Constructing a digital twin of the social network of Amsterdam

Presentation outline

- Data explanation
  - Data sources main data structure
  - Layers
  - Characteristics
  - Structure
- Static network
  - Basic static network
  - Extensions
- Homophily
  - SDA Model
  - Our model
- Dynamics

Data

- Network data of Amsterdam
- Source: CBS 2019
- Included 4 network layers:
  - Household
  - Family
  - Neighbours
  - Work / School relationships
- Include 240 groups based on person characteristics

Computational Science

Layers

- Household
  - Household relationships
  - Symmetrical
  - Include:
    - Partner,
    - Housemate,
    - Housemate institute
- Family
  - Family relationships
  - Symmetrical
  - Include:
    - Aunt/uncle,
    - co-parent,
    - sister/brother

- Neighbours
  - 10 closes neighbour households
  - Not Symmetrical
  - Randomly chosen if multiple neighbours have same distance

- Work/School
  - Work, school and university relationships
  - Not symmetrical
  - If more than 100 persons work in the company the 100 geographically closest persons are chosen

Computational Science

Person characteristics

- 240 groups based on person characteristics:
  - Age group: [0-20), [20-30) … [80-120]
  - Ethnicity: Native, Moroccan, Turkish, Surinamese, Other
  - Education: 1, 2, 3
  - Gender: Man, Woman

Data structure

- 2 different kind of datasets:
  1. Agent dataset
  2. Connection dataset

Age	Ethnicity	Education	Gender	Amount(n)
[0-20)	Native	1	Man	30829
[50-60)	Moroccan	3	Woman	105

Age Source	Ethnicity Source	Education Source	Gender Source
[0-20)	Native	1	Man
[50-60)	Moroccan	3	Woman

Age Dest	Ethnicity Dest	Education Dest	Gender Dest	Amount(n)
[20-30)	Moroccan	1	Man	600
[0-20)	Moroccan	1	Woman	440

Computational Science

Creation basic static network

- Creating a static network based on datasets
- Basic random network

Algorithm 1 Initializing of social network

```
for row in Connections data frame do
    S ← Source group nodes
    D ← Destination group nodes
    C ← Connections (n)
    while i < C do
        N_S = random.choice(S)
        N_D = random.choice(D)
        if N_S ≠ N_D And N_D notin links(N_S) then
            Add Directional edge from N_S to N_D
            if r > U(0, 1) And N_S notin links(N_D) then
                Add Directional edge from N_D to N_S
                if S = D then
                    i += 1
                end if
            end if
        end if
        i += 1
    end while
end for
```

Computational Science

Making extensions on the static network

- Adding a scale-free parameter where 1 will give a scale-free network and 0 a normal distributed network (and exponential between 0 and 1)
- Adding neighbourhoods
- Adding Households from household data
- Overlapping the household and family network

