University of Amsterdam

Masters Thesis

Constructing a digital twin of the social network of Amsterdam

A research about the creation of the social network of Amsterdam and it's underlying dynamics

Author:
Kamiel Gülpen

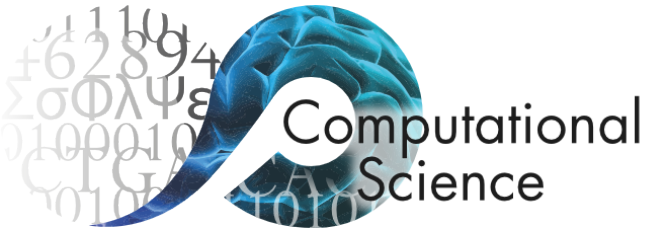
Supervisor:
Debraj Roy
Vítor Valsconcelos
Dhruv Mittal

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

June 2022


Computational
Science

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**rdős **R**ényi (random network)
BA	**B**arabasi **A**lbert
CBS	**C**entral **B**ureau for Statistics

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 [20]. 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[23], reciprocity[21] and transitivity[16].

Mining and analyzing social networks in diverse situations yields crucial insights for improving fundamental knowledge and understanding of human behavior [38]. Researchers have been able to explore social theories and effects such as the genesis of homophily[23], reciprocity[21] and transitivity[16] 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.

1.1 Research questions

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

Does the idea of homophily explain why citizens of Amsterdam connect with eachother, and what are the most important characteristics?

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)[15].

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

Overweight and obesity are major health issues all over the world [40]. Being overweight is linked to a variety of diseases and disorders, including diabetes and cancer, and has ramifications for mental health as well as social and economic consequences [29]. In 2019, 50.1 percent of Dutch adults aged 18 and up were overweight or obese [29]. This problem is more severe in Amsterdam than elsewhere in the country, particularly among children and young people [35].

A high obesity prevalence is one of the causes of a increasing percentage overweight and obese citizens. This prevalence and social acceptability leads to a continuous growth in children's BMI [37]. The emergence of such social norms and other social behaviors spreads by social contact[12, 13]. As a result, the network structure of who is connected to whom can have a significant impact on the rate at which a behavior spreads across a population [5, 13, 30]. Social networks are the pathways along which these social behaviours propagate. Studies of diffusion dynamics have demonstrated that the structure of a social network can have important consequences for the patterns of collective behavior that will emerge [3, 39]. Accurate descriptions of social interactions are thus critical for shedding light on the most significant mechanisms at work in these processes.

This paper proposes a model which can create a family of interaction networks for the city of Amsterdam based on data of the Central Bureau of Satistics (CBS). The goal of the construction of these networks is to accurately describe the social interactions

of the citizens of Amsterdam via a multiplex network of four layers: households, family, neighbors, and work/school ties. First the most common construction methods of a network are discussed, followed by an elaborate analysis of the properties of social networks. Secondly, methods for applying the ideas of network construction algorithms to the problem at hand are discussed. Afterwards additional data sources are used to create a more detailed and representative network of Amsterdam. Finally, the models outcome are thoroughly examined using a sensitivity analysis. This model and its output, can be used for future research to investigate the spread of obesity in the city of Amsterdam together with a wide range of other social contagious phenomena.

3.2 Background

3.2.1 Network engineering methods

Network science aims to build models that reproduce the properties of real networks. The three most popular network construction algorithms for making artificially real world networks are random networks, small-world networks and scale-free networks [3, 19, 39].

3.2.1.1 Random networks

The fundamental question pertaining to our understanding of how networks are formed was first asked by the mathematicians Ál Erdős and Alfréd Rényi [1]. In their famous paper: *On the evolution of random graphs*[19] they propose a random network generating algorithm (ER) that is able to create a random network based on the fact that, at first glance, networks do appear to be random. In a network, created by the ER algorithm, each node connects to another other in a random manner. A random network consists of N nodes where each node pair is connected with probability p. To construct a random network following steps are followed:

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.

3.2.1.2 Small-world networks

The small-world construction algorithm (SW) was first proposed by Duncan J. Watts and Steven Strogatz [39]. Their model is an extension on random networks, so that it creates a small-world property and a high cluster coefficient. The small-world network algorithm works as follows:

1. Start with a ring of N vertices each connected to its k nearest neighbours by undirected edges.
2. Choose a vertex and the edge that connects it to its nearest neighbour in a clockwise sense.
3. Reconnect this edge to a vertex chosen uniformly over the entire ring at random with probability p, with duplicate edges forbidden.
4. Repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed.
5. Consider the edges that connect vertices to their next-nearest neighbours clockwise.
6. Repeat step 1 to 5 untill each edge in the original lattice has been considered once.
7. As there are $\frac{nk}{2}$ edges in the entire graph, the rewiring process stops after $\frac{k}{2}$ laps.

3.2.1.3 Scale-free networks

The scale-free network construction algorithm, also known as the Barabási-Albert (BA) algorithm, is firstly proposed by Albert-László Barabási and Réka Albert [3]. Their algorithm is based on the idea that most of the networks, observed in the real, world are scale-free. This indicates that the node degree follows a power-law distribution, where most of the nodes have a small amount of edges while a few of the nodes having many. The scale-free network is based on two characteristics, which are Growth: networks are the result of a growth process that continuously increases the amount of nodes and Preferential Attachment: networks new nodes tend to link to the more connected nodes. The algorithm works as follows:

1. Begin with m_0 nodes, with links between them picked at random, as long as each node has at least one link.
2. Growth: 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.

3. Preferential Attachment: New nodes tend to connect to more linked nodes.

The probability $p(k)$ that a link of the new node connects to node i depends on the degree k_i as

$$p(k_i) = \frac{k_i}{\sum_j k_j} \tag{3.1}$$

3.2.2 Social networks

When creating an artificial social network, it is critical to consider the properties of a social network in general. The method used to generate a network, and thus the underlying algorithm algorithm used, has a significant impact on these network properties [11]. In this the section social network metrics are defined, which can later be used to utilize and describe the characteristics of social networks in depth.

