Part I

Foundations

"Then two wonders happened at the same moment. One was that the voice was suddenly joined by other voices; more voices than you could possibly count. They were in harmony with it, but far higher up the scale: cold, tingling, silvery voices. The second wonder was that the blackness overhead, all at once, was blazing with stars. They didn't come out gently one by one, as they do on a summer evening. One moment there had been nothing but darkness; next moment a thousand, thousand points of light leaped out – single stars, constellations, and planets, brighter and bigger than any in our world. There were no clouds. The new stars and the new voices began at exactly the same time. If you had seen and heard it, as Digory did, you would have felt quite certain that it was the stars themselves which were singing, and that it was the First Voice, the deep one, which had made them appear and made them sing."

C.S. Lewis, *"The Magician's Nephew" (1955)*[1]

[1]THE MAGICIAN'S NEPHEW by CS Lewis ©copyright CS Lewis Pte Ltd 1955.

1 Introduction

> *"The future is something which everyone reaches at the rate of sixty minutes an hour, whatever he does, whoever he is."*
>
> — **C.S. Lewis**
> "The Screwtape Letters" (1942)[1]

1.1 Motivation

More than ever before, we are connected. In the European Union, 63% of Internet users between the age of 16 and 74 used social media in 2016. If you look at the younger generation between the age of 16 and 24, 9 in 10 Internet users participated in social media according to Eurostat[2]. Some effects of the adoption of social media usage has been reflected in recent events with global interest. In the United States of America, the social media based Obama campaigns of 2008 and 2012 dragged millions of people to debate and participate in political life. In 2011, the Arab Spring shed light on the impact of social media for civic acts.

The increasing use of social media and the impact it's causing in society has attracted the interest of scientists from many different fields, for example from the social sciences, psychology, neuroscience, computer science, data science, etc. Some of these studies are interested in making predictions, such as who is going to win the next elections or how the market of shares is going to behave next week. Other studies are interested in evaluating the impact of interventions for the promotion of behaviours in groups, e.g. increasing the amount of physical activity or motivating a healthy food intake habit. All these processes exemplified above are connected by networks, and require more than social network analysis alone to explain the phenomenon behind the cases listed above.

The exchange of ideas, beliefs, sentiments and behaviour in networks can be better understood by social contagion theory. Social contagion can be observed in people's interactions and has been confirmed by recent neuro-cognitive discoveries [26, 27]. These discoveries shed light on the mechanisms of the brain that can help explain why and how people connected in a social network affect each other through their interactions. The social and health sciences have also shown consistently that relationships can shape people's mentality and/or behaviour [15, 38].

[1]THE SCREWTAPE LETTERS by CS Lewis © copyright CS Lewis Pte Ltd 1942.

[2]http://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/DDN-20170713-1

This thesis intends to expand the knowledge about social contagion and come up with realistic and trustworthy explanations that can be simulated or verified with real data concerning the spread of physical activity behaviour, political positioning and panic in disaster situations. This is done by creating new ways of learning about real life through a network-oriented perspective. The knowledge developed here can be applied to find potential applications to assist people and provide a better quality of life. The work presented here requires a multidisciplinary framework where computational models are used to develop realistic analyses of real life situations. This is done by unfolding possible ways to understand, explain and predict individual habits that impact and build group behaviour in different scenarios.

Therefore, this thesis aims to understand, model and predict different sorts of behaviour through cognitive models and social contagion in social networks. The models developed here can be applied broadly from promoting healthy lifestyle to reactions to web media posts. Data analysis is also used for understanding how the physical activity of a group of adults and a group of children evolve over time.

The following section of this chapter introduces the main concepts and basis for the presented work. Section 1.3 presents the research questions that have guided this work. Section 1.4 presents the methodology applied in order to develop the research and obtain the results. Section 1.5 presents the structure of the other chapters of this thesis.

1.2 Background

The concepts presented below are important to contextualize and enrich the understanding of the following chapters of this thesis. On section 1.2.1 data analysis methods and its applicability to this research is discussed. Section 1.2.2 will explain the theory behind social contagion and how it can be used to explain reality within many contexts. Section 1.2.3 will present the modeling techniques used throughout the thesis to represent scenarios involving the social relations and cognitive aspects of humans. Lastly, Section 1.2.4 will explain the relationship between research on social networks and research on health. It is also shown how this research contributes to the task of predicting the effect of interventions on people's health and understanding the dynamics of social contagion regarding physical activities.

1.2.1 Data Analysis

Recently, much attention has been given to quantitative research due to the increasing amount of data and the development of powerful computational tools for data analysis. Besides statistics, the ground for data analysis has become stronger with the advances in computer science. The popularity of social media websites, combined with the increase of mobile device usage and online communication has facilitated the gathering of people's behaviour data. Learning from personal data has become less intrusive as people leave their "fingerprints" when using social media, or tracking devices, like physical activity devices or smartphones freely.

The use of these big databases of personal interactions and information has for many years been of interest to computer science, social sciences and health sciences. Even though the amount of data is abundant, the quality of the data can present serious limitations to the exploration and analysis processes [31]. To find patterns in data it is required that the data is manipulated, processed, cleaned and crunched. The task of explaining data or validating theoretical models presents a number of limitations that can be challenging for any data analyst to overcome, i.e. missing data, low periodicity of the data collection process, etc.

