Part V

Discussion and Evaluation

It was the Unicorn who summed up what everyone was feeling. He stamped his right fore-hoof on the ground and neighed, and then cried:

"I have come home at last! This is my real country! I belong here. This is the land I have been looking for all my life, though I never knew it till now. The reason why we loved the old Narnia is that it sometimes looked a little like this. Bree-hee-hee! Come further up, come further in!"

C.S. Lewis, *The Last Battle (1956)*[1]

12 Discussions and Conclusion

The use cases discussed in this thesis show the relevance of studying human connections, as technology now enables humans to be connected as never before. This fact leads us to consider what the consequences of such an integrated world are, where it is possible to know what almost anyone is thinking, doing, eating, watching, listening to, etc. The sharing of all this personal information and how it shapes other people's behaviours, perceptions and emotions was the main motivation to explore models of social contagion provided throughout this thesis. Our main research question is:

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

To answer this question, we explored many different scenarios in which the spread of perceptions, emotions and physical activity behaviour is happening inside of a network structure. In this section, we summarize the results of this work and discuss what can be done going forward. Section 12.1 contains the discussion of the work presented in the previous chapters. Section 12.3 presents a debate surrounding the limitations of this work and the nature of the methods used. Section 12.4 presents potential future work to follow up on with the results provided by this research. Lastly, Section 12.5 highlights the contributions of the author of this thesis in each of the chapters presented previously.

12.1 Discussion of the contributions of this thesis

The research presented here contains several contributions to the state of the art on the computational modeling of behaviours, perceptions and emotions spread in social networks. The results are relevant as a methodological source for other studies that aim to model how people can change based on their relations and interactions. We aim to provide all the methodological details necessary to replicate and extend the models presented in this thesis. Our first contribution is an improved model for social contagion based on differential equations systems, detailed in Chapter 2. Throughout the thesis we explore many ways of applying this model to different contexts. The major part of this thesis is dedicated to the spread of physical activity behaviour. We used data sets from various groups of people randing from children to adults and from a small data set of 25 participants to a much larger data set of 5,000 people. Here we sought to explore and shed light on the dynamics of the changes in the physical activity levels of the participants. These explorations can be seen in Chapters 4, 5, 7 and 8.

Besides the analysis of the data sets, other problems were tackled to explain the social contagion effect based on the data collected, and to understand the effect of applying interventions to improve the physical activity levels of people. To find good candidates for interventions in a social network, we explored the dynamics of the social network being built over time in a health promotion program, as seen in Chapter 6. We also presented different methods to find intervention targets in a network of children in Dutch schools based on either their centrality in the network, or on their personal characteristics. This is seen in Chapter 8.

In Chapter 3 we proposed a method to find the traits of a group of young adults in a physical activity behaviour experiment. We try to understand how the relations in this group of people affect their behaviours over time. For that, it is important to find the traits of the participants that contribute to the social contagion effect, i.e. openness and expressiveness. The method provides an alternative to the limitations of self-reports which are a very biased source of data about people, as well as a pure machine learning approach that can be very impersonal and unconnected to the knowledge provided by real life information.

The model presented in the first part of the thesis can also be extended to different applications beyond healthy lifestyle and physical exercise. This is shown in Chapter 9 where we used the same method applied to a context of the spread of messages in disaster situations. A cognitive model was built based on a temporal-causal approach to account for the responses of people to receiving warning messages regarding potential disasters as they happen. Another model using this very similar approach was built in Chapter 10 in order to understand and predict people's reactions when interacting on web media with political posts. This work is based on a multidisciplinary investigation that embodies neuroscience and social psychology to validate the models.

Lastly in Chapter 11, this thesis presented a classification mechanism using machine learning techniques for political messages from Twitter (or Tweets). This classification is relevant, as the models proposed required that the context and the inputs for the individual are quantified. In this work, we proposed a mixed method for the classification of the Tweets where many methods are applied and a "jury" defines if the content is related to politics or not. The high accuracy ($\sim 97\%$) of this method is remarkable and provides a consistent way of quantifying the political content spread in the Twitter platform.