3.2.2.1 Density

The maximum possible number of edges in a simple graph is $n(n-1)$ for a directed graph and $\frac{1}{2}n(n-1)$ for a undirected graph. The density ρ of a graph is the fraction of these edges that are actually present:

$$\rho = \frac{l}{n(n-1)} \tag{3.2}$$

where l is the number of edges in the network. The density of a network lies in the range between 1 and 0. Most of the networks of interest are big enough such that that 3.2 may be reasonably approximated as $\frac{l}{n^2}$. A dense network is one in which the density tends to a constant as n $\rightarrow \infty$. As the network grows in size, the proportion of non-zero entries in the adjacency matrix remains constant. A sparse network is one in which ρ $\rightarrow$ 0 as n $\rightarrow$ 0. In other words, a network is sparse as l approaches a constant value as the network n grows towards a large value.

Sparsity, as discussed above, is a common phenomena in real-world networks, particularly in social networks with a large number of nodes. This is due in part to the fact that people lack the cognitive abilities and time required to interact with everyone in such a network. The so-called Dunbar number gives a estimation of the cognitive limit to the number of people a person can have and maintain stable social relationships[18]. As a result, an artificially designed social network should take this property into consideration.

3.2.2.2 Degree distribution

When we design a random network, as described in section 3.2.1.1, it shows that some nodes get numerous edges while others have few or no links. The degree distribution p_k, which is the chance that a randomly picked node has degree k, captures these differences.

The probability, p, that node i has precisely k linkages in a random network is described by Barabasi [2] as the sum of three components:

1. The probability that k of its linkages exist, denoted by pk.
2. The likelihood that the remaining $(N - 1 - k)$ linkages are broken, or $(1 - p)$ $N - 1 - k$.
3. The number of ways we may choose k links from the N - 1 possible linkages for a node, or $\left(\frac{N-1}{k}\right)$

Consequently, the degree distribution of a random network follows the binomial distribution, which is well approximated by a Poisson distribution in the limit of $k << N$.

$$p_k = \left(\frac{N-1}{k} p^k (1-p)^{N-1-k}\right) \tag{3.3}$$

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{3.4}$$

In a random society, all individuals are supposed to have a same number of friends. Therefore, if persons are randomly related to one another, we lack outliers: There are no famous people, and no one is left behind with only a few friends. This surprising finding is due to a fundamental aspect of random networks: in a large random network, the degree of the majority of nodes is in the limited range of $\langle k \rangle$. The BA method does generate these types of ourliers, since the algorithm is based on preferential attachment, there are both popular nodes that continue to grow in popularity and unpopular nodes. As a result, the degree distribution is scale-free.

$$p_k = k^{-\gamma} \tag{3.5}$$

Since Albert-László Barabási and Réka Albert published their algorithm and claimed that most networks are scale-free, there has been a lot of debate on this topic. The most vehement criticism came from a paper titled "scale-free networks are rare"[10]. This

paper contends that in real life, most networks, particularly social networks, are not scale-free. Despite the fact that this paper has received a lot of resistance [31, 36] most researchers do agree that most social networks are in fact not scale free [10, 31, 36]. It does not, however, imply that degree distributions in social networks are always Poisson-like and symmetric. On the contrary, they can come in a variety of shapes, particularly those that are significantly right-skewed [10, 25].

3.2.3 Small-world

The small world phenomenon, also known as six degrees of separation, has long fascinated the general public. It states that if you choose any two individuals anywhere on Earth, you will find a path of at most six acquaintances between them. This concept was initially proposed by the American psychologist Millgram, who became famous for his experiments demonstrating that two people on the planet are separated by an average of five connections [24, 34]. The small-world property got it's mathematical explanation when Ithiel de Sola Pool and Manfred Kochen wrote their paper "contacts and influence" [2, 17], in which they showed that for a network with a small-world property the distance (d) grows proportionally to the logarithm of the number of nodes N in the network:

$$d \propto logN \tag{3.6}$$

The small-world concept has received widespread support over the years and has been demonstrated in a variety of networks. It has been shown that the same property can be replicated when a random network is generated. For a random network with average degree $\langle k \rangle$, the expected number of nodes N(d) up to a distance d is given by

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + ... + \langle k \rangle^d \approx \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1} \tag{3.7}$$

We can identify the maximum distance, d_{max}, or the network's diameter by setting $N(d_{max})$ to N because $N(d)$ must not exceed the total number of nodes, N, in the network we can thus state that, assuming $\langle k \rangle >> 1$, that :

$$N(d_{max}) \approx N \tag{3.8}$$

$$\langle k \rangle^{d_{max}} \approx N \tag{3.9}$$

$$d_{max} \approx \frac{lnN}{ln\langle k \rangle} \tag{3.10}$$

The small-world property is also present for certain configurations of the barabasi albert algorithm. Cohen and Havlin showed [14] that, using analytical arguments, the diameter of a network scale-free is dependant on the exponent of the network. These exponents and their corresponding diameter is shown in 3.11.

$$\langle d \rangle \backsim \begin{cases} constant, & \text{if } \gamma = 2 \\ lnlnN & \text{if } 2 < \gamma < 3 \\ \frac{lnN}{lnlnN}, & \text{if } \gamma = 3 \\ lnN, & \text{if } \gamma > 3 \end{cases} \tag{3.11}$$

3.2.3.1 Clustering, Transitivity and Communities

The clustering coefficient is a key statistic for evaluating a network's local organization [39]. Social networks tend to have a high cluster coefficient while still being sparse as described in 3.2.2.1. The local clustering coefficient C_i, represents the density of triangles flowing through node i or, more precisely, the likelihood that two neighbors of node i are related [6]. It is defined as follows:

$$C_i = \begin{cases} \frac{\text{nr of triangles passing though node } i}{k_i(k_i-1)/2} & \text{for} \quad k_i > 1, \\ 0 & \text{for} \quad k_i = 0, 1 \end{cases} \tag{3.12}$$

The global cluster coefficient can be obtained by taking the average of all the local cluster coefficients:

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{3.13}$$

An alternative global clustering coefficient known as 'transitivity' is occasionally presented. Transitivity is a property that considers triples of nodes in a graph or, in other words, measures the density of triangles (three nodes connected to each other by three edges in the network, implying that every node is fully connected to the remaining two nodes) in a network [38]:

$$T = 3\frac{\text{Triangles}}{\text{Triads}} \tag{3.14}$$

One of the distinguishing characteristics of social networks is, next to a high transitivity, their tendency to be divided into modules (also called groups, clusters or communities). One way to find the strength of a networks division is by looking at the its measure which is called the modularity of a network. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. The modularity of a network partition is assessed using the modularity maximization method by comparing its number of links to a null network model, which has an equal number of nodes, links, and degree distribution but with random links among the nodes. A modularity function that can measure the quality of the introduced partitions Q as a community:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta\left(c_i, c_j\right) \tag{3.15}$$

The expected fraction of links is measured with the assumption that the probability that nodes i and j are connected is $\frac{k_i k_j}{2m}$. Here, k_i andk_j are respectively the output and input strengths of nodes i and j, $2m$ is total strengths of the network, and δ is Kronecker's delta where c_i c_j equals 1 if nodes i and j are in the same community and equals 0 otherwise. A_{ij} is the adjecency matrix of the concerned network.