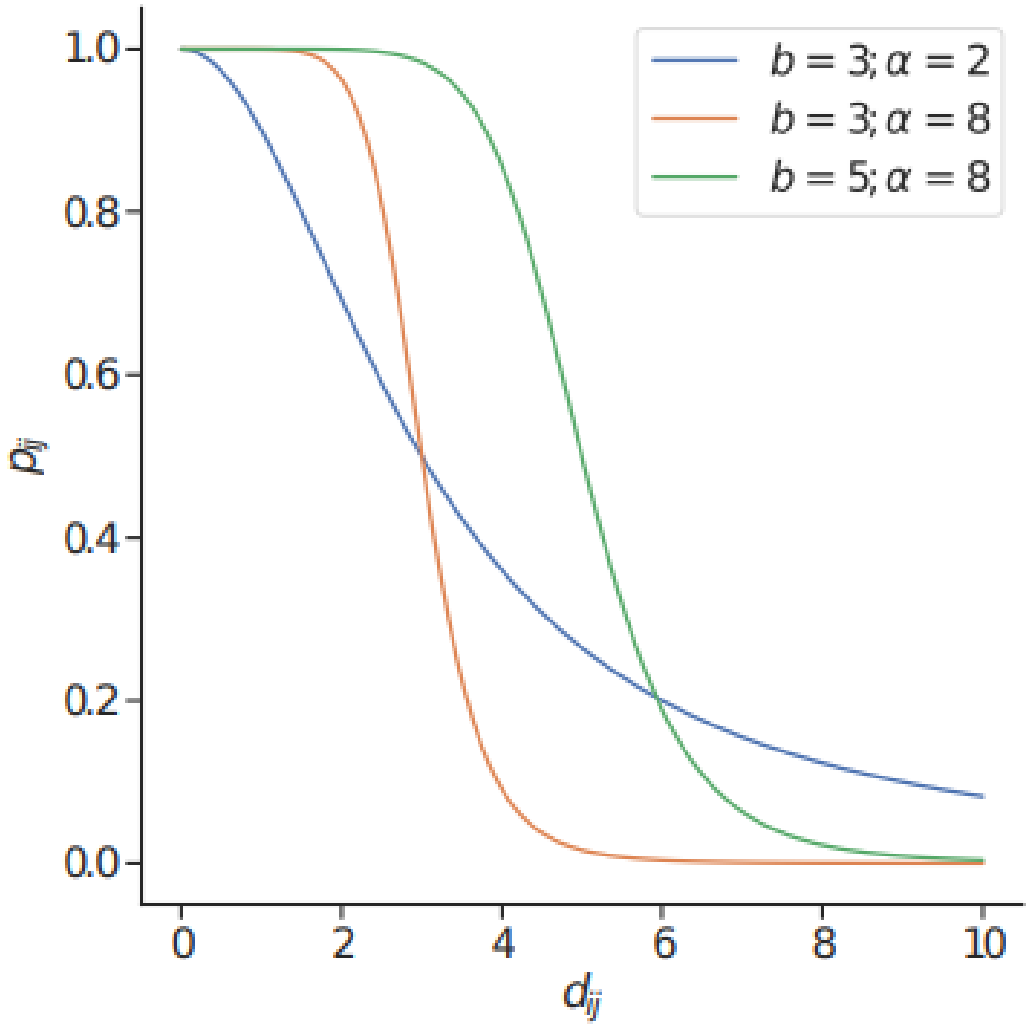
Homophily

- Social structure and social networks are related through the fundamental principle of homophily
- Agents that are similar with respect to some significant social features are more likely to be somehow connected than dissimilar agents

- How this idea would be presented in the data

Social Distance Attachment model

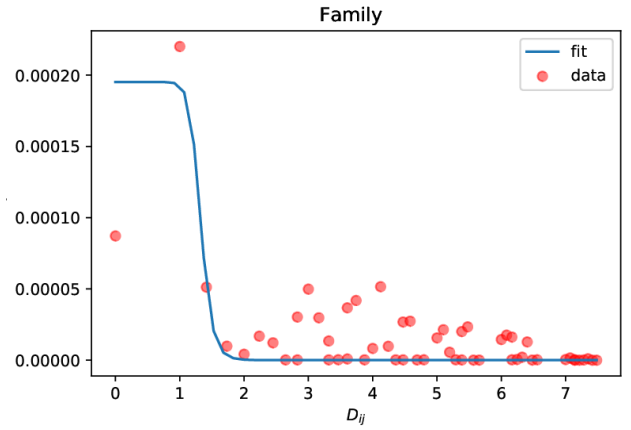
- $p_{ij} = \frac{1}{1+[b^{-1}d(x_i,x_j)]^{\alpha}}$
- p_{ij} is the probability of connection between i and j
- b is the characteristic distance
- d is the distance between i and j
- α is the level of homophily