Other attempts to simulate social contagion can be found in related literature. [18] presents a review in different models of people involved in a dynamic of opinions exchange. For the models presented in the work of Hegselmann, Krause, et al., the sum of all the connections between the agents ($a_{ij} \geq 0$) is 1 due to the mathematical framework of the models. The models presented in this thesis do not have the same structure. Even though it would be possible to change the scale by some normalization of the original weights for the edges, the representation wouldn't be realistic towards the original purpose of the work. Hegselmann, Krause, et al.'s models also permit that loops exist. That means that an agent i can disregard everyone's opinions, meaning that $a_{ii} = 1$ and $a_{ij} = 0$ for $j \neq i$. The result of [18] is a stochastic matrix for the connections, with all the values of the rows summing up to 1. In our work, we use deterministic equations to address the changes in the state of the agents and in the edges' weights. [18] reviewed many works with stochastic models [10, 21, 13, 14] and analyses a time-variant for situations "where agents put in the course of time more and more weight on their own opinion and less weight on the opinion of others". The only resemblance with some of the works listed by Hegselmann, Krause, et al. is developed by [20], where bounded confidence is used to define to whom the agents are going to consider the opinions. Some similarities with this approach can be seen in Chapter 8's model based on the works of [2] and [15] , where the threshold is used to define to what extent the behaviour of the agent is affected and prone to change. The same strategy can be found in [9, 23], all of them related to opinion adjustments, and not to behaviour, as proposed in part of this thesis. The temporal-causal model with differential equations presented in Chapter 2 and used throughout this thesis includes the time scale as a factor of the model together with a speed factor, not presented in any of the works explored in [18]. The nonlinear equations used are also deterministic, to which specific mathematical analysis are required. This work explores new and more advanced ways of addressing other problems with some similarities with the group opinion formation, but respecting the complexity of phenomena not considered as the target applications of the analysis performed by [18].

Flache et al. [12] also review three different classes of models for social influence, proposing new frontiers for future research. The models of this thesis are more related to the models of *assimilative social influence*, where individuals always influence each other towards reducing opinion differences. Chapter 8 presents a new approach by including a threshold to the influence received that would be closer to the class of models with *similarity biased influence*.

[12] claim that "more empirical work is needed testing and underpinning micro-level assumptions about social influence as well as macro-level predictions". This thesis presents many different ways of validation of a contagion model through empirical data in an integration of theory and reality, an empirically-based computational approach for social contagion model. As highlighted above, the model presented in Chapter 2 is distinguished from other classic models studied in previous decades, and proper mathematical and simulation analysis are provided to understand the characteristics of this temporal-causal model. [12] present three patterns for models of social influence with different convergence characteristics for the opinions of the agents, namely consensus formation, clustering and bi-polarization. The model presented in Chapter 2 could generate any of the three classes, depending on the context and on the dynamics of the network structure.

The model for political positioning presented in Chapter 10 and the model for disasters reactions presented in Chapter 9 are not validated in real data, but based on literature. As Flache et al. [12] state, referring to other attempts to model social influence, "in many contributions authors derive the theoretical assumptions they make both from fundamental psychological theories about social influence and from empirical evidence, thus 'calibrating' models in a broad sense". The political positioning model is an attempt to summarize the knowledge on political psychology and political neuroscience in order to fill some of the gaps on the understanding of political positioning shaping. This is a new approach for ABMs that could help to answer Robert Axelrod's question "if people tend to become more alike in their beliefs, attitudes, and behavior when they interact, why do not all such differences eventually disappear?" [1] by bringing the complexities of the human brain into consideration when evaluating the mechanisms of perceptions (or beliefs) formation.

Centola [8] data set has some similarities with the data presented in Chapters 5, 6 and 7. Even though Centola [8] gathers data of people in a health promotion network, there are strong differences between what was done by Centola [8] and this work. Firstly, [8] is interested in comparing two different network topologies in order to verify which one spreads the health-related behaviour farther. The participants of the experiments were designed to be part of one topology or another (a clustered lattice network or a random network). The topologies were fixed before the participant joined the network, and the number of neighbours was the same across conditions. Participants couldn't see the identity of their peers. Differently, this thesis does not take into account the network topology as a factor in the behaviour spread, even though the topology can be considered relevant to fully understand the social contagion phenomenon. The data set used in this thesis contains two populations and permitted the participants to connect with others without restrictions. Therefore, the resulted network contains all the connections desired by the people in it without limitations. For this reason, Chapter 6 explores the formation of the network in order to verify what type of network is created when it is built from the beginning.

