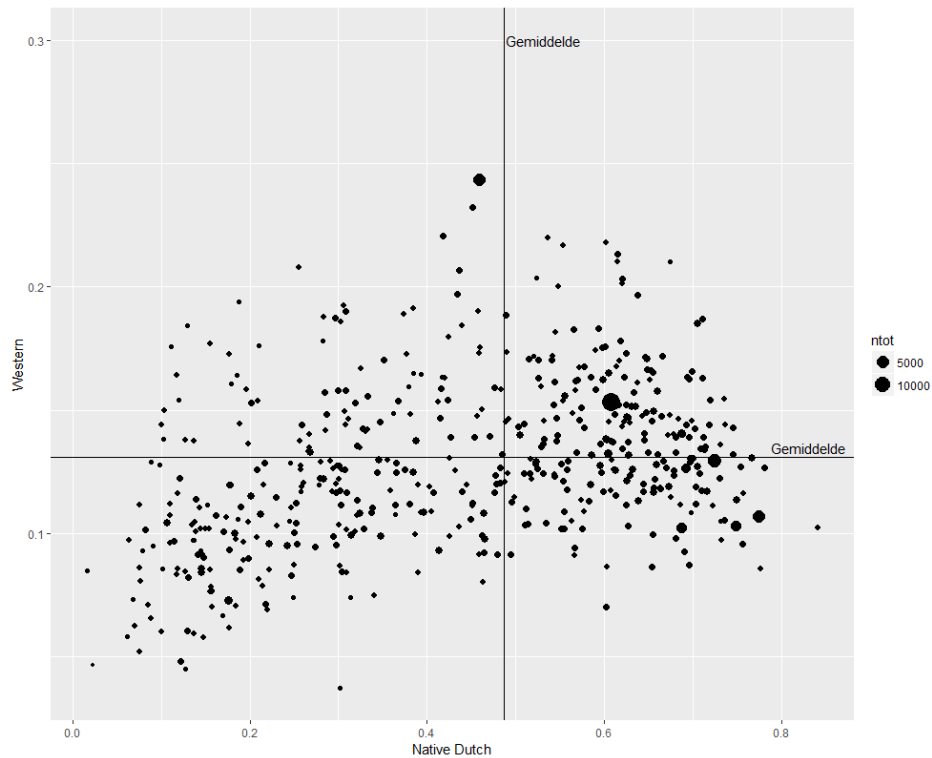
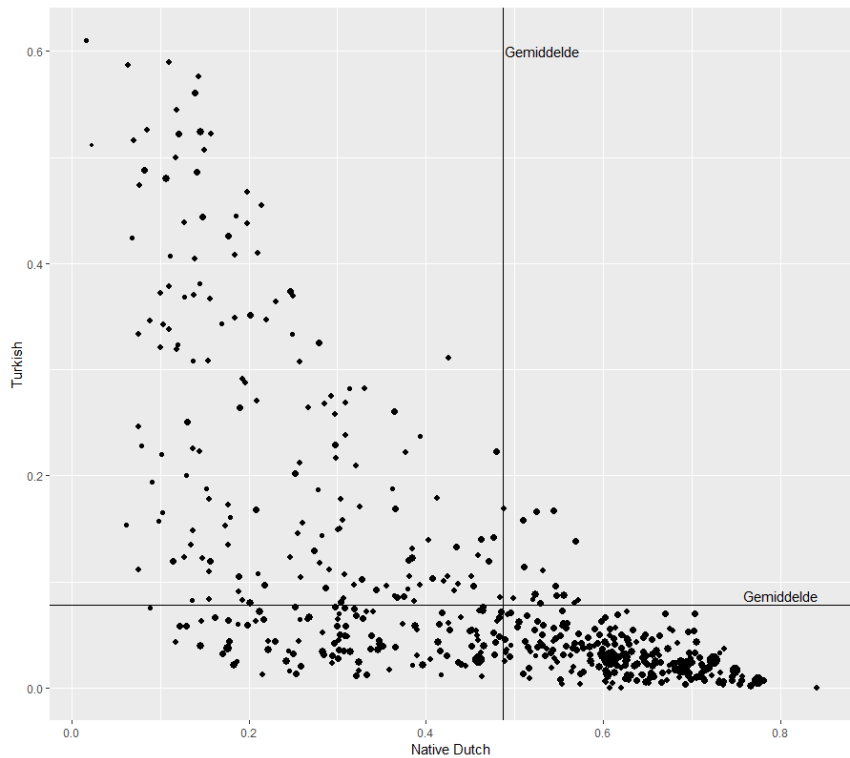
Doel: detecteren groepen mensen die onderling sterker verbonden zijn dan met rest van inwoners:

- Sturen kinderen naar zelfde scholen
- Trouwen onderling
- Werken bij zelfde bedrijven
- ...



Vorm van segregatie

Rotterdam - communities



Toekomst

Uitbreiding naar heel Nederland

Andere maten voor segregatie

- Op persoonsniveau: in welke mate is men verbonden met gelijksoortige personen

Gebruikte software/tools

- R:
 - igraph-package¹
 - Zelf geschreven code
- C++ implementaties van algoritmes
 - Hebben POSIX omgeving nodig (memory mapping)
 - Make/BASH

¹ en een hele berg andere packages

Hardware/OS

- Geheugen
 - Community detection (In R 64 GB niet genoeg voor Amsterdam; in C++ ca 25 GB voor Amsterdam)
 - Individuele segregatie: eigenwaardes van adjacency matrix (ca. 15Mx15M, maar sparse: 800M elem.)
- Huidige implementaties kunnen zeer beperkt gebruik maken van parallelisatie
 - Eén methode gebaseerd op uitrekenen eigenwaardes (is er bijv. parallelle BLAS? Integratie met R?)



Samenvatting

Data:

- Vertices/nodes: 17M records; 2 GB
- Edges: 800M records¹; 27 GB
- Code

Nodig:

- Linux (geheugen wrsch > 128 GB; wrsch 256 GB genoeg)
- R (+ evt R-Studio) + set packages
- C++ compiler + tools zoals Make/BASH
- (Python3 + NetworkX)

¹ als we het aantal collega's beperken; anders factor 50 groter



Vragen

- Is er een grafische omgeving?
 - Analyse van resultaten binnen R/RStudio: plots etc.
- Is het mogelijk/makkelijk om code te importeren?
 - Ontwikkeling van C++/R-code buiten remote access omgeving
- Hoe werkt installatie van R-packages?
- Importeren van hulpbestanden (bijv. kaartmateriaal)?
- Zijn zaken als geheugen en #cores flexibel?