Optimising modularity is NP-hard, and consequentially many heuristic algorithms have been proposed. One of the most popular algorithms to optimise modularity is the so-called Louvain algorithm[7] and leiden algorithm[33].

3.2.3.2 Degree Correlation

It is observed that nodes are not connected randomly across a wide range of real-world networks, but that there is a considerable tendency for nodes of high degree to connect to either nodes of high degree (assortative networks) or nodes of low degree (disassortative networks)[6]. Positive assortativity is commonly observed across a wide range of social networks[25]. The Pearson correlation coefficient of degrees between pairs of linked nodes is used to calculate the degree coefficient of a network [6]. We can define a matrix e as the joint probability matrix of the degrees, such that e_{ij} is the fraction of all edges that join a node with degree i with a node j. Since e is the joint probability matrix for the whole network, it holds that

$$\sum_{ij} e_{ij} = 1 \tag{3.16}$$

Furthermore, we define the quantities a_i and b_j as

$$\sum_{i} e_{ij} = a_i \tag{3.17}$$

$$\sum_{j} e_{ij} = b_j \tag{3.18}$$

such that a_i and b_j are the fractions of edges that start and end at nodes with degree i and j respectively. Using this, the degree correlation coefficient is defined as

$$r = \sum_{k,k'} \frac{kk' \left(e_{k,k} - q_k q_{k'}\right)}{\sigma^2} \tag{3.19}$$

where σ is the standard deviation in the distribution of e and r is a value between -1 and 1, where a positive value means that the network can be seen as assortative and a negative value indicates a disassortative network.

3.2.3.3 Reciprocity

The tendency toward reciprocity is considered to be a quite basic effect and a feature of most social networks [38]. Reciprocity, r, is a metric for directed networks that assesses the proclivity of two nodes to create mutual connections with one another. This measure can be calculated by computing the ratio of mutual connections in the network to total connections:

$$r = \frac{L_{bi}}{L} \tag{3.20}$$

3.2.3.4 Multiplex networks

Multiplex networks are those in which the same set of nodes or actors are linked to one another via various types of links. In the case of Amsterdam, a social network can for example be represented by a multiplex in which the same set of people might interact via their Household (first layer), via their family (second layer), via their work/school

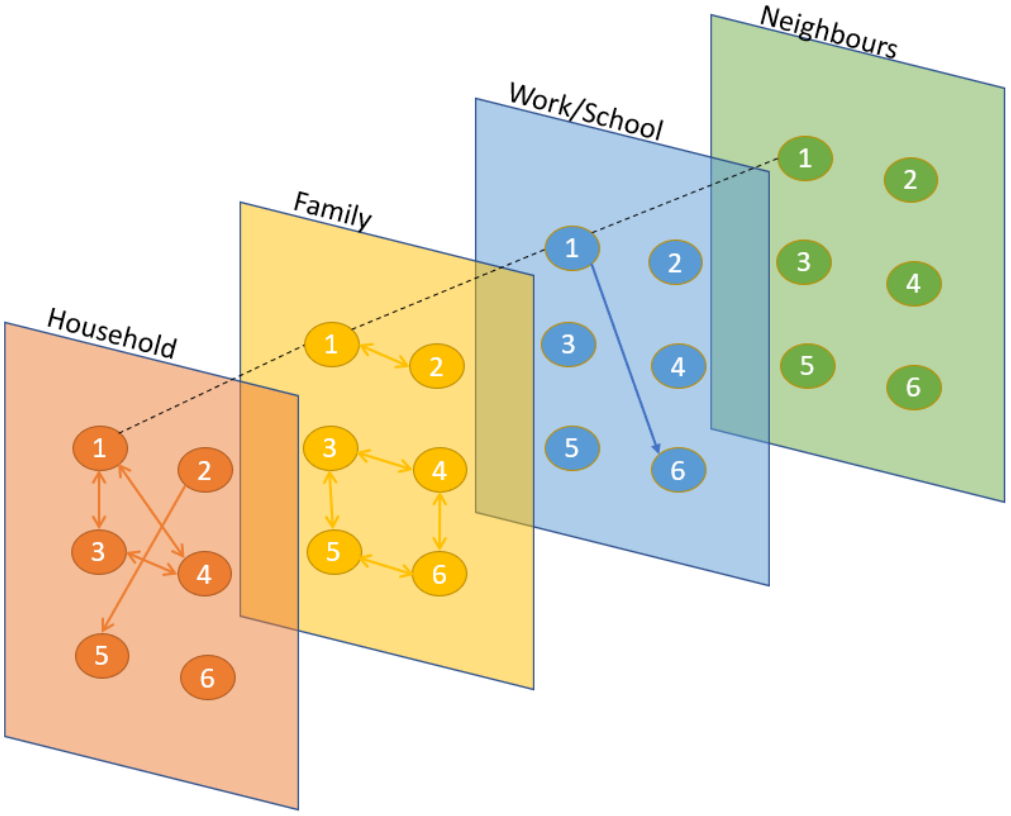


FIGURE 3.1: Visual representation and example of a multiplex network

relationships (third layer) or their neighbourhood (fourth layer), see Figure 3.1. Multiplex networks are widely used for describing social networks and has it's origin in the social networks [6].It has even been claimed that reducing a social system to a monolayer network is an extremely crude approximation of reality [22].

A relevant feature of many multiplex networks is their link overlap[6]. This means that a significant number of pairs of nodes are connected in multiple layers. Multiplex networks with significant link overlap are ubiquitous and include multiplex airport networks, online social games, collaboration and citation networks [6].