Most of the effort presented in this work is towards generating the network of people in real life rather than defining how the network should be set in order to better perform the spread of behaviour. Another consistent difference is in the fact that the health-related behaviour adopted by the participants in [8] is the adoption of a certain program and not an objective measure of their physical activity as done in this thesis.

Centola [7] presents a similar set up for the network system to the one explained above in [8], but aims to verify if connecting people with similar characteristics (creating homophilic ties) would increase or decrease the chances of obese and non-obese people to adopt a diet agenda after being exposed to other peers adopting it. The network structure was fixed, and the individuals were distributed in two groups: a homophilious group where the characteristics of the individuals (BMI, age, sex, etc.) were taken into account to populate the network, and, a random group where the individual attributes were not considered to place the nodes in network. [7] showed that homophily significantly increased adoption both among the obese and non-obese members of the community. The results presented by [7] are really relevant to strengthen the concept of homophily in social relations and spread of behaviour. This thesis presented some homophily analysis in a network built for a health promotion program in Chapter 6, but as stated above, the network wasn't fixed. Besides, the aim of the work wasn't to compare how far the diffusion of one specific program adoption goes. The participants of the data set presented in Chapter 6 had the option to opt in or out the network system alone. The results found in this thesis account for the quantified spread of behaviour using the PAL of the individuals over time. Therefore, instead of a binary condition (adopt or not adopt a diet agenda, for example), this thesis presented a concrete work on studying the quantification of the behaviour spread, a new exploration in the vast literature investigated in this field.

12.2 Research questions

The aim of this thesis is to understand *how we can create and validate computational models that explain social influence and contagion in social networks*. To achieve this goal, 3 subquestions were raised to guide our investigation.

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?

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?

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 following subsections aim to explain how the research questions connect with the work presented throughout this thesis.

12.2.1 Research question 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?

We have shown that temporal-causal models are very suitable for modeling social contagion in networks. The spread of any sort of behaviour, perception or emotion should have a temporal dimension to account for the dynamics of the change. That means, it takes time for someone to adapt to new perceptions, or to be affected by an external agent's emotions. The causal effects of the interactions between the nodes is also relevant, as the contagion is only possible between agents connected to each other and able to expose themselves reciprocally. Our proposed model is in line with other social contagion studies that show that "the empirical research has tended to confirm that the hypothesis that human behaviour clusters in both space and time even in the absence of coercion and rationale" [22].

For a model to be useful, it is essential that it provides a correct representation of the process in reality. Finding a mathematical representation for social contagion is a challenge, as there is no guarantee that the structure chosen to explain a certain event is the most suitable one. To provide an accurate representation of reality, effort is required to validate it. This can be done by showing that empirical data can be explained by the proposed model. Differential equations are used in many other investigations with the intent to provide time-dependent dynamics of some phenomenon [11, 19]. In Chapter 2 we presented our model to account for the social contagion of behaviour. For this we have adapted a previous model developed by [4] and provided evidence that the model was stable enough to explain contagion of behaviours, perceptions and emotions on a bigger scale than the previous model.

This was a good improvement, mainly because it removed the instability of the changes in the states of the individuals in the simulation of bigger groups of nodes. The new model provides more realistic results as the changes in the states are not abrupt or out of the limits defined for the scenarios.

We explored many methods to model the network connections of the individuals in our data sets. To create a realistic model that accounts for the changes caused by the relationships of people, it is also relevant to consider how the connections between people are determined. The relationships of a person are the way through which social contagion happens. It is through the interaction with friends, family and other people that behaviour, perceptions and emotions are changed. We proposed a few methods to describe and quantify the relationship of groups of people in different scenarios in order to fill the model with a reliable estimate of how strong the ties are between the individuals in different data sets. In Chapter 4, we applied some questionnaires to a group of young adults to define what the network structure was during the experiment. The task was to define how strong the connections are, and who influences whom in an oriented network. That is, two peers can have different levels of relationship, and therefore levels of influence, when looking at how the first affects the second, and vice-versa. In Chapter 5 however, the edges of the network structure are not oriented. The social network structure provided by the health promotion program did not permit that we assess the level of relationship between the participants in detail, providing only the friendship requests and the acceptance of them. For this data set the edges were dynamic. That is, they were generated over time, as the program was being run. Because of the dynamics of the connections, we proposed a method where a network is generated per each day of the experiment, so the changes in the network structure can be reflected in the spread of contagion and therefore in the model proposed. To better understand how the structure of the data set used in Chapter 5 is changed, we investigated the social network characteristics over time in Chapter 6. In Chapter 6 we presented the dynamics of the connections in this network, the number of edges increasing over time and the changes in the position of nodes, as those are factors that can affect the spread of behaviour and potential predictions using the contagion model. All the studies on the network connections and structure dynamic are a contribution to the building of reliable and consistent models. The network structure is extremely relevant for the validity of the results of the simulations using the models proposed in this thesis. The design of realistic network edges will certainly provide good understanding and more accurate predictions for the state of the network.

