University of Amsterdam

Masters Thesis

Constructing a digital twin of the social network of Amsterdam

Author: Kamiel Gülpen
First name SURNAME

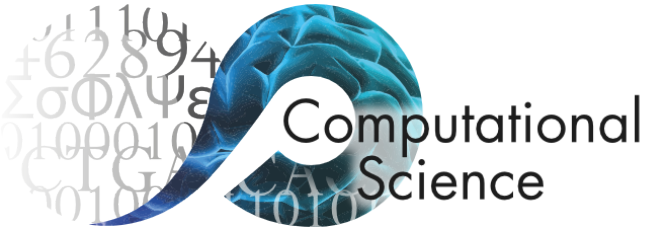
Supervisor:
supervisor name

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Computational Science

in the

Computational Science Lab
Informatics Institute

May 2022



Declaration of Authorship

I, First name SURNAME, declare that this thesis, entitled 'Your Thesis Title' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the University of Amsterdam.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Your signature

Date: 1 August 2020

"The world works as a network, we function within networks and there is no way we can understand this world without using network science."

Albert-László Barabási.

UNIVERSITY OF AMSTERDAM

Abstract

Faculty of Science
Informatics Institute

Master of Science in Computational Science

Your Thesis Title

by First name SURNAME

Include your abstract here Abstracts must include sufficient information for reviewers to judge the nature and significance of the topic, the adequacy of the investigative strategy, the nature of the results, and the conclusions. The abstract should summarize the substantive results of the work and not merely list topics to be discussed.

Length 200-400 words.

Acknowledgements

Thank the people that have helped, supervisors family etc.

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations

CSL	**C**omputational **S**ceince **L**ab
UvA	**U**niversitiet **v**an **A**msterdam
ER	**E**rdos **R**enyi
BA	**B**arabasi **A**lbert

Chapter 1

Introduction

The analysis of complex networks, often known as network science, is a multidisciplinary discipline that arose primarily from sociology and graph theory in mathematics. Fundamentally, it is concerned with the understanding of relational phenomena through the study of large networks of interconnected components, and its applications include physics, bioinformatics, epidemiology, eco- nomics, sociology, psychology, and many other fields. Among the numerous network models available, a subgroup of social network models focuses on how individuals interact in diverse fields. As a result, these models make various attempts to reflect the structure of social networks [12]. These social networks provide one of the most useful analytical and theoretical frameworks for studying social phenomena, such as such as the genesis of homophily[15], reciprocity[13] and transitivity[10].

Mining and analyzing social networks in diverse situations yields crucial insights for improving fundamental knowledge and understanding of human behavior [20]. Researchers have been able to explore social theories and effects such as the genesis of homophily[15], reciprocity[13] and transitivity[10] using data from social networks. Data also aids in the development of data-driven models of human interactions, which may be used to characterize the many processes occurring in a particular population, such as information dissemination, coordination, consensus building, the transmission of contagious diseases or the creation and breakdown of friendships. Accurate descriptions of social interactions are thus critical for shedding light on the most significant mechanisms at work in these processes, such as understanding the elements influencing whether a rumor will spread or what the best strategies are to control the spread of a virus or contagious diseases such as obesity. Network Science characterizes network structures to increase our understanding of complex systems, as it is assumed that the underlying network structure of a complex system encodes information about its function.

Real-world social networks have a set of distinctive structural traits that do not appear as frequently in other types of complex networks (i.e. technological, informational, biological). They are often sparse, with non-trivial clustering coefficients, positive degree assortativity, and low average path lengths (small-world property) [18]. As a result, it is critical to build social networks with these ideas in mind.

1.1 Research questions

Can we make a usable and relatively accurate representation of the social network of Amsterdam?

What are the main drivers of connection making in different layers of the social network of Amsterdam?

How will the social network of Amsterdam look like in the future?

1.2 Research approach

The work proposed in this thesis is based on data obtained from the Central Bureau of statistics (CBS) which has network data on four layers: households, families, nighbours and work/school relations. Three separate but subsequent studies are conducted using this data.