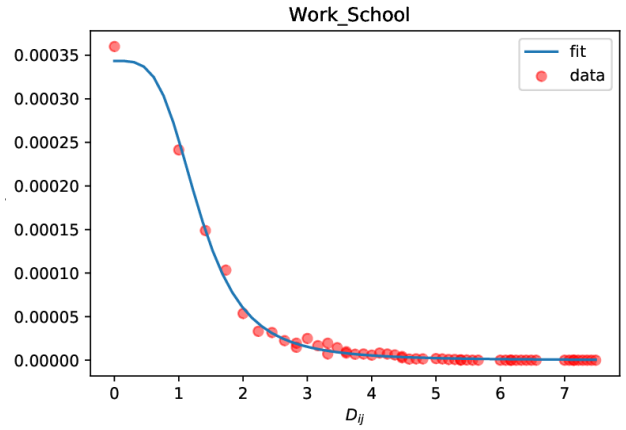


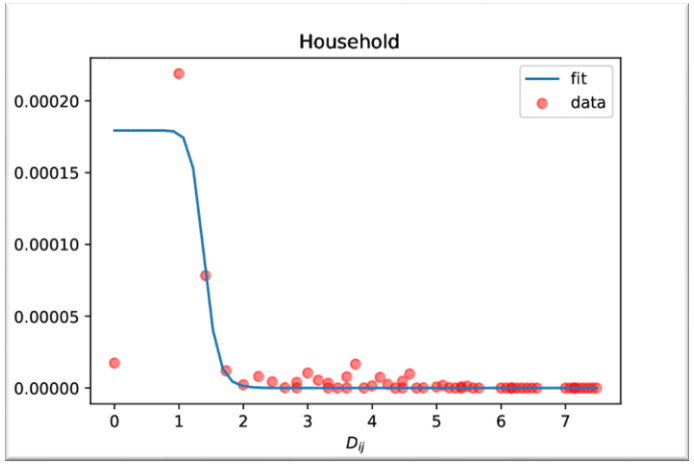
Computational Science

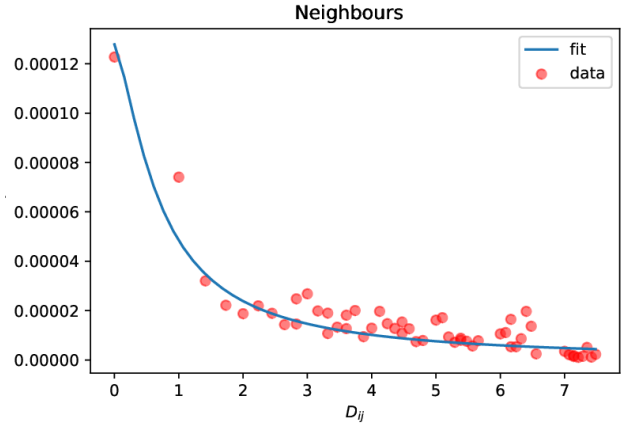
Using this model to fit the data

- $p_{ij} = \frac{a}{1+[b^{-1}d(x_i,x_j)]^{\alpha}}$
- a is a free variable which is the maximum probability in the original formula