In this thesis we have shown a few methods to gather information about personal traits and how to quantify them. Besides knowing how the network structure of the people is, it is also relevant to know two specific personality traits of the individuals: openness and expressiveness. These traits are the indicators of the potential change in a person (openness) and of the level of extroversion and articulateness (expressiveness), and are important to determine the amount of influence received and sent to the neighbors of an individual. Therefore, finding these two traits is also important to make the proposed contagion model more accurate. In Chapter 3 we proposed a new method to combine the use of a Big Five questionnaire applied to a group of young adults with some optimized algorithms to better define the personal characteristics in the context of our contagion model for spread of behaviour. These methods try to improve the results obtained and shown in Chapter 4, where the selfreported questionnaires were used alone to define the openness and expressiveness of the participants.

We have shown that we can use dynamic models to analyze which network questions are more relevant to ask in order to obtain a realistic representation of the social connections of individuals. In Chapter 8 we used self-reports from children in the schools from The Netherlands to build the network structure. The children were asked in many ways how they perceive their classmates. 16 questions in total were used to understand who the popular children are, who dresses well, who is a good person to talk about food, who the friends of the participants are, etc. From these nomination questions we came out with three different oriented networks based on different sets of questions. We were interested in learning if raising less questions to the children would provide us similar networks. This was an attempt to diminish an overload on the participants by asking less questions and keep the confidence about their connections built from the data. Also in Chapter 8 we used a different mathematical model to account for the change in the behaviour of the children. This mathematical model turned out to be very sensitive to the threshold parameters, which resulted in necessary adaptations to keep the results realistic. The original model for this work is based on the research described in [2, 16].

Finally, this thesis presented a body of work on how to model a social network for social contagion using different mathematical methods. Some of the problems encountered when modeling the spread of behaviour are studied in Chapters 2 , 4, 5, 6, 7 and 8.

12.2.2 Research question 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 Physical Activity Level (PAL)?

In the previous research question we have shown that this thesis provides methods and ways to model a contagion model that accounts for the spread of behaviour, perceptions and emotions. The more realistic a model becomes, the more it becomes a reliable measurement for predictions and interventions.

We have shown in Chapter 4 that our social contagion model can predict around 80% of the cases if the PAL of an individual is going to increase or decrease based on the information provided for the network structure and the daily PAL collected for a period of one month. The high accuracy for the contagion model in this scenario implies that the social spread of behaviour is real and can be modeled using the model proposed. The results of this chapter also provided the information on the amount of physical activity behaviour spread in a social network of young adults.

Chapter 7 shows that the model presented in Chapter 2 performs better at describing the pattern seen in a data set of approximately 2,400 people than a simple linear model. We have shown in this chapter that some of the dynamics of the PALs in the network can be explained by social contagion processes. First we provided a simple linear model that has been derived by a regression analysis. Then we compared with the social contagion model mixed with a steady increase in the PAL due to the community effect (see Chapter 5) and showed that the model outperforms the simple linear model. As far as we know, this is the first analysis of the ability of a computational model of social contagion to capture the pattern of physical activity levels in a community over time.

To predict changes in the behaviour of a group of children in Dutch schools, we presented a study on the many scenarios where interventions are applied to the participants. We have shown that selecting individuals based on their socio-economic situation to increase their PALs is comparable to using social network optimization algorithms in the data set collected by the MyMovez Project [3]. The results are shown in Chapter 8, which presents the exploration on the spread of behaviour in a group of schools in the Netherlands. The children were tracked using Fitbit devices and answered questions regarding their social relationships with their classmates. In this work we generated the initial network and state for each of the participants and applied 5 different interventions in a percentage of the class increasing the PAL of the individuals that were selected. This research demonstrates how we can use data to measure the social contagion and provide predictions according to the model used.