The first study investigates how a detailed and representative network of Amsterdam can be built with the CBS data. During this study a family of networks will be presented which can help legislation and other research where there is a need for the social network of Amsterdam. This family of networks are made by using different algorithms such as random network generation or scale free network generation.Following the introduction of these algorithms and the development of a more generic version of the networks, we can move on to the next phase, which is zooming in on how the social networks appear at various levels using existing literature and other data, which enabled the development of a more sophisticated and diverse family of Amsterdam social networks, allowing laws, interventions, and other network dynamics to be tested across multiple networks.

The second study is a more in-depth data examination/data-analysis of the CBS data. The main question we want to answer here is understanding on which basis people connect with one another on different layers of Amsterdam. To answer this question we rely heavy on the existing theory of homophily and try to contribute to the existing

literature around this. Based on this theory we tried to develop a model which depicts the dynamics of human connections in Amsterdam.

The final study looks at how we can make Amsterdam's social network more dynamic by combining the ideas from the first two studies to see how the network evolves over time.

1.3 Research contribution

The provided network can be used to investigate the probable mechanisms of the establishment and maintenance of various attitudes (social norms) about body weight within groups of varying socioeconomic level (SES)[9].

Furthermore, complex network analysis of the network can assist in identifying the reasons of such socioeconomic inequality, in addition an ABM based on this network allows for the investigation of potential interventions.

Chapter 2

Data

- The proposed network is based in data from CBS

- THerfore we need to describe the data in detail

- First a general data description, where does the data come from, what are the thoughts behind the data, next up data investigation, what does the data say

- Spatial data?

- Data about households

2.1 Data description

- Describe data based on drive file

2.2 Data investigation

- Show description of the data: amount of links per layer, how many nodes are connected, average amount of connections per node.

- Show interesting plots and result on how data is laid out, what are interesting plots? Maybe show heatmap plots?

- Maybe already show things such as Theil's index gebasseerd op het vraagstuk van segegratie

- Theils index - Segregation index - Other indexes

2.3 Data conclusions

- Give general conclusion on the data - Waar zijn we mee bezig?

Chapter 3

Constructing a network of Amsterdam

3.1 Introduction

Every city on the planet exhibits collective behavior. In 2019, for example, there was a storming of the Capitol in Washington and weekly protests on the Amsterdam Dam. To influence these collective behaviors, it is necessary to first understand them and how they emerge.

Most collective behaviors spread through social contact. From the emergence of social norms (Centola, Willer, and Macy 2005), to the adoption of technological innovations (Coleman, Katz, and Menzel 1966), to the growth of social movements (Marwell and Oliver 1993; Gould 1991, 1993; Zhao 1998; Chwe 1999), social networks are the pathways along which these "social contagions" propagate. Studies of diffusion dynamics have demonstrated that the structure (or topology) of a social network can have important consequences for the patterns of collective behavior that will emerge (Granovetter 1973; Newman, Barabasi, and Watts 2006). Accurate descriptions of social interactions are thus critical for shedding light on the most significant mechanisms at work in these processes. [6–8].

Although this type of structural approach to social understanding is still a very powerful tool for uncovering the hidden structures underlying social activities, it has become increasingly clear in recent years how a monodimensional analysis is unable to account for an increasing number of phenomena [14]. A multilayered network gives more insight into such phenomena and is therefore more interesting to study and construct.

This paper proposes family of interaction networks in the city of Amsterdam. The proposed network consists of four layers: households, family, neighbors, and work/school ties.

These layers represent the amount of connections between groups, as indicated in Chapter 2; however, in this study, the network will be constructed on meso level instead of macro level, with agents connecting to each other rather than merely the groups which they represent. In order to achieve this multiple theories and algorithms have to be stuedied to make a

3.2 Social networks

Network science aims to build models that reproduce the properties of real networks. Most networks we encounter do not have the comforting regularity of a crystal lattice or the predictable radial architecture of a spider web. Rather, at first inspection they look as if they were spun randomly (Figure 2.4). Random network theory embraces this apparent randomness by constructing and characterizing networks that are truly random.