Besides the limitations imposed by the social media websites regarding the amount of data that someone can collect (if they can), other problems can emerge when scientists are creating their own environment for data collection. To collect longitudinal data, for instance, regularity is required on the data collection and very detailed information about the dynamics of the traits and relationships within the population observed. At the same time, it is a very hard task to track subjects in a very depicted way, as the quality of the data collection implies more intrusion and subsequently higher taxes of participants dropping out of the experiment, or feeling violated in their privacy.

Ethical aspects are also very relevant when developing research on data analysis. Charles [12] gives some good examples on how data analysis can be used to shape consumers' habits and its side effects. Besides increasing the profit of companies, cases such as the mega company Target is one of the examples on how not to use data analysis. Target collected data on their consumers regarding their grocery shopping habits through the data provided by membership cards and credit cards for more than a decade. The data analysis department developed algorithms to detect if the costumer was single or married, if they have kids at home, or even if they were moving house, or pregnant. The pregnancy algorithm would trigger vouchers for items related to the baby care for the mother to be. In one situation, the father of a teenager girl started receiving vouchers for pregnancy items after his daughter used his credit card for shopping, bringing up a very embarrassing (if not unethical) situation for the company.

In this thesis we use a couple of data sets related to physical activity and other behaviours of heterogeneous groups of people. All the data sets were collected respecting the privacy of the participants and following research methods that will be further explained in Section 1.4.

This current section presented the data analysis background and relevance for part of the data oriented research done and shown in this thesis. The knowledge of data analysis techniques and its limitations were important to define the methods used in Chapters 3, 4, 5, 6, 7, 8 and 11. The techniques developed in data analysis and its methods are combined with the network-oriented models throughout this thesis to bring up knowledge on how can we understand, model and predict human behaviour changes through social contagion models.

1.2.2 Social contagion

Social contagion is the concept that individuals tend to follow the same ideas, sentiments and/or behaviour as those with whom they communicate, their social network. This theory assumes that people don't need to have the intention or awareness to affect others. The process happens involuntarily [44]. In this sense, the relationships, or ties, are very important when studying social contagion, as it is through the connections that people are influenced in their attitudes.

The assumption that social contagion is real in our daily lives led many scientists to research how it can be understood within a large range of fields, such as the spread of obesity [14], smoking behaviour [13], alcohol consumption [29], happiness [21], depression [42], divorce [34] and new products diffusion [28].

The recent discovery of the mirror neurons and the role they play in our brains also gave a bold and solid basis to scientifically claim that we have a biological mechanism that explains why we cry when watching a drama or reading a novel, or why we reproduce violence when raised in a violent environment [26]. These neurons are important for the imitation learning system. That is, humans can learn from imitating others, or from actions done by other individuals. Therefore, mirror neurons mechanisms provide the ground to model and validate scenarios of people interacting and exchanging emotions, opinions, information and letting themselves be affected by others. Other mechanisms are also important to understand how people react when receiving information of others' actions. The suppression mechanism, for instance, helps us to avoid shameful situations, like firing back when your boss is angry about some mistake made. To combine all the mechanisms of the brain without losing the knowledge from the other mechanisms, many scientists chose to apply network-oriented models, as they are equipped for loops and potential cross-relations between different parts of the human cognitive system. These models can become naturally complex, just as the brain is.

Most of the work in this thesis is based on the understanding and modeling of social contagion. The current knowledge about mirror neurons, social contagion, neurological responses for threatening situations is extensively applied. This research brings all these topics together, trying to understand how people behave when interacting on web media (Chapter 10), or in a disaster situation (Chapter 9). These tools are also used for modeling physical activity behaviour spread, both in adults and kids, attempting to validate the social contagion model proposed to predict potential improvement or degradation of people's lifestyles (Chapters 4, 5 and 7). A cognitive model for political positioning presented in Chapter 10 brings knowledge from neuropolitics and psychology together to understand how political opinion is shaped and affected by the burst of daily information received though social media.

In Chapter 2 a contagion model is proposed based on the previous work of Bosse et al. [9]. This model is the basis for the other simulations throughout the thesis.

1.2.3 Modeling

Modeling has been present since the first registries of humans from the stone age onwards. There are registries of new models created and used to improve every-day life by the time of 2.000 BC in at least three cultures, including the Egyptians and the Indians [45]. The art of making representations of reality, or copying it was further transferred to the mathematical field with the representation of numbers using bones, up to our present time, where computational models are used to simulate all sorts of real scenarios.

Computational models can be understood as models that use computers as the tool to be simulated. The power provided by the new computational machines created in recent years permit complex and massive simulations to be performed in a feasible amount of time. Helbing [24] adds that computational models can complement classical research methods in the social sciences providing a tool to test if theoretical models and frameworks explain observed phenomena and provide reliable explanation of the reality. Computational models and simulations are very useful for testing hypothesis, performing analyses of many sorts of scenarios, and for prediction. Usually these models are accompanied by real-life data for validation purposes. Furthermore, computational models can be useful in guiding data collection, revealing dynamic analogies, demonstrating trade-offs, decision support and many other reasons [20].

This thesis is focused mainly on creating and testing models that can explain human behaviour, perceptions and emotions concerning many aspects, from physical activity to political opinions. Different modeling approaches are also used according to the context where it is applied, from individual cognition characteristics to a group's healthy lifestyle increase. Therefore, here we explain the importance of these modeling approaches and what the key concepts are for each of them.