12.2.3 Research question 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?

Our hypothesis is that the use of temporal-causal network models can be extended to many other contexts and applications that involve people and the spread of behaviours, perceptions and emotions through social ties. As the scope of the model presented is not restricted to the physical activity, we dedicated a part of the thesis to show how to apply the same methods to the spread of messages in disaster situations (Chapter 9), and for cognitive modeling of political positioning changes while interacting with web media posts (Chapter 10).

Chapter 9 explores the spread of emotions in situations of disaster. For this, an Agent-Based Model was introduced accounting for the expressed emotion and potential actions of a person when they receive a message warning about a potential disaster scenario. Many simulations were performed to show that the model simulates what it is expected to happen when a tense and negative phone call or a happy and positive text message are received by the agent.

Chapter 10 presents a cognitive model which is also based on differential equations and temporal-causal relations, to explain how a political perception (or positioning) of a person is affected by exposure to web media tweets. This paper is based on research results from neuroscience and social psychology, and also considers that the social contagion of opinions plays an important role in the shaping of political positioning. Chapter 11 is a complementary work of Chapter 10, as it provides a machine learning classifier to categorize the tweets that would be used as inputs for our cognitive model.

Thus, we have shown that the model can be extended to other contexts where the social contagion is observed. Many other applications can be derived from the methods used in this thesis, incorporating the social contagion whenever it is relevant to explain the phenomenon studied. Section 12.4 of this current chapter will give good examples of potential future work following up on the methods and results obtained here.

12.3 Limitations of this work

The work presented in this thesis is relevant for the understanding of the social contagion processes and the modeling of social networks to explain and predict the spread of behaviours, perceptions and emotions. The task of creating realistic models is a very complex task that requires knowledge from the other fields. We dedicate this section to bring up the limitations of this work that require special attention.

When evaluating the spread of PAL, we were faced with many problems to define a fair and solid method to understand the results. To validate a contagion model is a difficult task that requires a good data set with sufficient participants, reliable information about personal characteristics and a complete longitudinal trace of the behaviour studied (in this case, the physical activity behaviour). The work presented in this thesis tried to address all these problems, but still faces the problems that the data sets don't provide all the necessary information. Most of the work here could be improved by collecting data for a longer time period. Unfortunately, to reach all the requirements of a perfect model would require more time and resources than was available.

In Chapter 4 we collected data over 30 days of a small group of people (25 young adults). One could claim that more data should be necessary to validate the results, or that a longer period of time should be taken to provide a better longitudinal perspective in the changes. Although these remarks are relevant and true, collecting behavioural data is not an easy task. For Chapter 4, the short period of time was relevant to reduce the seasonal effect on the data, but didn't avoid the lack of other potential explanations for the changes in the PAL of the participants, such as weather, injuries, holidays or other motives for people to exercise less (or more) than usual.

Although Chapter 4 presents a small data set, more explorations were done in Chapters 3, 5 and 7 using a data set of originally 5,000 participants over a time period of more than 300 days. In this data set, we had to acknowledge the fact that a high number of participants had dropped out or didn't record the data properly, causing missing data and sparseness of the data points. That means, even with a lot of data available, we struggled to build a reliable analysis. The conclusions were based on a very strict cleaning process to filter out any dirt from the data set (i.e. participants who dropped out, individuals that started the plan in other time windows, etc.), but it certainly limited our understanding of the whole population's behaviour. Therefore, we acknowledge that the data management and filtering were bottlenecks that could have an influence on our analysis and outcomes. Nevertheless, we worked with the aim to provide reliable explanations for all the decisions made and for all the steps taken to derive the final results.

The approach used to define personal characteristics in our models can be considered too generic. It would require additional work on how to define the personal traits (e.g. openness and expressiveness) to create a more reliable metric for these characteristics. In Chapter 4 for instance, we adapted the Big-5 questionnaire [17] to account for the openness and expressiveness of the participants. In Chapter 7 we set all the expressiveness and opennesses to 0.5 due to the lack of information about the personal traits. We believe that better results would be obtained if this data was provided through a very specific, well tested and validated process. Due to the difficulties on getting data related to the personal traits of individuals, we proposed a new method in Chapter 3. This method needs further investigation and more study using a different data set.