Engenieering artificial social networks Determining how to construct an artificial social network may be challenging since the underlying structure, and hence its attributes, vary substantially based on the approaches that are used to construct such a network [5]. Asserting the qualities the intended network should have is thus necessary before selecting an algorithm. The many qualities of a network and which methods build networks with these features will be addressed in this section.

3.2.1 Random Networks

A random network is a network in which nodes connect to each other in a random manner. The mathematicians ál Erdős and Alfréd Rényi have played an important role in not only understanding the properties of these networks, but also developing an algorithm that creates them[11]. Their method for generating a random network is widely used. As a result, a random network is also known as an Erdős-Rényi network.

A random network consists of N nodes where each node pair is connected with probability p. To construct a random network we follow these steps:
1) Start with N isolated nodes.
2) Select a node pair and generate a random number between 0 and 1. If the number exceeds p, connect the selected node pair with a link, otherwise leave them disconnected.

3) Repeat step (2) for each of the $\frac{N(N-1)}{2}$ node pairs.

A random network has a binomial distribution, well approximated by a Poisson distribution in the $k << N$ limit.

$$\left(\frac{N-1}{k}p^k(1-p)^{N-1-k}\right) \tag{3.1}$$

$$p_k = e^{-<k>}\frac{<k>^k}{k!} \tag{3.2}$$

$$N(d) \approx 1+<k>+<k>^2+...+<k^d> = \frac{<k>^{d+1}-1}{<k>-1} \tag{3.3}$$

N(d) must not exceed the total number of nodes, N, in the network. Therefore the distances cannot take up arbitrary values. We can identify the maximum distance, d_{max}, or the network's diameter by setting:

$$N(d_{max}) \approx N \tag{3.4}$$

assuming $<k>>> 1$

3.2.2 Small-world networks

3.2.3 Scale free network

Begin with m_0 nodes, with links between them picked at random, as long as each node has at least one link. The network evolves in two stages, which are: 1. Growth, Real networks are the product of a continual growing process that increases N. Therfore, networks grow indefinitely by adding new vertices. At each timestep we add a new node with $m(\leq m_0)$ links that connect the new node to m nodes already in the network 2. Preferential Attachment, new nodes tend to connect to more linked nodes. In contrast, nodes in random networks pick their inter - action partners at random. So new vertices attach preferentially to previously well-connected locations. The probability $p(k)$ that a link of the new node connects to node i depends on the degree k_i as

$$p(k_i) = \frac{k_i}{\sum_j k_j} \tag{3.5}$$

[1, 16].

node distribution is

$$P(k)_{scale-free} = k^{-\gamma} \tag{3.6}$$

3.2.4 Real-world social networks

After stirring up the game other scientist have also looked at this problem with different tools, one Other characteristics of real social networks are: - Real world social networks s. Furthermore, social networks are at best weakly scale free

n contrast, social networks present a different picture. Like the corpus overall, half of social networks lack any direct or indirect evidence of scale-free structure (50% Not Scale Free; Fig. 5b), while indirect evidence is slightly less prevalent (41% SuperWeak). The former group includes the Facebook100 online social networks, and the latter includes many Norwegian board of director networks. However, among the categories representing direct evidence of scale-free structure, more networks fall into the Weakest (48and Weak (31%) categories, but not a single network falls into the Strong or Strongest categories. Hence social networks are at best only weakly scale free, and even in cases where the power-law distribution is plausible, non-scale-free distributions are often a better description of the data. The social networks exhibiting weak evidence include many scientific collaboration networks and roughly half of the Norwegian board of director networks[4].

3.2.5 Multilayer networks

- Random networks

3.3 Amsterdam as a social network

- Amsterdam segregation is only based on spatial data not on data such as work relationships or family ties

- This network will give a deeper understanding how the social network in Amsterdam lays out and can show emergent characteristics

- Other research is done based on assumptions of social networks in Amsterdam

- This network can be used for further research where a social network of Amsterdam is acquired.