The total overlap O$[\alpha, \beta]$ between two layers α and β is defined as the total number of links that are in common between layer α and layer β and can be defined as

$$O^{[\alpha,\beta]} = \sum_{i<j} a_{ij}^{[\alpha]} a_{ij}^{[\beta]} \tag{3.21}$$

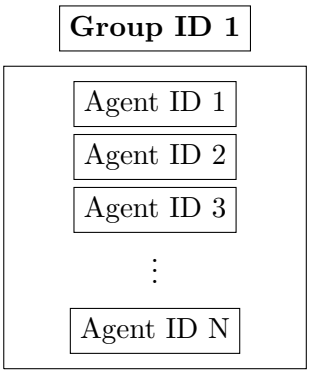


FIGURE 3.2: Visual representation of a group with N agents.

3.3 Methods

3.3.1 Generating a network

In this section, the methods used to create a family of artificial networks that represent Amsterdam's social network are discussed. First, we will go through the agents' initialization. Second a new method is introduced that includes an algorithm which is capable of constructing a network based on the fact that agents are classified into groups (based on characteristics) and that these groups have a predetermined number of connections with one another. This network is heavily inspired by the ER and BA algorithms and captures both ideas in one algorithm.

3.3.1.1 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. In total we got 240 groups where each group has its own group ID. To build a network with all of the agents represented as nodes, we must go through each group and through the nodes in the corresponding group. The pseudocode of the algorithm that creates each node can be viewed in Algorithm 1. The main objective of this algorithm is to to assign a N amount of agents to a group, where N is the amount of persons of that group that live in Amsterdam. A visual representation of the output can be observed in figure 3.2

The initialization of the nodes begins with setting an agent ID to 0, this ID is given to the first agent. 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. The number of nodes are used to determine the amount of times we loop through N amount of agents that are in the particular group. These agents are added to a dictionary with the group ID as its key,

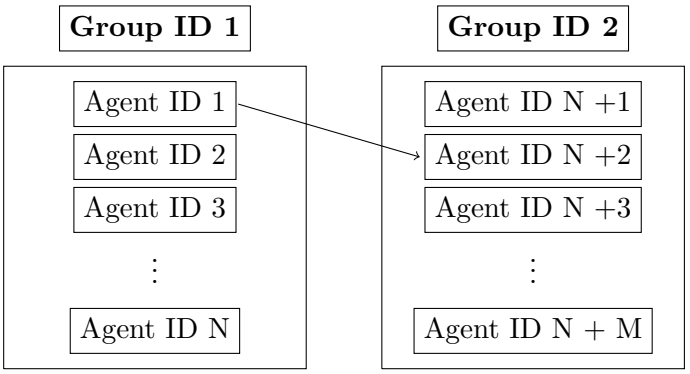


FIGURE 3.3: Visual representation of how a possible connection between two agents of two groups can be realized. Group ID 1 is the source group and Group ID 2 the destination group

so that we can keep track of which nodes are assigned to which group. The agent ID is updated after a agent is added to a group. After an group dictionary has been filled with N agents, the algorithm moves on to the next group, and so on until all groups have the appropriate number of agents.

Algorithm 1 Initializing of agents

$ID_{agent} = 0$
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
  group = $C_{ethnicity}$, $C_{agegroup}$, $C_{education}$, C_{gender}
  for agent in range(N) **do**
    add ID_{agent} to group
    update ID_{agent}
  end for
end for

3.3.1.2 Network generators

After the agents have been initialised, we can start making a network based on the data described in Chapter 2 of each layer. The network is a directed network meaning that the edges between agents point in only one direction, the data states how many edges there are between 2 groups of agents and a layer is complete when all the predetermined connections between all the combinations of groups are made. Figure 3.3 shows a visual representation of an edge between two agents. The strategy of choice determines how each node is selected to create an edge.

The random strategy is the most straight forward strategy for selecting agents. The two group random network (2GRN) algorithm is implemented to successfully generate a random network. A random network is initialized by choosing the source agents from the source group and agents from the destination group in a random manner. This algorithm examines the number of edges that exist between each group combination, i.e. 200 edges between group 1 and 2, 600 edges between group 5 and 7 etc., and selects an agent at random from the source and destination groups for each edge. A connection between the two agents is formed if there is not already an edge between the two agents and if the agents are not the same.

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 3.22.

$$p_{group} = \frac{1}{N_{group}} \tag{3.22}$$

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

$$\langle k \rangle_{group} = p_{group} * \frac{1}{n} * \sum_{j}^{n} c_{ij} \tag{3.23}$$

Where p is the probability of being chosen as described in equation 3.22, c_{ij} is the amount of connections between group i and j and n the amount of groups.

The in-degree distribution between two groups is approximately the same as equation 3.3 in section 3.2.2.2 and thus

$$P_{random} = \left(\frac{N-1}{\langle k \rangle} p^{\langle k \rangle} (1-p)^{N-1-\langle k \rangle} \right) \tag{3.24}$$

A second strategy of picking the nodes is based on preferential attachment. To make a network based on preferential attachment the two group scale-free network (2GSFN) algorithm is used, fow which the probability that a destination node is chosen to form an edge depends on it's degree . A destination bin (DB) must be created in order to add a preferential attachment property to a two group network constructor. The addition of the DB leaves us with three groups: the source and destination group, which include all the agents of that group and the DB that starts with a single randomly chosen agent from the destination group. To make an edge between two agents we first choose an

agent randomly from the source group and then, also at random, from the DB. The agents connect if there is not already an edge between the two agents and if the agents are not the same. Afterwards the chosen destination node is added to the bin together with a randomly sampled node from the destination group. This makes the probability for a destination node being chosen:

$$p_{group} = \frac{\sum_i n_i}{\sum_j n_j} \tag{3.25}$$

Where n_i are the number of times node i is represented in the DB and n_j the total amount of nodes inside the DB. The corresponding in-degree distribution between the two groups follows the distribution described in equation 3.26.

$$P_{scale-free} = k^{-\gamma} \tag{3.26}$$

Because the amount of nodes and connections are pre-determined there is a closed system and not a indefinite growing network. It does, therefore, not follow a perfect scale-free property. A drawback from such a closed system is that all nodes will have the same degree when, the amount of connections >> the amount of nodes, due to the fact that two agents can only connect once. We do 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:

$$E[edges] = \sum_{n_{edges}} \frac{k_{i_{max}}}{\sum_j k_j} \tag{3.27}$$

The highest expected value of edges for the highest connected node is equal to $\frac{n_{edges}}{2}$, because each time this node is chosen and thus added to the bin, another randomly selected node is also added to the bin. This means that whenever the number of connections is twice as large as the number of nodes, we will almost certainly see a power law.