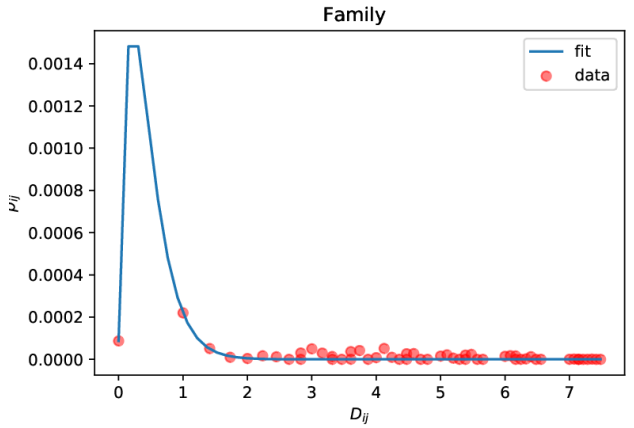


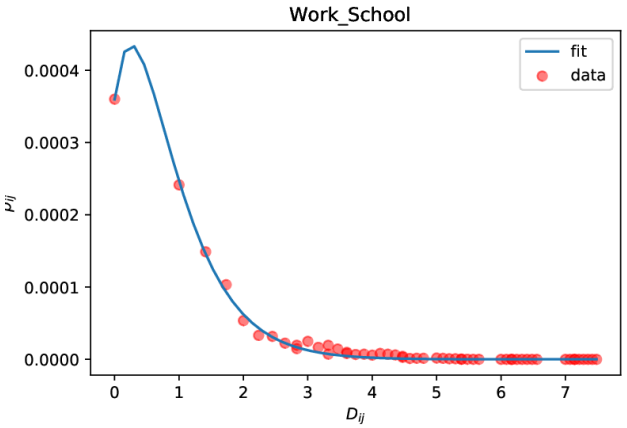


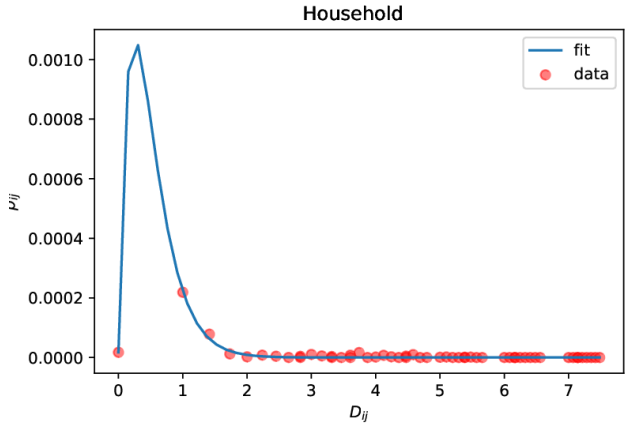
Computational Science

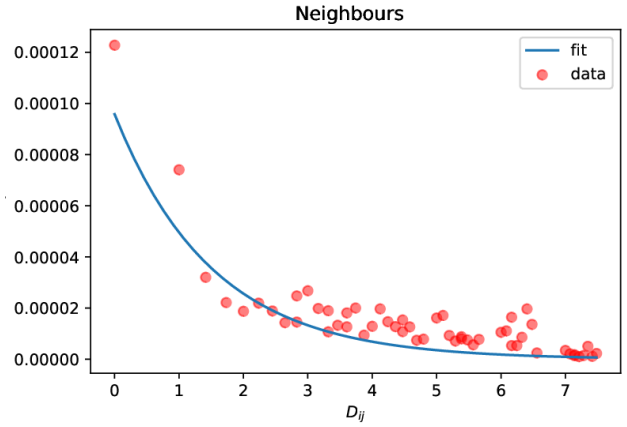
Redefining the formula

- Formula does not fit all layers
- Rewrite so it can fit all layers
- $p_{ij} = (a + \beta * D_{ij})e^{-\alpha * D_{ij}}$









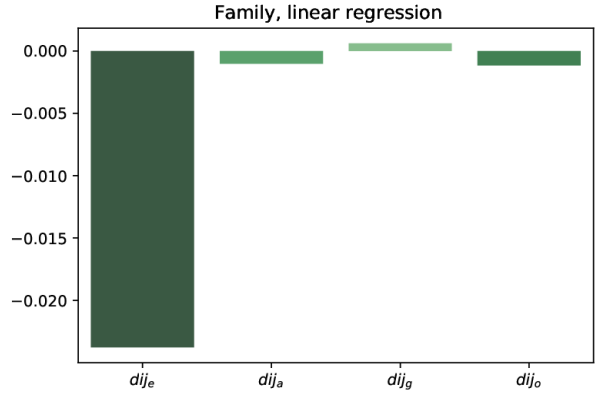
Computational Science

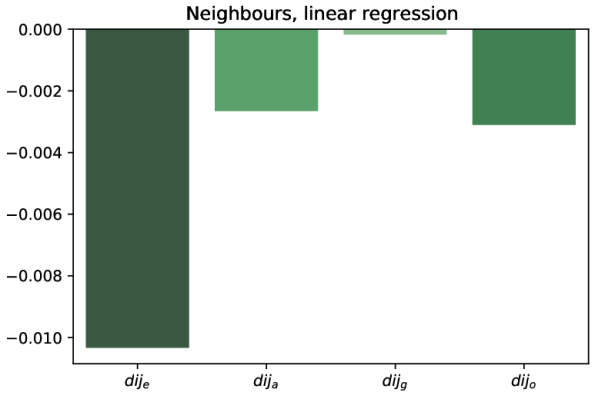
Weighted function

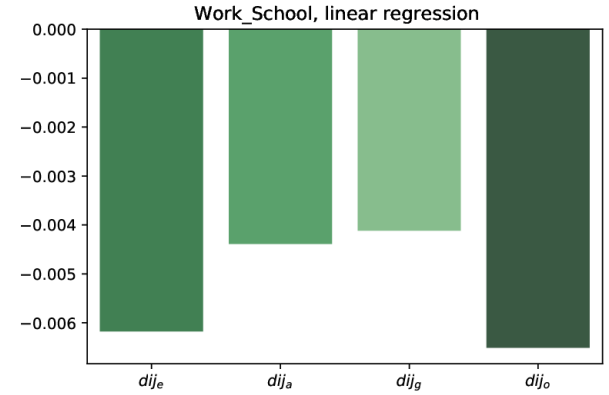
- Distance of agents are based on different characteristics
- Not all characteristics evenly important (McPherson, Smith-Lovin & Cook 2001)
- Weighted function is introduced to find the characteristic importance
- $p_{ij} = \left(a + \beta * \sum_c^n W_c D_{ijc}\right) e^{-\alpha * \sum_c^n W_c D_{ijc}}$