3.4 Network algorithms

In this section, we will go over all of the methods used to create a family of networks that represent Amsterdam's social network. First, the agents' initialization is discussed, including the algorithm that drives this initialization and rationale behind certain choices. After the explaination of the agent initialization the network realizing algorithms, that connect the previously initialised agents, are discussed. These algorithms are built on the random networks and scale-free networks discussed in the previous section. Lastly other types of data that can be used to make a network more representative of Amsterdam are investigated and implemented into the network.

3.4.1 Agent initialization

As mentioned in Chapter 2, an agent is a person in Amsterdam who exhibits four distinct characteristics: age group, ethnicity, gender, and educational level. To build a network with all of the agents represented as nodes, we must walk through all of the groups and all of the nodes in each group, remembering the characteristics of each node. This algorithm's pseudocode is given in algorithm 1. The main objective of this algorithm is to got agent ID - agent group pairs such each agent has a agent ID and a group ID. We can later use these group ID's to make the given connections that are represented in the data.

The initialization of the nodes begins with setting an agent ID to 0, which is the first ID received by an agent when the initialization begins. Following the initialization of the first agent, we loop through the entire agent data frame (see Chapter 2). In each row, we can determine the group size as well as the characteristics of that particular group. To speed up the modeling process, a hash function is used to hash the group into an integer, which can be seen as it's group ID. For example, the group with the characteristics: Male, [0-20], Native, 1, it will have the group ID 1 etc. This simple hash has significantly sped up the entire model.We use the number of nodes to create a for loop with the range function so that we can go through all of the agents in the data frame and add them to a dictionary with the group ID as the key. This way each all the agents of a certain group are grouped together which makes it possible to sample

agents from a certain group. The agent ID is updated when the agent is added to the it's group. This will continue until all agents have been initialized and joined a group.

Algorithm 1 Initializing of random network between groups

$ID_{agent} = 0$
for row in Agent data frame **do**
  N ← Amount of agents in group
  $C_{ethnicity}$ ← Ethnicity of group
  $C_{agegroup}$ ← age group of group
  $C_{education}$ ← education of group
  C_{gender} ← gender group of group
  hashroup = hash(($C_{ethnicity}$, $C_{agegroup}$, $C_{education}$, C_{gender}))
  nodes[hashroup] = array
  for agent in range(N) **do**
    add ID_{agent} to nodes[hashroup]
    ID_{agent} += 1
  end for
end for

3.4.2 Network generators

To obtain the first set of networks, a random graph algorithm is used to connect the network's nodes across all four layers. These networks are constructed using the data described in Chapter 2. The data shows the number of connections that exist between two groups. To completely connect a layer, the predetermined number of connections between all the listed groups must be implemented. A random network generator is implemented to successfully generate such a network. First we loop through the dataframe, where each row gives information about the 2 groups that are connected with each other and the amount of connections there are between these two groups.

Secondly a source node is chosen by taking a random sample from the source group bin. This works a bit different for the destination node. The destination node is randomly picked from a selection bin where a fraction, k_f, of the amount of nodes from the original destination bin are represented. A random sample is taken from the selection bin to choose the desiantion node. After both nodes are chosen there will be checked if they are not the same node and if they do not already have a connection, if both conditions have been met there will be layd a edge between the two nodes. After the connection between the two nodes is made the destination node will be added again to the destination bin together with a randomly chosen node with the probability $1 - k_f$. This algorithm can be observed in Algorithm 3. As we can see Algortihm 3 the in-degree distribution is heavily depended on the parameter k_f, and it's effect is described in equation 3.7, where it can be observed that if the parameter k_f is equal to 1 a

random ER network will be generated and if k_f is equal to to $\frac{1}{N_{group}}$ (which is an initial destination bin of 1 node) the network distribution will be scale-free.

$$P(k) = \begin{cases} P_{random}, & \text{if } k_f = 1 \\ P_{right-skewed} & \text{if } k_f \leq 1 \text{ and } k_f \geq \frac{1}{N_{group}} \\ P_{scale-free}, & \text{if } k_f = \frac{1}{N_{group}} \end{cases} \tag{3.7}$$

When k_f is 1 and thus a randomly generated network is implemented, the probability of a node being chosen to make a connection is dependent on the amount of nodes its group is represented by. For this reason the probability of a node being chosen is described in 3.8.