Agent-Based Model (ABM) is a modeling method in which agents and their interactions with each other or the environment are accessible as a program or in a physical structure, like a robot. In ABM, the agents can be representing people, animals, government agencies, groups of cells or even countries, and have their own characteristics and actions [18, 39]. The agents can also be entities that have no representation in real life, with the purpose to gather information from the environment, for instance.

ABM is a very good method to use especially in social sciences and behaviour theory tests. In these cases, the behaviour of the entities and dynamics of the interactions are formalized by algorithms containing equations and decision rules, making the modeling of the behaviour more flexible [24]. The flexibility provided by this method allows simulations to provide heterogeneity, where each agent can have its own characteristics and behaviour, and the possibility of stochastic scenarios where random events can be simulated. The flexibility of ABMs can be extended to use it with other kinds of models [19, 40, 47]. ABMs have been extensively used in simulations of the spread of diseases [40], policies for food production [46], emotions [8, 10, 48], social influence and opinion formation [33, 50], etc.

Complex systems is a new science that still lacks a solid definition. Latora et al. [32] roughly defines complex systems as "a system made by a large number of single units (individuals, components or agents) interacting in such a way that the behaviour of the system is not a simple combination of the behaviours of the single units". The main characteristics of complex systems is the lack of a central control. That is, the overall behaviour of the system is not derivable from the individual entities within it. In this case, the interactions between the components of the system are extremely relevant to better understanding what is happening in the whole complex system [51]. The complex systems are used for modeling and prediction in animals' behaviour, economics, politics, neuroscience, etc. [3, 5, 32, 51]. Over the years there has been an effort on understanding the characteristics of the individuals in the many contexts, as well as how they interact with each other. The net of interactions and how the individuals affect each other is named as the backbone of the complex system. When the backbone is described in terms of nodes and edges we have the complex network of the complex system [32]. The **complex networks** are studied within a new science discipline, the network science. It is basically the study of social interactions, social web media, the neurons in the human brain, and any other systems composed by a large amount of interconnected nodes through very complex arrangements. It is a multidisciplinary field, where the challenges are mainly related to the understanding of the evolution of connections, or in the spread of information throughout the system. As the networks are context-based, the knowledge from this field can be applied to studies in diverse fields, e.g. from virus and biology to global political patterns within countries [32].

The complex network tools are very useful for behavioural observations and analysis. The advances of social network analysis and complex systems brought ideal tools to evaluate these scenarios beyond the pure statistical analysis as in the past.

Understanding the dynamics of complex networks provides knowledge about the many features that can describe social relationships, such as who befriends whom and why. It is also possible to predict potential individual problems, like loneliness, or burnout, based on the amount (or the lack) of interaction with other people at work, at home or with friends. Collective behaviour can also be studied using these tools, like reactions in emergency situations, or the spread of ideas. It is possible to understand what the means are where ideas are more efficiently diffused, like political ideas, or simple information about facts happening in the world.

Ultimately, complex networks are not a completely new discovery, but a new way of looking at the world and interpreting it. It is a change in perspective that allows scientists to build more realistic views of what we call reality, and to provide tools that allow us to read this reality within new boundaries. The contexts studied in this thesis are very suitable for a complex network representation, as it is going to be shown throughout the following chapters. The spread of behaviours, perceptions and opinions requires knowledge about the individuals in a group and the connections between them.

Within the context of the complex networks, Treur [54] presents network-oriented modeling using temporal-causal models as a very suitable alternative to represent the complexity of human and social processes. Science has built more information on cognitive, affective and social neuroscience fields, permitting that more complexity can be added to the modeling of the mechanisms of the brain [54]. Network-oriented modeling can be inspired by neurological and social events. This paradigm intends to bring together processes that are traditionally studied separately, like cognitive and emotion processes, or individual and collective phenomena. It is known that many human processes involve sub-processes that run in parallel and/or in cyclic to the main process. When using network-oriented modeling, it is possible to represent all the sub-processes without isolating or ignoring part of them. It also permits that a time dimension can be incorporated to the model, which can be useful for timing the processes and generating more realistic simulations.

For this sort of modeling, states (nodes) and connections (edges) are the main actors. States contain activation levels that will determine when the state is triggered or not. The combining functions of the states reflect the aggregated impact caused by the other states connected to a specific state. The connections determine causal relations between the states, and influence the volume of information transmitted between states depending on its weight. The use of network-oriented modeling with **temporal-causal networks** incorporates the semantics of the network necessary to account for the dynamics of the states and connections whereas providing a temporal dimension to the simulations [53]. Even though two different models can have the same isomorph graph representation for the nodes and edges, as exemplified by Treur [53], more information is needed to provide the dynamics of the process being modeled. The information includes how the connections between the nodes are interpreted, or what does a node mean. Without the semantics of the mechanisms present in the structure of the network, we could consider it as a simple graph theory problem. The semantics permit that the dynamics of the states can reproduce diffusion or contagion within a network. The same way, the semantics provide dynamics to the network structure, permitting that adaptive or evolving networks can be represented as such. All there dynamics are connected through causal relations, as they represent the chain of interactions between nodes and/or edges. The perspective of changes in the phenomena studied in this thesis depends also on a temporal structure as each state (or node) affect each other over time. To incorporate the dynamics and temporal aspects of such a modeling approach, differential equations are used, and the network generated with the semantics of these network characteristics are called **temporal-causal network** [53].