Social networks are rarely scale free or random, but tend to be left skewed [10]. Therefore, the ideal algorithm would be a combination of both the 2GRN and the 2GSFN. This algorithm, simply called the two group network (2GN) algorithm, is described in psuedocode in Algorithm 2. The 2GN algorithm works a lot like the 2GSFN algorithm but differs from it on two main points, which are the fraction of agents that are initially in the DB and the chance that a chosen agent is added to the DB. The fraction and

chance are combined to one parameter called k_f, where the fraction of initial agents in the DB equals k_f and the the chance of being added to the DB corresponds to 1 - k_f. So when k_f is equal to 1, the DB is initially filled with all the agents of the destination group and is thus equal to the destination group and that the chosen node has a 1-k_f percent chance of getting added tot the DB, which makes the DB unchanged. This means that when k_f has the value of 1 it is the same algorithm as the 2GRN algorithm and thus generates a P_{random}. Whenever k_f is equal to $\frac{1}{N_group}$ there will only be one agent in the destination bin and the chance of being added to the destination bin is 1-$\frac{1}{N_{group}}$, which is, in most cases, approximately 1. This means that for a k_f value of $\frac{1}{N_{group}}$, the algorithm works like the 2GSFN algorithm. For all the values of k_f between 1 and $\frac{1}{N_{group}}$ the algorithm generates left-skewed in-degree distribution (see 3.28). The in-degree distributions of the networks created by the 2GN algorithm for $k_f = 1$,$k_f = 0.1$ and $k_f = \frac{1}{N_{group}}$ are shown in figure 3.4.

$$P(k) = \begin{cases} P_{random}, & \text{if } k_f = 1 \\ P_{right-skewed} & \text{if } \frac{1}{N_{group}} \leq k_f \leq 1 \\ P_{scale-free}, & \text{if } k_f = \frac{1}{N_{group}} \end{cases} \tag{3.28}$$

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_f * 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_f < random.uniform() then:
                Add N_D to D-bin
                N_DRandom = random.sample(D)
                Add N_DRandom to D-bin
            end if
        end if
        i += 1
    end while
end for
```

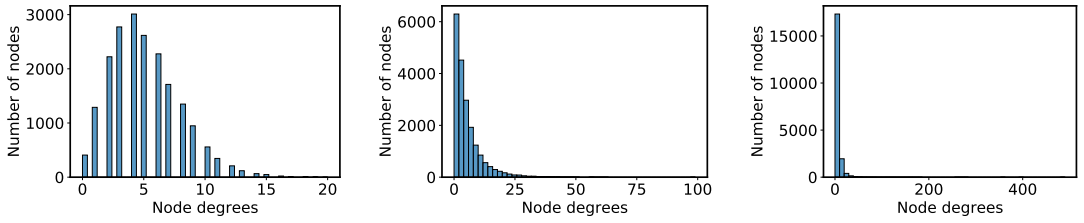


FIGURE 3.4: In-degree distributions of a dummy network with a source and destination group of 20000 nodes and 100000 edges. The Figure shows three degree distributions when $k_f = 1$(left), $k_f = 0.1$(middle) and $k_f = \frac{1}{N_{group}} = \frac{1}{20000}$ (right)

3.3.2 Reciprocity parameter

Since reciprocity is one of the basic features of most social networks, a reciprocity parameter has been implemented. This parameter, r, can range between 0 and 1, where the value equals the probability that, when an edge is laid between the source agent and the destination agent, an edge will also be created between the destination agent and the source agent.

3.3.3 Layer specification

Generating networks based on Algorithm 2 gives a general representation of Amsterdam's social network. The main issue with using only this algorithm is that the various layers are treated as if they are the same, while they should be treated differently.

The four layers each have different properties, such as multigroup networks in which the agents in one group are not connected to another, a varying reciprocity value and a influence of popularity. Other important properties such as a high transitivity and modularity value are unlikely when using only algorithm 2 as well as a edge overlap between different layers. Each layer is therefore subjected to a more in-depth analysis to map the network as accurately as possible on each layer and between layers.

3.3.3.1 Neighbourhood layer

The neighbourhood layer can be seen as a semi-multigroup model in which the network is partly dependent on the neighbourhood membership[28], this implies that agents from spatially different neighbourhoods have a small chance of being connected while people from the same area have a high chance of connection.

To make sure each agent can almost exclusively connect with agents from its neighbourhood we need to introduce an addition to the initial groups discussed in section 3.3.1.

Instead of creating groups based on only the agents characteristics, we should add groups based on area codes such that an agent has both a group ID and a area code. This way we can impose the restriction that two nodes can only be connected if they live in the same neighbourhood. The distribution of agents across neighbourhoods must be done in accordance with the total number of people and the fraction of each group living in a area. For this distribution the BBGA dataset[1] is used to find the total number people and the fraction of each group living in a certain area. Three different scales are used to describe the separate areas of Amsterdam in different sizes, these scales are: 22-areas, districts and neighbourhoods, with 22 ,98 and 480 areas respectively (see figure 3.5). The algorithm for the distribution of agents over the areas is shown in Algorithm 3.

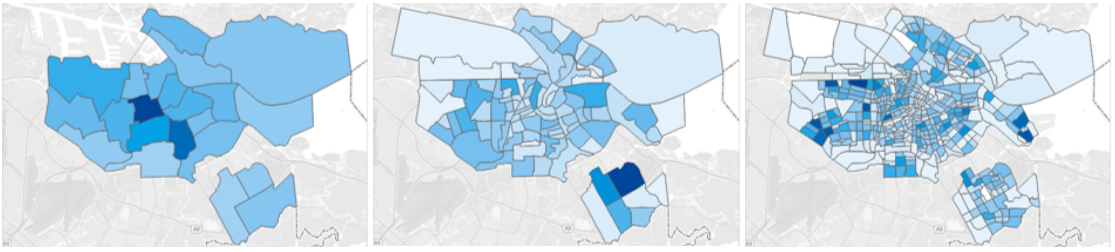


FIGURE 3.5: Amsterdam divided into 22 area's (left), districts (middle) and neighbourhoods (right).

Algorithm 3 Distribution of agents over areas