$$p_{group} = \frac{1}{N_{group}} \tag{3.8}$$

The amount of connections a node eventually has depends on the amount of edges it's group has with another group as well as the amount of agents that are represented in it's own group. Therefor the average degree of a group is

$$< k >_{group} = p_{group} * \frac{1}{n} * \sum_{i}^{n} c_{group_i} \tag{3.9}$$

Where p is the probability of being chosen as described in equation 3.8 and c_{group} is the amount of connections a group has with another group i and n is the total amount of different groups the group has connections with. The degree distribution is easily calculated for each group because the $< k >$ is a fixed value. Therefor the degree distribution of a group can be described as

$$P_{random} = \left(\frac{N-1}{< k >} p^{<k>} (1-p)^{N-1-<k>} \right) \tag{3.10}$$

When kf is $\frac{1}{N_{group}}$ and thus generate a scale-free network, the probability of a node being chosen to make a connection is the same as probability as p(k)(3.5) because the probability of being connected now depends on it's degree. Also the degree distribution described in equation 3.6 is the same for certain node - connection pairs. Because the amount of nodes and connections are pre-determined there is a closed system in which the connections of the nodes can not rise till infinity. Also a node can not connect twice with the same node so when the amount of connections >> the amount of nodes all

Algorithm 2 Initializing of random social network between groups

```
for row in Connections data frame do
    S ← Source group nodes
    D ← Destination group nodes
    C ← Connections between groups
    D-bin ← random.sample(D, k_fraction * 100)
    while i < C do
        N_S = random.sample(S)
        N_D = random.sample(D-bin)
        if N_S ≠ N_D And N_D notin links(N_S) then
            Add Directional edge from N_S to N_D
            if k_fraction < random.uniform() then:
                Add N_D to D-bin
                N_D_Random = random.sample(D)
                Add N_D_Random to D-bin
            end if
        end if
        i += 1
    end while
end for
```

nodes will have the same amount of connections. We expect to see a power-law like distribution when the expected amount of edges of the most connected node does not surpass the amount of nodes in the system. The expected amount of nodes can be easily calculated as follows:

$$E[edges] = \sum_{n_{edges}} \frac{k_{i_{max}}}{\sum_j k_j} \tag{3.11}$$

The highest expected value the highest connected node can have for being chosen is equal to 0.5, because each time this nodes gets chosen and thus added to the bin, another randomly chosen node gets added to the bin as well. This means that whenever the amount of connections is two times as large as the amount of nodes we well definitely see a power law.

Because social networks are rarely scale free but tend to be left skewed we will have the best chance of getting a real world social network is when k_f will be between $\frac{1}{N_{group}}$ and 1.

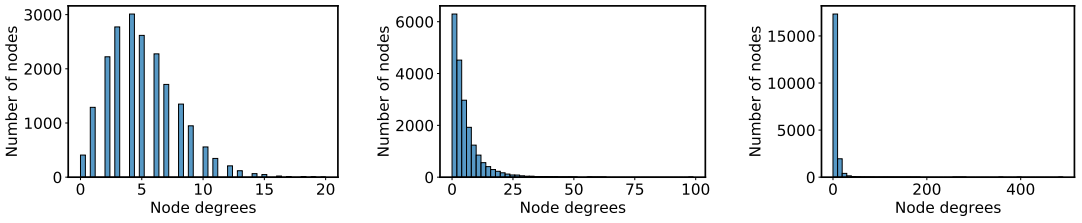


FIGURE 3.1: This figure shows the in-degree distributions of a dummy network which contains 20000 nodes and 100000 edges. The Figure shows three degree distributions when $k_f = 1$(left), $k_f = 0.1$(middle) and $k_f = \frac{1}{N_{group}} = \frac{1}{20000}$ (right)

3.5 Layer specification

Generating the networks based on Algorithm 3 might give a good general representation of the social network of Amsterdam, the biggest problem of this implementation is that the different layers are being handled as if they are the same. The same approach has been taken for all four layers while one of the mayor values and things we should do is to handle each layer of the multilayer network in a different way. Another problem that occurs when only implementing the networks based on these algorithms is that there is a lack of multigroups where a group is not connected to another group. This may be accomplished by include geographical data, household data, and a reciprocity parameter. In this section the implementations of each layer is described.