The applicability of our models should also be understood from the perspective of the context. Modeling behaviour change in groups is highly dependent on the context, i.e. the time of the year when data was collected, the level of details provided by the participants, the commitment of the participants to the data collection, the socio-economic characteristics of the population, etc. Therefore, it is important to be aware that our models and our validation studies are relevant in the context of the research we have done. Results cannot be generalized to any population, so for other populations it is important to verify if the characteristics of the individuals within the population, network and tools are similar to the ones used in our work.

Some of the explorations presented in this thesis do not have any data, due to the nature of the models. This is the case as seen in Chapters 9 and 10. Cognitive models are based on brain states that are not easily accessible by any existing tool so far. In these cases, we created data based on our assumptions and the expected functioning of the states modeled. It is important to acknowledge that as the state of the art in these fields become more advanced and provide better details about the phenomena studied, changes in the models may be needed to incorporate the new discoveries.

12.4 Future Work

This thesis presents an exploration on the modeling of social contagion in social networks within different contexts. This section describes potential future investigations based on the work presented in the previous chapters.

In Chapter 2, we presented a new proposal for the contagion model. In this proposal, the speed factor calculation is adapted to avoid unrealistic changes in the states of the nodes. Further investigations could explore how the new speed factor alternatives affect the results of previous research, and how they can be combined with the model for emotion contagion spirals from [5]. It would also be an interesting work to verify how this model can be validated in other contexts such as the perception and emotions' contagion.

The new method to define traits of people in a network for sharing their physical activity achievements presented in Chapter 3 could be improved in several ways. As the data set is quite small, we would have to test the method for finding the personality traits of people using optimization and machine learning algorithms within a bigger data set. This method can also be very useful for other applications, such as for the personal traits of users of web media, or for finding relevant traits of people in programs for behaviour changes (i.e. drug addiction, leadership training, etc.). Using the method proposed could help to divulge other results and therefore improve the understanding of the advantages and limitations of this method. Further applications using the method proposed could be also developed. For instance, an application for defining potentially depressed people in a network based on self reports and on machine learning techniques could be helpful identifying nodes that can be a support for people who are struggling with depression.

In Chapter 4 we collected data of the PAL of a group of young adults over a period of 30 days in order to understand the dynamics of the social contagion of behaviour in their network of relations. A longer and bigger set up would be required to provide stronger findings that would support the spread of the behaviour in the network. Other improvements could include considering the changes in the connections over time to account for the real changes in the relationships, and to try to apply interventions in some of the nodes to propagate a positive behaviour faster. The results presented are good indicators that the social contagion can explain changes in a group of people. Therefore, it can be expected that applications seeking to find good candidates for interventions in order to improve the lifestyle and health of a population can take advantage of the results obtained by this research. The social contagion model can be used to predict the increase/decrease of the PAL and provide inputs for a potential change in the structure of the network, or some interventions in the people's states.

Even though the data set shown in Chapters 5 and 7 provides a much bigger sample with more people and a longer experiment, many challenges are still to be tackled in future investigations. It is important to study the effect of other factors on the physical activity level, such as the community size and structure. That way, research can further uncover phenomena that are at the basis of the beneficial effects of online social networks in health promotion programs. For this, the results from Chapter 6 can be a good starting point. The study of the dynamics of the network creation provides insights to how the ties are formed and how the position of the nodes in the network are changed. The combination of degree measurements for the nodes and the density of the ego-network can be used to identify people who are potentially influential in their network in further work. The results show that continuously monitoring the evolution of a network is important to identify such people. Future work can use the outcomes of the network analysis to form the basis for automated (health) interventions that exploit the social network for changing behaviours of individuals. This could lead us to future discoveries about leadership, spread of emotions or any other application related to a network's topology and dynamics.

Besides modeling social contagion and the spread of behaviour, applying interventions in a group of people is also a task that requires deep investigation. In Chapter 8 we use an adapted model to select children from Dutch schools with the aim to improve the PAL of the whole group. Future work has to be carried out in order to verify if our proposed contagion model is a better predictor for the behavioural change in the data set than the currently used diffusion model. Other strategies can also be drawn to define better ways to select the intervention nodes in the network. The data set used in this work can be studied further and additional findings could better explain the dynamics of the network, or improve the mapping of the relationships questions to a network. The use of the method proposed in Chapter 3 could possibly also be used in this data set to find the personality traits of the children.