This thesis presents works that can be categorized as ABMs, complex systems and network-oriented modeling with temporal-causal networks. All the tools provided by these areas are very useful when modeling how social contagion changes the characteristics of a group of individuals connected to each other. Most of our studied scenarios contain people in networks where their physical activities are being shared (i.e. Chapters 3 to 8). Some of the scenarios are related to the spread of information, like messages in disasters (i.e. Chapter 9). Some of the systems simulated describe the cognitive aspects and the brain of humans when interacting with information in web media (i.e. Chapter 10). In each case, the states (nodes) are defined according to the context where the model is built, together with the ties between the states. The areas explored by this research are neuroscience, psychology and health sciences.

1.2.4 Health and Social networks

Much of this thesis is dedicated to explore how a person's healthy lifestyle is affected by their connections in a network. It also tries to predict intervention targets that can enhance the overall physical activity of a group. The issue of the spread of physical activity throughout networks is a relevant topic that addresses the use of social networks to promote health in a population.

Social network studies have existed for decades, but have received more interest and publicity in recent years, due to the growth of the Internet and other communication devices that use network as a core feature [6, 11, 16, 17, 57]. Social network studies have also been important to explain behaviours that were unexplainable using the classical theories of behaviour, like why one person quits smoking, while another person doesn't [13].

Almost every health topic can be described as a social network [55]. Examples include HIV/STDs transmission networks [37, 43], drug addiction [56], smoking [2, 13], suicide [4, 41] and obesity [14, 15]. Many other studies provide models that can account for health issues in many aspects, from the spread of diseases to the diffusion of good preventive information, from depression and loneliness tendencies to suicide ideation.

Including the relations as part of the model is important to fill the gaps that other studies in behaviour lack. As social beings, we each have different traits, character and opinions. But we are also largely influenced by our relationships when it comes to how we behave or what we believe. Knowing who is a friend of whom, and how much time is spent between people can tell a lot about many aspects that could not be addressed by simply knowing the characteristics of someone. It is known and much study has shown that there is a tendency for people to befriend others who are like themselves. This is called *homophily* [16, 36, 52]. This fact takes us to one of the biggest challenges when using social network models to explain people's behaviours: while random sampling was acceptable in the classical behavioural science, to address the effects of the network on individuals requires a more careful and strict approach.

Random sampling can remove the entire net of relationships of the individual, requiring that other methods to collect network relations are applied, like ego centric data collection techniques [16, 25]. Chapter 4 used this technique to address the social network of a group of bachelor students, starting from one individual's contacts.

Besides the difficulties in data collection, social network models require that the connections present a reliable metric, where you can differentiate acquaintances from spouses, and spouses from best friends. The frequency of contact is also relevant, as it is the dynamics of the network. Not only the individuals change, but also the connections.

Other prominent studies on connections of people is related to potential interventions in order to change people's behaviour. This kind of work is becoming increasingly important, as the health situation of the population is a big issue for governments and society in general. We attack this problem in Chapter 8 when trying to predict which kids are the most ideal to apply interventions in a network of students from schools in the Netherlands. For this purpose, it is relevant to know who are the more isolated individuals, who are the bridges, who are the most connected, or the ones who are more centralized and verify which selection of nodes delivers better results.

This thesis dedicates a big part for understanding the spread of health behaviour throughout social networks. The aspects of data collection, analysis and difficulties in this sort of research are addressed in Chapters 4, 5, 7 and 8.

1.3 Research questions

The aim of this research is to understand how behaviour, perceptions and emotions are spread in social networks on many levels, from social interactions to the cognitive system and its reaction towards the external world. We use artificial intelligence methods combined with data analysis and neuroscience. Ultimately, we want to address the following question:

How can we create and validate computational models that explain social influence and contagion in social networks?

This overall question can be subdivided in the sub-questions below.

1. How can we design and use temporal-causal models based on networks to better understand and describe social contagion taking into account personal characteristics?

The first question consists of a process of depicting a phenomenon (the social contagion) in many parts. The first part consists of the design of social science experiments to collect social network data that can provide us with a temporal-causal structure, both for the states of the individuals and for the connections between them. The second part consists of treating the data collected, running the model and adjusting the parameters to find a better fit between the designed model and the empirical data. The third part is based on the reflection on the important aspects that describe the studied phenomenon. It involves adapting the models to better describe reality, or propose new methods that incorporate other disciplines (i.e. social sciences) in the process of building up traits of individuals and the net.

2. How can we predict changes in behaviour using the relationships and data related to physical activity and how can we measure social contagion in a social network regarding people's PAL?

Besides describing the dynamics of interactions in a social network, predicting the future of the structure of the states and connections is also very important. In this work we used the Physical Activity Levels (PALs) as the diffusion unit of behaviour. In this sense, we are interested in verifying and predicting how the PALs are affected based on the social connections in a network. To achieve the results we wanted, a combination between differential equations and agent-based modeling in complex systems were used. The quantification of the amount of social contagion is still a very open research question, and we therefore explore it in order to shed light in this matter.

From the group's perspective, we also want to know if connectivity is a factor that can describe the increase (or deterioration) on PAL of people. And if that's the case, we want to identify influential nodes in the network in order to promote healthy lifestyle and optimize the spread of behaviour.

3. What are potential applications for modeling behaviour in social networks and how can we apply the knowledge of temporal-causal network modeling using different contexts and methodologies?