```
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
```

A down side of treating the neighbourhood network as a multigroup model is that agents of neighbouring areas 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 area. Also, if a node cannot connect to another node in its own area, we will keep moving to a neighboring area, until we find a connection.

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

TABLE 3.1: Attributes of household table.

abbreviation	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]

3.3.3.2 Household layer

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 at the same time and thus cannot have any connections with the agents from another household. This means that households can, be viewed as multigroup model in which your network is entirely dependent on your household membership [28].

Another critical aspect of the household network is that the households composition needs to be representative for the city of Amsterdam. To meet this requirement, the DNB Household Survey (DHS) dataset[2] from Centerdata was used. In their Household Survey panel, members annually provide detailed information, providing 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. It is therefore reasonable to believe that this data set accurately represents the composition of Dutch households. Teppa and Vis[32] 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 compositions. These attributes can be rewritten by the agent attributes described in Chapter 2 to make it competible with the data.

Although the DHS dataset provides an accurate representation of the household compositions, they do not show the right household distribution of Amsterdam. The statline dataset of the CBS is used[3] to get this distribution, which is showed in table 3.2.

[2]https://www.dhsdata.nl
[3]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

Based on the household composition and distributions we can introduce a new way to make the household network. The way this method works is as follows: first we sample the amount of households based on their size, (i.e 17401 households of size 5+, 32883 households of size 4 etc.). Then we sample from the DHS data a household which corresponds with the household size . Finally we can take agents from our agent dataset which fit in the description of the household members and put them in a household. A multigroup model is realized after all the households are filled .

We must ensure that the layers correspond to each other because we are working with a multiplex network. So, in the case of a household, we must ensure that all agents in the same house live in the same neighborhood. In order to realize this property, an agent in a household may only build a household connection to agents who live in the same region.

3.3.3.3 Family layer

People in the family layer, unlike those in the household layer, can belong to up to two families: a agent's own and the family of a potential spouse. As a result, this network is more entangled than a multigroup network. One factor to consider while designing this network is that many of the previously described household connections are also family, and hence many nodes require a link on both layers.

This is accomplished by retrieving the number of sampled families from the household, which is, according to the CBS, 140,000 households [4]. These connections in the household layer are reused for the family network.

3.3.3.4 Work/school layer

The work/school layer is the final layer. To obtain a better idea of how people are clustered in this layer, we separated the agents into those who work and those who

[4]https://opendata.cbs.nl/statline/CBS/nl/dataset/71486NED/tablefromstatweb

attend to school.To distinguish between the two groups, we must first determine how many agents aged 0 to 20 are enrolled in primary, secondary, and vocational schools. The following is a detailed approach:

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 statistics 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 statistics to see 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 statistics to see 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. Label these agents as students and assign them to "schools" which only they can attain.

Students are picked based on the process described above. In this layer, these students are grouped together in schools so that they cannot interact with agents who are working. There is a small number of working agents, such as teachers, who work on such schools. As a result, we allowed a fraction of working agents into these schools.

The schools are formed under the assumption that the ethnic character of the school corresponds to the ethnic composition of the neighborhood in which it is located. Although this is true for 8 out of 10 primary schools [8, 9], it is not necessary true for secondary schools and vocational education (MBO). Based on this assumption, we use the same data as in section 3.3.3.1 to create a school composition and assign agents to schools accordingly.

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

3.4 Sensitivity-analysis

Sensitivity Analysis (SA) is a widely used tool for determining the sensitivity of a models output to input parameters. This section will describe a Global Sensitivity Analysis (GSA) of the model using the PAWN method. The cumulative distribution functions (CDFs) and model indices are generated to provide an insight into the relative importance of each of the network statistics described in section 3.2.2. All four networks and the influence of the parameter settings, k_f, r, and layer specification, on the various layers are examined for the SA.