The results provided by the research in Chapter 8 can be used to define a good set of questions in order to generate the network of a classroom of children and teenagers from the same school. It also provides knowledge of what happens when some intervention is applied in the network. Therefore, more research can still be done to verify if the interventions caused the changes that were expected by the model proposed, and if so, how can these results be repeated in other groups (i.e. different social-economic situation contexts).

Additional work can also be carried out by finding other contexts to apply the temporal-causal modeling method. Chapters 9 and 10 proposed two cases for the use of contagion principles, and from these models other real applications can also be created. The model for the spread of messages in disaster situations could be extended to simulate a population in a situation of disaster or panic. The political cognitive model can also be extended such that it is suited for a network to investigate the spread of political opinions in social networks. Further validation using empirical data would also be useful to provide stronger evidence that the models are correct. The work presented in Chapter 11 is a starting point for the creation of real world applications using the model created in Chapter 10.

Understanding and modeling social contagion still has many unanswered questions and potential applications. The results obtained by this research are useful for future applications for health lifestyle promotion, understanding perceptions and modeling the spread of emotions involving new technologies that aim to use the personal traits and the connections of people to encourage them to be more active, eat healthier or both. These technologies could be extended to applications that account for loneliness, depression, drug addiction, or any other behaviour that can be considered shaped within a social context.

12.5 Contributions and chapters overview

This section presents the contributions of the author in each of the research works presented throughout the thesis and the chapters overview.

Chapter 2 has been published as: ***Fernandes de Mello Araújo E.***, Treur J. (2016) Analysis and Refinement of a Temporal-Causal Network Model for Absorption of Emotions. In: Nguyen NT., Iliadis L., Manolopoulos Y., Trawi´nski B. (eds) Computational Collective Intelligence. ICCCI 2016. Lecture Notes in Computer Science, vol 9875. Springer, Cham.

My contribution to this chapter includes the proposal of a new measurement for the speed factor in the mathematical equation that accounts for the aggregated impact of the other nodes. I also coded the simulated scenarios, plotted the graphics, discussed the results and wrote part of the manuscript in partnership with the other author. This paper presents an improvement for the emotion contagion model previously created by Bosse et al. [6] that is used to account for the spread of behaviour in many other chapters of this thesis.

Chapter 3 has been published as: ***Eric F. M. Araújo***, Bojan Simoski, and Michel Klein. 2018. Applying machine learning algorithms for deriving personality traits in social network. In Proceedings of ACM SAC Conference, Pau, France, April 9-13, 2018 (SAC'18).

My contribution to this chapter includes the proposal of the new method to define personality traits, the coding of the analysis and algorithms, the plotting of the graphics, the discussion of the results and writing of the manuscript. In this chapter we present a method that combines intake questionnaires with some optimization algorithm in order to better tune the traits of openness and expressiveness of a group of young adults in an experiment of sharing behaviour.

Chapter 4 has been published as: ***Eric F. M. Araújo***, Anita V. T. T. Tran, Julia S. Mollee, Michel C. A. Klein. Analysis and evaluation of social contagion of physical activity in a group of young adults. In Proceedings of the ASE BigData & SocialInformatics 2015. ACM, 2015.

My contribution to this chapter includes the proposal of the experiment, the supervision of the data collection, the configuration of devices to track the PAL of the participants, the coding and analysis of the results, the discussion of the results and the writing of the manuscript. This work aimed to use a social contagion model based on temporal-causal relations to predict the changes in the PAL of 25 young adults from the same course. The results showed that in more than 80% of the cases, the contagion model is a good predictor for the increase or decrease of the PAL based on the social relations and personality traits of the individuals in the data set.

Chapter 5 is an extension of the paper published as: A. Manzoor, J. S. Mollee, ***E. F. M. Araújo***, A. T. V. Halteren and M. C. A. Klein. Online Sharing of Physical Activity: Does It Accelerate the Impact of a Health Promotion Program?, 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, 2016, pp. 201-208.