The third question is related to the extension of this approach to other contexts, such as the description of phenomenons at a cognitive level. Here we want to investigate if the method is useful for other contexts besides the PAL. Thus, we have to understand internal (cognitive) reactions for people in various situations, e.g. in disasters, or when interacting with web media. We also want to investigate the cognition behind political opinions using temporal-causal models based on networks. As part of this process, we also need some tools to help in the data collection, such as classifiers for political content.

1.4 Research Methodology

This section explains the methods used throughout this thesis to achieve the understanding to answer the questions presented in the section above.

1.4.1 Systematic review

As stated by Akerjordet and Severinsson [1], "a systematic review involves the identification, selection, critical analysis and written description of existing information". Multidisciplinary research demands a huge effort on reading and understanding the state of the art of many phenomena involved in any sort of simulation or analysis. Due to the many disciplines involved throughout this work, systematic review was used constantly in order to understand concepts and developments from the fields of neuroscience and psychology.

The models created for political positioning (Chapter 10) for sharing behaviour on web media (Chapter 11) and for disaster situations (Chapter 9) required extensive research on many different disciplines. We used a large body of literature to build models that describe the cognition of people interacting with web environments or a person's reaction in a specific context of disaster information spread. The literature shed light on the knowledge about the main mechanisms behind these scenarios and how much is known about the brain functioning when it comes to these topics.

The studies on health and physical activity also demanded a large amount of work to understand the methods to quantify and interpret lifestyle of people. A good amount of systematic review was done especially when trying to understand the dynamics of changes in the behaviour of children in Chapter 8, and also when developing measurements for physical activity level among young adults in Chapter 4.

Other studies related to social sciences were also relevant to assess personal traits of participants in experiments created in Chapter 3. For this reason, we invested significant time in building knowledge on which methods are used and how efficient they are when handling social experiments.

1.4.2 Data collection, processing and analysis

Many governments, companies and researchers became intrigued with the usefulness of the data provided by users not too long after the emergence of the Internet in the 1990's. The increasing number of websites and the boom of social media increased on a large scale the amount of data available, and this was mainly behavioural and relational data. Even a simple email can contain contextual information and provide data for tracking some kinds of behaviour or relations between sender and receiver. A discussion over a post on Facebook involves many agents that can be connected as 'friends' or not, as well as showing the tone and reactions to every response.

Besides from all the online data provided on the Internet, offline data has also increased with the miniaturization of the sensors and electronic devices [22]. Activity trackers can follow people without any need for connection throughout the day, providing a good amount of information about how active people are.

In some parts of the work in this thesis we used Fitbit One devices as a way to collect data from participants. In other parts of the work shown in this thesis, other devices were used, always aiming to gather data of people in social relations. This data went through the process of data analysis, described below.

To collect relational data for part of our data set we used questionnaires such as the Big Five Inventory, and based our methods on previous works from social sciences. When the connections happen in a virtual environment (i.e. the Internet), we used the data from the logs of ties built over time. The information of the connections is very important to build the social networks, as they are the means through which the spread of behaviour happens.

The data analysis required for this thesis involves the use of statistical methods to evaluate differences between groups of people connected or not connected (i.e. Chapter 5), or linear regression for the matter of comparing with other models and finding trends on the individuals' PAL data (i.e. in Chapters 4, 7 and 8).

The pre-processing and cleaning of data is also very important when analyzing data. In many of the following chapters the reader will learn how the data was selected to draw conclusions about the behaviour of the participants in the experiments. The choices about how to select the data was done with the utmost care to not lose information and at the same time bring a reliable interpretation of the phenomena studied.

Most of the work was done using Python and Jupyter notebooks [30]. Python is a very powerful tool as it permits that data can be read, cleaned, processed, and statistically analyzed without having to change the environment. Python also provides tools to save results, plot graphics, generate agent-based modeling codes, etc. [35]. The Jupyter Notebooks facilitate the visualization of the code and the results. It also permits the graphics to be extracted straight from the interface of the code. Some of the Chapters will provide the link to the notebooks that were used for the data analysis, which also facilitates the verification and questioning of the methods used. We also used R for some of the statistical analysis, such as the multiple linear regression model and the linear mixed model of Chapter 5.

1.4.3 Social network analysis

As discussed in Section 1.2.4, 'social networks' are how we name the representation of actors and interactions in a formal way. In other words, social networks are the way relational data is stored or represented. Differently from the classic sets of data, which are focused on the attributes of people, the study of social networks requires a new set of techniques that incorporate the relational characteristics of the data [16, 49]. This is where Social Network Analysis (SNA) becomes important. According to Crossley et al. [16], SNA is "a set of interconnected concepts, theories and techniques developed for the most part within a relatively cohesive, interdisciplinary research 'network', devoted to gathering and analysis of relational data".

The two essential elements of social networks are a set of nodes and a set (or sets) of ties. Both elements can contain specific attributes, and the presence of multiple sets of ties can be useful to represent what is called multilayer networks. Some concepts are also important for the SNA. The connected components will tell us which subset of nodes are linked by a path, and the centrality can tell us about the level of centrality of each node in the network. Centrality can be measured in terms of degree, paths, eigen vectors and other metrics that can describe structural characteristics of the nodes. For a good understanding of these metrics and the concepts behind SNA, see [7], [25] or [49].