3.4.1 PAWN-method

The PAWN method [26] is a moment-independent method for performing GSA. It is described as producing robust results with small sample sizes [27] for factor ranking and screening.

The distribution of model outputs is investigated rather than their variation, as is common in other commonly used GSA approaches such as the sobol-method. The PAWN method distinguishes itself from other moment-independent approaches even further by characterizing outputs by their CDF rather than their probability distribution function. Because the CDF for a given random variable is typically normally distributed, PAWN is more appropriate when outputs are highly skewed or multi-modal, which variance-based methods may produce unreliable results for.

PAWN quantifies the variation in output distributions after conditioning an input to characterize the relationship between inputs and outputs. If distributions coincide at all S conditioning intervals, a factor is deemed non-influential. The distance between distributions is measured using the Kolmogorov-Smirnov statistic.

The median value is the most frequently reported value.

3.4.2 Parameter settings

The parameter settings, as described in Table 3.3, show the possible range of values for each layer's k_f, r, and $specification$ parameters, the $*$ represents the ranges parameters that cannot be adjusted if the specifidation parameter is not False. The values of the parameters k_f and r range from 0 to 1, whereas the value of the parameter $specification$ can only be True or False for the Work/school, Household and Family layer and 22-Areas, Districts, neighbourhoods and False for the neighbourhood layer.

TABLE 3.3: Parameter per layer. * = only when specification is False

Nr	Layer	k_f	r	specification
1	Work/school	[0,1]	[0,1]	[True,False]
2	Neighbours	[0,1]*	[0,1]	[22-A,Dist, neigh, False]
3	Household	[0,1]*	[0,1]*	[True,False]
4	Family	[0,1]*	[0,1]*	[True,False]

The rational behind this is the fact that the specification parameter provides a more detailed description of a layer. As a result, some parameters can only vary when the specification is set to False. For example, the connections in the households and family layer, as described by the data in Chapter 2, are symmetrical and thus is the reciprocity 1 when the specifications are True. A similar reasoning can be thought of for the parameter k_f, as no households, family members, or neighbors are likely to have significantly more connections than others in each of the corresponding layers, so when specifications is True we keep the parameter k_f for these layers constant at 1. The work/school layer differs in that these connections can be influenced by popularity, as evidenced by the study on collaborations [4] and interactions in karate club [41] and thus can vary when the specifications are True.

3.4.3 Results

3.5 Discussion

This paper demonstrates a method for constructing a social network of Amsterdam with various layers in numerous ways, all of which together form a family of networks based on the same or partially the same data. The method heavily relies on the 2GN algorithm, which allows us to adjust the degree distribution based on a parameter and create a network between two groups.

The results attained by the method shows that differences in parameter and layer specification can result in significant differences in network statistics. It observed, for example, that the parameter k_f influences the in-degree distribution while the r parameter influences the reciprocity in a network. Furthermore, it can be established that configurations representing a social network were discovered that are reasonably consistent with the theory and can thus be regarded as representative.

However, it should be noted that assumptions were made, particularly in the case of the layer specifications. For example, the ethnic composition of schools is assumed to

correspond to the ethnic composition of neighborhoods, and the DHS dataset is used, which is representative of the Netherlands but may differ from that of Amsterdam. Workplaces were also chosen at random, with no consideration given to the composition of people in businesses. As a result, this network should be regarded as an approximation of Amsterdam's social network rather than the truth. The data also contains a number of unusual values, such as the outliers in the ethnic group "other." The people who fall into this category are very diverse, so classifying them under one denominator is not entirely correct. Finally, this network excludes important social connections such as friends, sports and online influences. While many of your friends are at work or school, friends outside of these institutions are excluded, and thus some strong influences are overlooked.

Although this family of networks is not a perfect representation of Amsterdam's social network, it is a good approximation. As a result, these networks can be used to mimic the effects of different laws, such as a lockdown. It can also be used to track the spread of socially contagious processes like obesity through Amsterdam. By using different network configurations rather than just one network, such a spread can be properly mapped and the model tested for robustness.

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. Linked: The new science of networks, 2003.

[2] A.-L. Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.

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

[4] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4):590–614, 2002.

[5] V. Barash. The dynamics of social contagion. 2011.