Results weighted function

- Ethnicity
  - Family , household & Neighbours ➔ biggest influence
  - Work/School ➔ Second highest influence
- Age
  - Family , household & Neighbours ➔ Medium influence
  - Work/School ➔ High influence
- Gender
  - Family, household ➔ Positive influence
  - Work/School, Neighbours ➔ Lowest influence
- Education
  - Family, household, neighbours ➔ med high influence
  - Work/school ➔ Highest influence







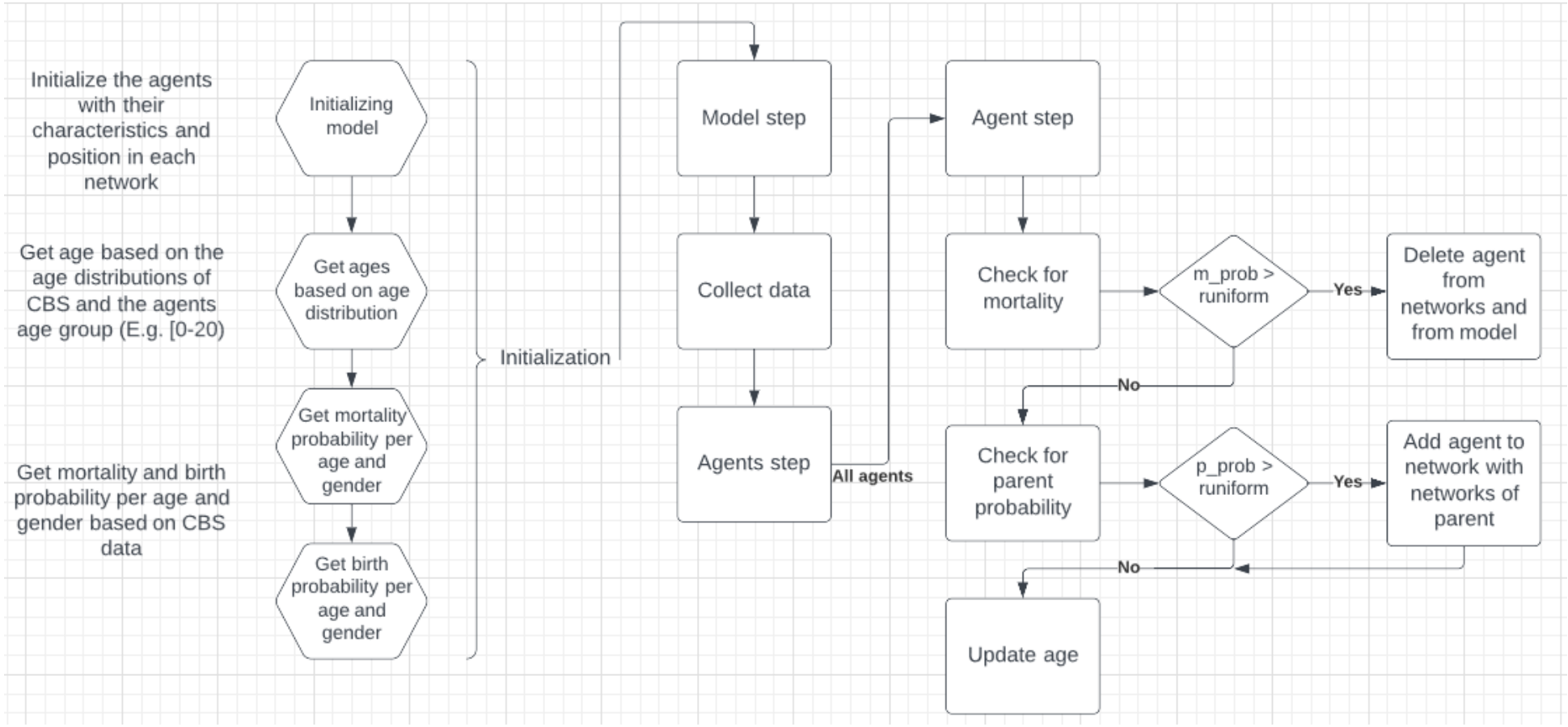
Computational Science

Dynamics

- An Agent-based approach is used to include the dynamics

- Each person is an agent in the Agent-based model
- Use Static networks as starting point
- Use probability function as for making new connections
- Introducing other dynamics such as death and birth

Agent based model flow chart



Computational Science

Discussion