In this work, SNA is present in several chapters as ways to better understand the structure of the networks where social contagion is going to be evaluated. In Chapter 4 we evaluate the structure of a network of young adults from the same undergraduate class collected using a snow-ball method [23]. In Chapter 6 we evaluate the structure of a relational data set of individuals in a health promotion program in order to understand the dynamics of the creation of ties and the evolution of the whole network. In Chapter 8 we apply SNA to understand the structure of the network of children in Dutch schools for a study on health and physical activity propagation through the ties of the kids.

1.4.4 Computational modeling

Section 1.2.3 explained the modeling concepts used in this thesis. Chapter 2 covers the explanations of the core of the contagion model used in many of the chapters in this thesis. The models presented here are based on a network-oriented approach, meaning that the mathematical representation of the events are done using nodes as agents or states, and the ties represent the relationship or the influence between two edges. Some of the studies are in the context of health, and some are related to behaviour and cognition, as explained below.

We applied the model presented by Bosse et al. [9] for the spread of physical activity in networks in several chapters:

- In Chapter 3: to find the traits of the agents in the models using parameter tuning techniques.
- In Chapter 4: we try to verify if the model can predict the increase or degradation of PAL for a group of young adults from the same class.
- In Chapter 7: we try to fit the model to account of a big network of participants in a health promotion program.
- In Chapter 8: to predict behaviour change among children from Dutch schools.

The concepts of network-oriented model are also applied to different contexts in Part IV of this thesis. In Chapter 9 we define a model to describe people's reactions when receiving messages related to disaster situations from a neuro-psychological perspective. Chapter 10 addresses the political positioning on web media through a cognitive model based on psychological and neuroscientific discoveries about the brain.

1.5 Structure of this thesis

This Section presents the organization of this thesis. As a cumulative research, the chapters can be read individually or as a whole body of work.

1.5.1 Part I: Foundations

The first part of the thesis aims to give the background and introduce the reader to the topics of social contagion, social networks and complex networks. It is expected that the reader can grasp the fundamental theories that guide the other chapters and that provide the basis for the whole work.

Chapter 1 (current chapter) presents the literature review and the questions that are addressed in this work. It also describes the methods used to build up the contributions of the thesis.

Chapter 2 presents the adaptations proposed in the absorption model of emotions previously created. The goal of this chapter is to present the mathematical structure of the contagion model and the improvements proposed in order to deliver more stable results when applying it within the different contexts.

1.5.2 Part II: Data Analytics in Networks

Part II is dedicated to the studies regarding physical activities in networks. For this part, we are interested in understanding how physical activity levels (PAL) are spread in networks and how we can quantify the PAL and predict future states based on social contagion models.

Chapter 3 is about a method proposal to derive personality traits based on empirical data and social contagion. The method includes machine learning algorithms to fine tune the parameters and contrast them with methods from social sciences based on intake questionnaires.

Chapter 4 presents a social experiment involving social contagion of PAL in young adults. The work contains the data collection, analysis and study of the applicability of the social contagion model on this data set.

Chapter 5 shows the results of a data analysis of a data set with originally around 5.000 participants in a health promotion program. The group was connected through a social network where they could visualize their friends' PALs and compare themselves with their peers. For this work we evaluate if individuals that chose to participate in the social network present better improvement in their PAL (in comparison to those who opted out) in the following two scenarios: (1) people who are connected in the network from day 1 of the program, and (2) people who are willing to be in the community program. This work gives attention to one of the methodological issues when selecting data.

Chapter 6 is a Social Network Analysis (SNA) of the data set from Chapter 5. In this work we are interested in understanding how the dynamics of the connections happen over time using SNA tools, like centrality measurements and correlations.

1.5.3 Part III: Using Social Contagion Models for Explaining Physical Activity

This part is dedicated to propose ways of tuning and validating the social contagion model in different contexts. By validating we mean that we try to use the social contagion model to explain the data collected in two different sets.

Chapter 7 will use the same data set from Chapters 5 and 6. In this chapter we want to fit the data about the PALs of the individuals and their connections to our social contagion model shown in Chapter 2. Chapter 7 also presents all the decisions regarding the cleaning of the data and the process of generating the simulations together with the results obtained.

Chapter 8 applies the same strategy using an alternative model to explain the changes in PAL in a network of children from schools in the Netherlands. The model used here is also based on social contagion, but with a different mathematical framework from our social contagion model.

1.5.4 Part IV: Use of Contagion Models for Perceptions and Emotions

This part contains other works in contexts other than PAL that can also rely on social contagion phenomena. The main goal of this part is to show that the approach applied for the PAL is expandable to many possible investigations, including neurological and cognitive explorations.

Chapter 9 presents a temporal-causal model for the spread of messages in disasters. It is based on cognitive knowledge about responses from the brain and human interactions when in a context of the spread of information about disasters.

Chapter 10 presents a cognitive model for political positioning of people when interacting on social media websites.

Chapter 11 presents the construction of a classifier of messages for the political model presented in Chapter 10. A machine learning method is proposed to classify social media messages (tweets) between political or non-political.

1.5.5 Part V: Discussion and evaluation

Part V of this thesis aims to discuss the contributions of this work and explain how the research questions presented in Chapter 1 were answered throughout the chapters in this thesis. This discussion and evaluation is present in **Chapter 12**.