To implement this a simple reciprocity parameter, r, is introduced. This parameter ranges between 1 and 0, with 1 indicating full reciprocity and 0 indicating no reciprocity. This parameter is easily implemented by simply including a uniform probability. If the parameter r is greater than this probability, the edge is undirected; if it is less than this probability, the edge is directed.

3.5.1 Household layer

As described in Chapter 2 the household data is fully symmetrical, For this reason, a r value of one is assigned. One of the aspects of a household network that distinguishes it from other networks is that the agents in one household have no relationship with the agents in another, i.e. when an agent belongs to a household, the same agent cannot possibly belong to another household and thus cannot have any connections with the agents in this household. This results in a multigroup model in which your network is entirely dependent on your group membership[17].

TABLE 3.1: Attributes of household table.

Symbol	Description	Type	Values
no hhold	household ID	int	[6-89986]
positie	Agent's position in household	string	child, head, spouse
no mem	Household Member ID	int	[1-10]
gebjaar	Birth year	int	[1925-2019]
oplmet	Education level	String	e.g. HBO, University, VMBO
aantalhh	amount of agents in a household	int	[1-10]
aantalki	amount of children a household	int	[1-8]

Another critical aspect of the household network is that the households themselves be representative. To meet this requirement, the DNB Household Survey dataset[1] from Centerdata was used. In the DHS Panel members annually provide detailed information for the DHS Household Survey (DHS), providing researchers with a rich set of background information on many aspects of the respondents' lives such as Individual characteristics and household structure. Despite the fact that the panel is based on a random sample, unit nonresponse and selective attrition suggest that some groups are better represented than others. Participation, in particular, is associated with higher socioeconomic status. CentERdata provides education, income, gender, and age weights that can be used to correct for this in the analysis. Teppa and Vis[19] describe the data set and their methodology in more detail. Table 3.1 shows the household attributes which are the most interesting for the houshehold aggregations. These attributes can be rewritten by the agent attributes described in Chapter 2 to make it competible with the data.

To get the right household distribution of Amsterdam statline dataset of the CBS is used[2]. The distribution of households are represented in table 3.2.

Based on the household aggregations and distributions we can introduce a new way to make the household network. When this household network is created we discard the old household network. The way the new household network is generating the amount of households described in table 3.2, an sampling from the DHS a household which corresponds with the size of the household. Than we can take agents from our agent dataset that fit in the description of the household. We can do this untill all the agents have a household and a multigroup model is realized.

[1] https://www.dhsdata.nl
[2] https://opendata.cbs.nl/statline/CBS/nl/dataset/71486NED/tablefromstatweb

TABLE 3.2: Distribution of the amount of households per household size

Household size	Amount
1	248069
2	124787
3	43208
4	32883
5 or more	17401

3.5.2 Family layer

The network in the family layer is fully symmetrical just like the household layer, this means that for this layer the r parameter is set to 1. People in the family layer, unlike those in the household layer, can belong to up to two families: your own and the family of a potential spouse. As a result, this network is more entangled than a multigroup network. One attribute that we do have to taken in to consideration when making this layer more complete is the fact that a lot of earlier mentioned household connections are also family and thus a lot of nodes need a connection on both layers.

To do this the amount of sampled families can be retrieved from the household layer and be re-used in the family network. This way the families are as well represented in the family layer as in the household layer.

3.5.3 Neighbourhood layer

The neighbourhood layer is not fully symmetrical which means that the r parameter is not equal to 1. Tho this network is not fully symetrical it is quite stright forward that there is still a high reciprocity in neighbourhoods. Therefore we use a r parameter of 0.5 or higher for this layer.

Just like the household layer, the neighbourhood layer can be seen as a multigroup model. Where each group is represented by its neighbourhood instead of it's household. These neighbourhoods can be observed in figure 3.2. When approaching the network in this way we can make sure that people of spatially different neighbourhoods don't have a connection on neighbourhood level with agents that live in the east of Amsterdam.