[6] G. Bianconi. *Multilayer networks: structure and function*. Oxford university press, 2018.

[7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[8] 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.

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

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

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

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

[13] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[14] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical review letters*, 90(5):058701, 2003.

[15] 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.

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

[17] I. de Sola Pool and M. Kochen. Contacts and influence. *Social networks*, 1(1):5–51, 1978.

[18] R. I. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of human evolution*, 22(6):469–493, 1992.

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

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

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

[22] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[23] 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.

[24] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[25] M. E. Newman. Random graphs with clustering. *Physical review letters*, 103(5): 058701, 2009.

[26] F. Pianosi and T. Wagener. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67:1–11, 2015.

[27] F. Pianosi and T. Wagener. Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling & Software*, 108:197–207, 2018.

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

[29] RIVM. Overgewicht nederlandse bevolking. *https://www.volksgezondheidenzorg.info/onderwerp/ou context/samenvatting*, 2021.

[30] M. Rogers Everett. Diffusion of innovations. *New York*, 12, 1995.

[31] M. Serafino, G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J. R. Banavar, and G. Caldarelli. True scale-free networks hidden by finite size effects. *Proceedings of the National Academy of Sciences*, 118(2), 2021.

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

[33] V. A. Traag, L. Waltman, and N. J. Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.

[34] J. Travers and S. Milgram. An experimental study of the small world problem. In *Social networks*, pages 179–197. Elsevier, 1977.

[35] H. van Kolfschooten, R. Neerhof, A. Nijboer, A. de Ruijter, M. Visser, et al. Juridisch instrumentarium voor een gezonde voedselomgeving in de stad. 2020.

[36] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov. Scale-free networks well done. *Physical Review Research*, 1(3):033034, 2019.

[37] Y. Wang, H. Xue, H.-j. Chen, and T. Igusa. Examining social norm impacts on obesity and eating behaviors among us school children based on agent-based model. *BMC public health*, 14(1):1–11, 2014.

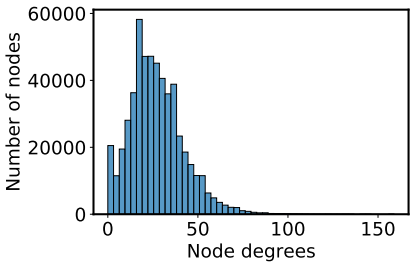
[38] S. Wasserman, K. Faust, et al. Social network analysis: Methods and applications. 1994.

[39] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

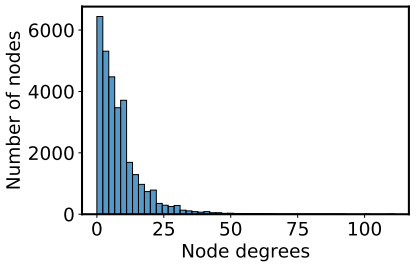
[40] WHO. Obesity and overweight. *https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight*, 2021.

[41] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.

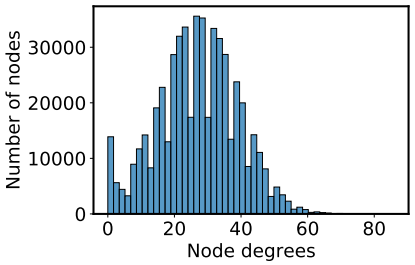
.1 Constructing a network of Amsterdam



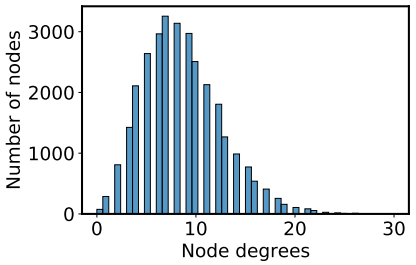
(A) Configuration 1, whole network



(B) Configuration 1, network within one group

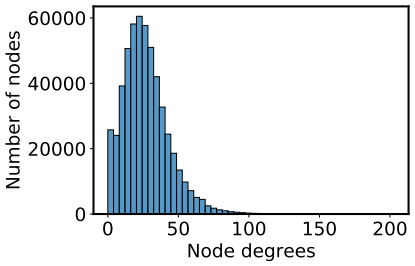


(C) Configuration 2, whole network

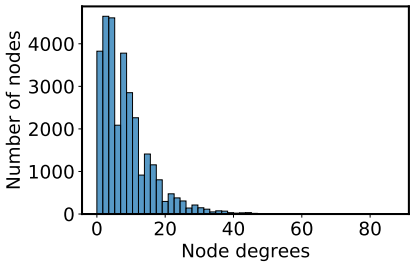


(D) Configuration 2, network within one group

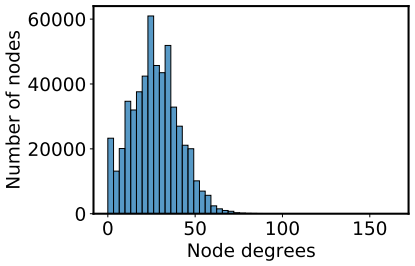
FIGURE 1: Difference between configuration 1 and 2 in terms of in-degree distribution, for both the whole network (A,C) and the within a group (B,D).



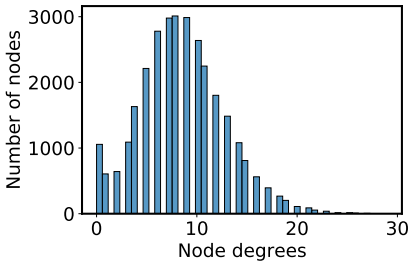
(A) Configuration 3, whole network



(B) Configuration 3, network within one group

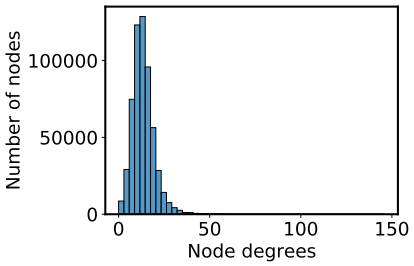


(C) Configuration 4, whole network

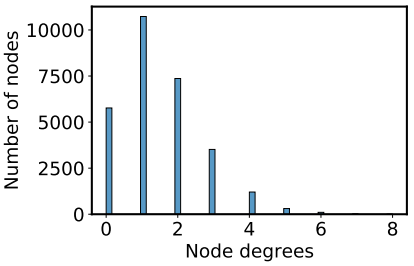


(D) Configuration 4, network within one group

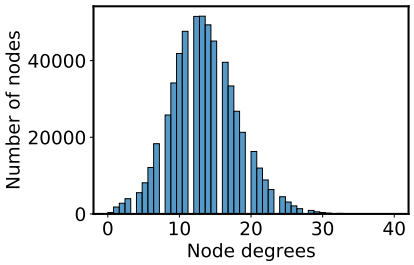
FIGURE 2: Difference between configuration 3 and 4 in terms of in-degree distribution, for both the whole network (A,C) and the within a group (B,D).



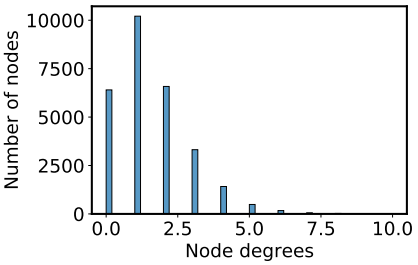
(A) Configuration 5, whole network



(B) Configuration 5, network within one group

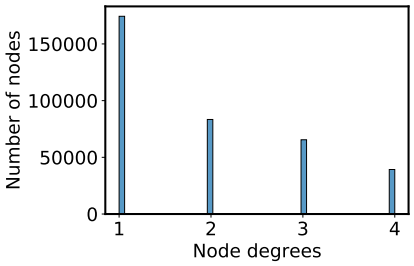


(C) Configuration 6, whole network

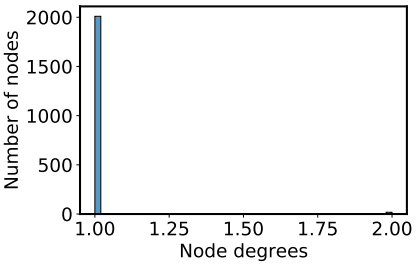


(D) Configuration 6, network within one group

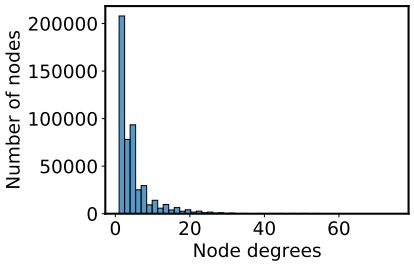
FIGURE 3: Difference between configuration 5 and 6 in terms of in-degree distribution, for both the whole network (A,C) and the within a group (B,D).



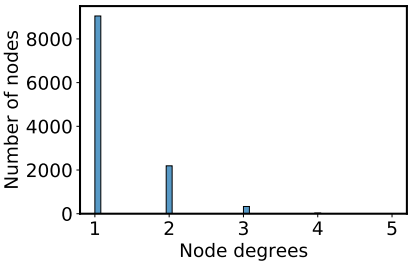
(A) Configuration 7, whole network



(B) Configuration 7, network within one group

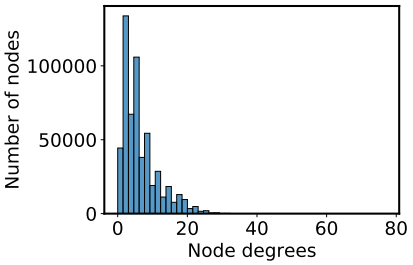


(C) Configuration 8, whole network

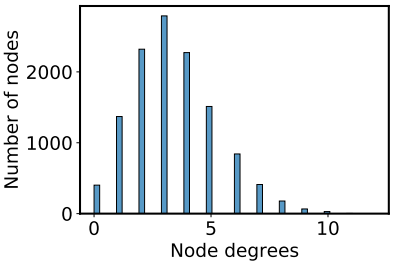


(D) Configuration 8, network within one group

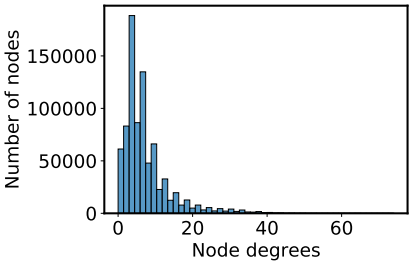
FIGURE 4: Difference between configuration 7 and 8 in terms of in-degree distribution, for both the whole network (A,C) and the within a group (B,D).



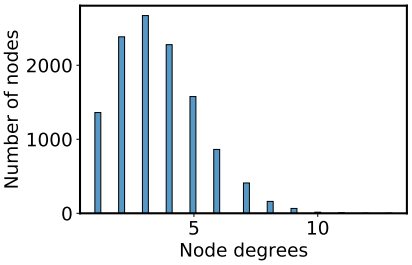
(A) Configuration 9, whole network



(B) Configuration 9, network within one group



(C) Configuration 10, whole network



(D) Configuration 10, network within one group

FIGURE 5: Difference between configuration 9 and 10 in terms of in-degree distribution, for both the whole network (A,C) and the within a group (B,D).