Bibliography

[1] Kristin Akerjordet and Elisabeth Severinsson. „Emotional intelligence: a review of the literature with specific focus on empirical and epistemological perspectives“. In: *Journal of clinical nursing* 16.8 (2007), pp. 1405–1416 (cit. on p. 12).

[2] Cheryl Alexander, Marina Piazza, Debra Mekos, and Thomas Valente. „Peers, schools, and adolescent cigarette smoking“. In: *Journal of adolescent health* 29.1 (2001), pp. 22–30 (cit. on p. 10).

[3] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008 (cit. on p. 8).

[4] Peter S Bearman and James Moody. „Suicide and friendships among American adolescents“. In: *American journal of public health* 94.1 (2004), pp. 89–95 (cit. on p. 10).

[5] Nino Boccara. *Modeling complex systems*. Springer Science & Business Media, 2010 (cit. on p. 8).

[6] Jeremy Boissevain. *Friends of friends: Networks, manipulators and coalitions*. St. Martin's Press, 1974 (cit. on p. 10).

[7] Stephen P Borgatti, Martin G Everett, and Jeffrey C Johnson. *Analyzing social networks*. Sage, 2018 (cit. on p. 14).

[8] Tibor Bosse, Rob Duell, Zulfiqar Ali Memon, Jan Treur, and C Natalie Van Der Wal. „Multi-Agent Model For Mutual Absorption Of Emotions.“ In: *ECMS* 2009 (2009), pp. 212–218 (cit. on p. 7).

[9] Tibor Bosse, Rob Duell, Zulfiqar Memon, Jan Treur, and C van der Wal. „A multi-agent model for emotion contagion spirals integrated within a supporting ambient agent model“. In: *Principles of practice in multi-agent systems* (2009), pp. 48–67 (cit. on pp. 6, 15).

[10] Tibor Bosse, Rob Duell, Zulfiqar Memon, Jan Treur, and C van der Wal. „A multi-agent model for emotion contagion spirals integrated within a supporting ambient agent model“. In: *Principles of practice in multi-agent systems* (2009) (cit. on p. 7).

[11] Ronald S Burt. „Autonomy in a social topology“. In: *American Journal of Sociology* 85.4 (1980), pp. 892–925 (cit. on p. 10).

[12] Duhigg Charles. *The Power of Habit*. Random House Publishing Group, 2012 (cit. on p. 5).

[13] Nicholas A Christakis and James H Fowler. „The collective dynamics of smoking in a large social network". In: *New England journal of medicine* 358.21 (2008), pp. 2249–2258 (cit. on pp. 6, 10).

[14] Nicholas A Christakis and James H Fowler. „The spread of obesity in a large social network over 32 years". In: *New England journal of medicine* 357.4 (2007), pp. 370–379 (cit. on pp. 6, 10).

[15] Ethan Cohen-Cole and Jason M Fletcher. „Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic". In: *Journal of health economics* 27.5 (2008), pp. 1382–1387 (cit. on pp. 3, 10).

[16] Nick Crossley, Elisa Bellotti, Gemma Edwards, et al. *Social network analysis for ego-nets: Social network analysis for actor-centred networks*. Sage, 2015 (cit. on pp. 10, 14).

[17] Alain Degenne and Michel Forsé. *Introducing social networks*. Sage, 1999 (cit. on p. 10).

[18] Joshua M Epstein. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, 2006 (cit. on p. 7).

[19] Joshua M Epstein. „Modelling to contain pandemics". In: *Nature* 460.7256 (2009), p. 687 (cit. on p. 7).

[20] Joshua M Epstein. „Why model?" In: *Journal of Artificial Societies and Social Simulation* 11.4 (2008), p. 12 (cit. on p. 7).

[21] James H Fowler and Nicholas A Christakis. „Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study". In: *Bmj* 337 (2008), a2338 (cit. on p. 6).

[22] Bruno Gonçalves and Nicola Perra. *Social phenomena: From data analysis to models*. Springer, 2015 (cit. on p. 13).

[23] Leo A Goodman. „Snowball sampling". In: *The annals of mathematical statistics* (1961), pp. 148–170 (cit. on p. 14).

[24] Dirk Helbing. „Agent-based modeling". In: *Social self-organization*. Springer, 2012, pp. 25–70 (cit. on p. 7).

[25] Marina Hennig, Ulrik Brandes, Jürgen Pfeffer, and Ines Mergel. *Studying social networks: A guide to empirical research*. Campus Verlag, 2012 (cit. on pp. 10, 14).

[26] Marco Iacoboni. *Mirroring people: The new science of how we connect with others*. Farrar, Straus and Giroux, 2009 (cit. on pp. 3, 6).

[27] Marco Iacoboni and John C Mazziotta. „Mirror neuron system: basic findings and clinical applications". In: *Annals of neurology* 62.3 (2007), pp. 213–218 (cit. on p. 3).

[28] Raghuram Iyengar, Christophe Van den Bulte, and Thomas W Valente. „Opinion leadership and social contagion in new product diffusion". In: *Marketing Science* 30.2 (2011), pp. 195–212 (cit. on p. 6).

[29] Rosenquist J, Murabito J, Fowler JH, and Christakis NA. „The spread of alcohol consumption behavior in a large social network". In: *Annals of Internal Medicine* 152.7 (2010), pp. 426–433. eprint: `/data/journals/aim/20203/0000605-201004060-00007.pdf` (cit. on p. 6).