To make sure each agent only can connect with agents from its neighbourhood we need to introduce an addition to the initial groups discussed in section 3.4. Instead of group connection based on characteristics we also need to make a restriction that two nodes can only be connected when they live in the same neighbourhood. To do this we need

to distribute all the nodes in a neighbourhood and thus give an area code. This has to be done accordingly to the amount of agents in each area and the distribution of agent types in a area. To make such a distribution the BBGA dataset[3] is used. This dataset gives a rich representation on how different kind of persons are distributed among the neighbourhoods. Three different scales are used to describe the seperate areas of Amsterdam in different sizes, these scales are: 22-areas, districts and neighbourhoods, see figure 3.2. From this dataset we can see how people with a certain education level, ethnic background and gender are distributed and can be combined to make a joint probability distribution for each area. based on this distribution we can assign for each area the right amount of persons.

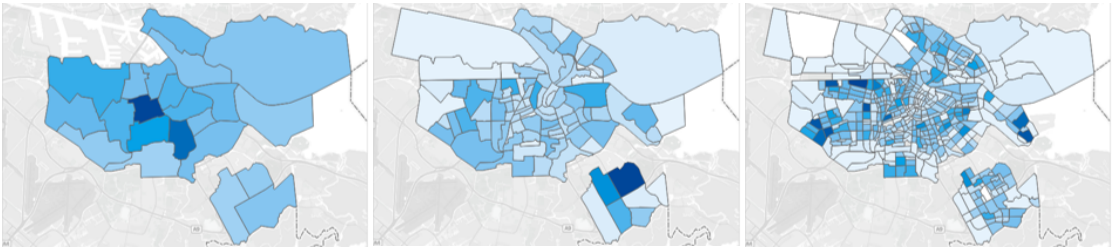


FIGURE 3.2: Amsterdam divided into 22 area's (left), districts (middle) and neighbourhoods (right).

Algorithm 3 Initializing of random social network between groups

```
for group in groups do
    nodes = nodes in group
    shuffle(nodes)
    shuffle(areas)
    i = 0
    while i < amount of nodes do
        node = i^th node
        if area is not full then
            add node to area
        else
            go to next area
        end if
    end while
end for
```

This also results in a problem, which is the fact that agents of neighbouring neighbourhoods can not have a connection while they might live next to each other. To resolve this issue a stochastic element is implemented which makes sure that a certain percentage of the population also has a connection with a neighbouring neighbourhood.

[3]https://onderzoek.amsterdam.nl/interactief/dashboard-kerncijfers

To make sure that each agent inside the same household also lives in the same neighbourhood the same layer lives inside the same area as it's made only possible for a agent in a household to make a household connection to agents who live in the same area

3.5.4 Work/school layer

The last layer is the work/school layer. To get a more specific view for how the persons are clustered in this layer we divided the agents in agents that are working and agents that go to school. to be able to see this how many agents between the ages of 0-20 are attending primary school, secondary school and vocational school.

1. Take all agents from 0-20
2. From these agents only take the agents that have a educational level 1 or 2
3. Use the stats from the BBGA to see how many of these agents are registered at primary education level and thus have a education level 1.
4. Look at CBS stats how many percent of the highschool students are not registered at the 4^{th}, 5^{th} or 6^{th} grade of HAVO/VWO and thus also have a educational level of 1.
5. Look at CBS stats how many percent of the high school students are registered at the 4^{th}, 5^{th} or 6^{th} grade of HAVO/VWO and how many students are regisered at MBO-2, MBO-3 and MBO-4, and thus have a educational level of 2.
6. Take label these agents as students and form "schools" which only they can attempt to.

Based on these criteria the students are chosen. These students are grouped together in this layer so that they can not interact with people who are working. There is a small fraction of working agents that work on such schools, such as teachers. Therefor we let a small percentage of the working agents also attempt these schools.