[30] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, et al. „Jupyter Notebooks-a publishing format for reproducible computational workflows.“ In: *ELPUB*. 2016, pp. 87–90 (cit. on p. 14).

[31] Gueorgi Kossinets. „Effects of missing data in social networks“. In: *Social networks* 28.3 (2006), pp. 247–268 (cit. on p. 5).

[32] Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017 (cit. on p. 8).

[33] Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. „How social influence can undermine the wisdom of crowd effect“. In: *Proceedings of the National Academy of Sciences* 108.22 (2011), pp. 9020–9025 (cit. on p. 7).

[34] Rose McDermott, James H Fowler, and Nicholas A Christakis. „Breaking up is hard to do, unless everyone else is doing it too: Social network effects on divorce in a longitudinal sample“. In: *Social Forces* 92.2 (2013), pp. 491–519 (cit. on p. 6).

[35] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012 (cit. on p. 14).

[36] Miller McPherson, Lynn Smith-Lovin, and James M Cook. „Birds of a feather: Homophily in social networks“. In: *Annual review of sociology* 27.1 (2001), pp. 415–444 (cit. on p. 10).

[37] Martina Morris. *Network epidemiology: A handbook for survey design and data collection*. Oxford University Press on Demand, 2004 (cit. on p. 10).

[38] Julianna Pacheco. „The social contagion model: Exploring the role of public opinion on the diffusion of antismoking legislation across the American states“. In: *The Journal of Politics* 74.1 (2012), pp. 187–202 (cit. on p. 3).

[39] Mario Paolucci. „N. Gilbert and KG Troitzch, Simulation for the Social Scientist“. In: *Journal of Management & Governance* 12.2 (2008), pp. 225–231 (cit. on p. 7).

[40] Jon Parker and Joshua M Epstein. „A distributed platform for global-scale agent-based models of disease transmission“. In: *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 22.1 (2011), p. 2 (cit. on p. 7).

[41] Bernice A Pescosolido and Sharon Georgianna. „Durkheim, suicide, and religion: Toward a network theory of suicide“. In: *American Sociological Review* (1989), pp. 33–48 (cit. on p. 10).

[42] J Niels Rosenquist, James H Fowler, and Nicholas A Christakis. „Social network determinants of depression“. In: *Molecular psychiatry* 16.3 (2011), p. 273 (cit. on p. 6).

[43] Richard B Rothenberg, John J Potterat, Donald E Woodhouse, et al. „Social network dynamics and HIV transmission“. In: *Aids* 12.12 (1998), pp. 1529–1536 (cit. on p. 10).

[44] Clifford W Scherer and Hichang Cho. „A social network contagion theory of risk perception“. In: *Risk analysis* 23.2 (2003), pp. 261–267 (cit. on p. 6).

[45] Hermann Schichl. „Models and the History of Modeling“. In: *Modeling Languages in Mathematical Optimization*. Ed. by Josef Kallrath. Boston, MA: Springer US, 2004, pp. 25–36 (cit. on p. 7).

[46] Pepijn Schreinemachers and Thomas Berger. „An agent-based simulation model of human–environment interactions in agricultural systems“. In: *Environmental Modelling & Software* 26.7 (2011), pp. 845–859 (cit. on p. 7).

[47] Frank Schweitzer. „Brownian agents and active particles. On the emergence of complex behavior in the natural and social sciences“. In: *Springer Series in Synergetics. Berlin, Germany, Springer* (2003) (cit. on p. 7).

[48] Frank Schweitzer and David Garcia. „An agent-based model of collective emotions in online communities“. In: *The European Physical Journal B* 77.4 (2010), pp. 533–545 (cit. on p. 7).

[49] John Scott. *Social network analysis*. Sage, 2017 (cit. on p. 14).

[50] Yilun Shang. „An agent based model for opinion dynamics with random confidence threshold“. In: *Communications in Nonlinear Science and Numerical Simulation* 19.10 (2014), pp. 3766–3777 (cit. on p. 7).

[51] Robert Siegfried. *Modeling and simulation of complex systems: A framework for efficient agent-based modeling and simulation*. Springer, 2014 (cit. on p. 8).

[52] Jeffrey A Smith, Miller McPherson, and Lynn Smith-Lovin. „Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004“. In: *American Sociological Review* 79.3 (2014), pp. 432–456 (cit. on p. 10).

[53] J Treur. „The Ins and Outs of Network-Oriented Modeling: from Biological Networks and Mental Networks to Social Networks and Beyond (Keynote Lecture)“. In: (2018) (cit. on p. 9).

[54] Jan Treur. „Network-Oriented Modeling and Its Conceptual Foundations“. In: *Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions*. Cham: Springer International Publishing, 2016, pp. 3–33 (cit. on pp. 8, 9).

[55] Thomas W Valente. *Social networks and health: Models, methods, and applications*. Vol. 1. Oxford University Press New York, 2010 (cit. on p. 10).

[56] Thomas W Valente, Peggy Gallaher, and Michele Mouttapa. „Using social networks to understand and prevent substance use: a transdisciplinary perspective“. In: *Substance use & misuse* 39.10-12 (2004), pp. 1685–1712 (cit. on p. 10).

[57] Thomas W Valente, Anamara Ritt-Olson, Alan Stacy, et al. „Peer acceleration: effects of a social network tailored substance abuse prevention program among high-risk adolescents“. In: *Addiction* 102.11 (2007), pp. 1804–1815 (cit. on p. 10).