The forming of the schools are made by using a big assumption which is that the schools ethnic composition corresponds to the ethnic composition of the neighbourhood it is in. Although this is somewhat true for primary schools [2, 3], it is not necessarily the case for secondary schools and vocational education (MBO). Therefore this assumption needs to be addressed and taken into account. Based on this assumption we us the same data as in section 3.5.3 to make a composition of the schools and classify agents to a school accordingly.

For the working agents we make so called "working places", which are made in a random manner. The quantity of companies and workshops is determined on the basis of a distribution, which can be either a normal distribution or an exponential distribution. People are randomly assigned to a company and can then only have a connection with someone from the same company.

3.6 Results

3.6.1 Statistics

- Distributions - Modularity - Transitivity - Reciprocity - Small-World - Assertive

3.7 Discussion

Chapter 4

Homophily within the social network of Amserdam

4.1 Introduction

4.2 Background

4.3 Methods

4.3.1 Retrieve the connection probabilities from data

4.3.2 SDA model

4.3.3 New model

$$p_{ij} = (a + \beta * D_{ij}) * e^{-\alpha * D_{ij}} \tag{4.1}$$

$$D_{ij} = -\frac{W\left(-\frac{\alpha p_{ij}}{e^{\frac{\alpha a}{\beta}}\beta}\right) b + \alpha a}{\alpha\beta} \tag{4.2}$$

TABLE 4.1: R^2 values for homophily fit on data. - means that V has a bigger R2 and + means 1 has a bigger R squared

Layer	0(base)	D V 1 *	D 0 V *
Household	0.99999989	-5.16e-09	0.054
Family	0.9999965	2.88e-06	0.056
Neighbours	0.999965	-2.79e-06	0.180
Work/school	0.99999969	-1.5938e-06	0.72

4.4 Results

4.4.1 Total probability

- Show the results we have acquired based on the probabilities

4.4.2 Most important characteristics

4.5 Discussion

Chapter 5

Dynamic network of Amsterdam

5.1 Introduction

5.2 Background

5.3 Methods

5.3.1 ABM

5.4 Results

5.5 Discussion

Chapter 6

Static network

Chapter 7

Discussion

Chapter 8

Conclusion and future work

Bibliography

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[2] W. R. Boterman. The role of geography in school segregation in the free parental choice context of dutch cities. *Urban Studies*, 56(15):3074–3094, 2019.

[3] S. v. F. Boterman Willem R, Cohen Lotje. Monitor diversiteit in het basisonderwijs. *Onderzoek, Informatie en Statistiek*, 56(15):3074–3094, 2018.

[4] A. D. Broido and A. Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

[5] J. Bruggeman. *Social networks: An introduction*. Routledge, 2013.

[6] D. Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[7] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.

[8] E. Cozzo, R. A. Banos, S. Meloni, and Y. Moreno. Contact-based social contagion in multiplex networks. *Physical Review E*, 88(5):050801, 2013.

[9] L. Crielaard, P. Dutta, R. Quax, M. Nicolaou, N. Merabet, K. Stronks, and P. M. Sloot. Social norms and obesity prevalence: From cohort to system dynamics models. *Obesity Reviews*, 21(9):e13044, 2020.

[10] J. A. Davis. Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review*, pages 843–851, 1970.

[11] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[12] T. Johansson. Generating artificial social networks. *The Quantitative Methods for Psychology*, 15(2):56–74, 2019.

[13] C. Kadushin. *Understanding social networks: Theories, concepts, and findings*. Oup Usa, 2012.

[14] M. Magnani and L. Rossi. The ml-model for multi-layer social networks. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, pages 5–12. IEEE, 2011.

[15] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[16] M. POSFAI and A.-L. BARABASI. *Network science*. Cambridge University Press, 2016.

[17] S. Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316 (5829):1298–1301, 2007.

[18] S. Talaga and A. Nowak. Homophily as a process generating social networks: insights from social distance attachment model. *arXiv preprint arXiv:1907.07055*, 2019.

[19] F. Teppa, C. Vis, et al. The centerpanel and the dnb household survey: Methodological aspects. Technical report, Netherlands Central Bank, Research Department, 2012.

[20] S. Wasserman, K. Faust, et al. Social network analysis: Methods and applications. 1994.