My contribution to this chapter includes the data analysis, involving the understanding of the data set, the management of the data, the building of the networks and the filtering of nodes, the coding of the tasks to obtain the results, the discussion of the results and the writing of the manuscript. This chapter presents a statistical analysis of a data set of a health promotion program where the participants could opt to participate in a community for sharing their PALs over time with their friends (connections). We were interested in understanding what the effect of being part of a community is and what the effect of being connected to other participants in the same experiment is.

Chapter 6 has been published as: ***de Mello Araújo, E. F.***, Klein, M., van Halteren, A. (2016, November). Social Connection Dynamics in a Health Promotion Network. In International Workshop on Complex Networks and their Applications (pp. 773-784). Springer, Cham.

My contribution to this chapter includes the discussion of the research questions, data manipulation, coding of the analysis, discussion of the results and writing the manuscript. This work is a social network analysis of the dynamics of the edges in a data set of people connected in a health promotion program, where they could share their daily PAL with their connections. We look for leaders in the network and the dynamics of a network generation.

Chapter 7 has been published as: Mollee, J. S., ***Araújo, E. F.***, Manzoor, A., van Halteren, A. T., Klein, M. C. (2017, March). Explaining Changes in Physical Activity Through a Computational Model of Social Contagion. In Workshop on Complex Networks CompleNet (pp. 213-223). Springer, Cham.

My contribution to this chapter includes the discussion of the research methods, the coding of the simulations, the generation of the graphics and results, the discussion of the findings and the writing of the manuscript. This work tries to explain the changes in the PAL of a group of people in a health promotion program by using the social contagion model combined with the effect found in being a part of the community. The results showed that the model outperforms a simple linear model.

Chapter 8 has been published as: **Araújo, E.**, Simoski, B., Woudenberg, T., Bevelander, K., Smit, C., Buijs, L., Klein, M. and Buijzen, M. *Using Simulations for Exploring Interventions in Social Networks - Modeling Physical Activity Behaviour in Dutch School Classes.* In Proceedings of 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2018), pages 414-425. ISBN: 978-989-758-323-0.

My contribution to this chapter includes the discussion of the research questions, the definition of the methods, the coding of the simulations, the generation of the results, the discussion of the findings and the writing of the manuscript. This work aims to find the best intervention methods to increase the overall PAL of a group of children in a Dutch school class. This paper was nominated to the best student paper award of the conference.

Chapter 9 has been published as: ***Fernandes de Mello Araújo E.***, Franke A., Hosain R.W. (2017) A Temporal-Causal Model for Spread of Messages in Disasters. In: Nguyen N., Papadopoulos G., Jedrzejowicz P., Trawinski B., Vossen G. (eds) Computational Collective Intelligence. ICCCI 2017. Lecture Notes in Computer Science, vol 10449. Springer, Cham 2017.

My contribution to this chapter includes the discussion about the model, as well as the formulation of the model, the discussion of the results and the writing and review of the manuscript. This work presents a cognitive model for the reactions caused by messages and calls alerting a disaster scenario. The model is based on neuroscience and psychology findings and is modeled to predict people's behaviour within these contexts.

Chapter 10 has been published as: ***E. F. de Mello Araújo*** and M. Klein, "A computational cognitive model for political positioning and reactions in web media," 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Oxford, 2017, pp. 414-422.

My contribution to this chapter includes the literature review and the elaboration of the model, the coding and simulation of the model, the discussion of the findings and the writing of the manuscript. This work presents a cognitive model that represents the political positioning change of a person when interacting with Tweets from web media. The work is based on neuroscience and social psychology findings, and is a great contribution to the understanding of how we can map changes in people's positioning.

Chapter 11 has been published as: ***Fernandes de Mello Araújo E.*** and Ebbelaar D. (2018). *Detecting Dutch Political Tweets: A Classifier based on Voting System using Supervised Learning*. In Proceedings of the 10th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, pages 462-469.

My contribution to this chapter includes the conception of the method to be used, the supervision of the student that collected the data and performed the machine learning techniques, the discussion of findings and the review and writing of the manuscript. This work provides a very accurate classifier for the task of distinguishing political and non-political tweets.

The following article was published but not included in the thesis: Mollee J.S., ***Araújo E.F.M.***, Klein M.C.A. (2017) Exploring Parameter Tuning for Analysis and Optimization of a Computational Model. In: Benferhat S., Tabia K., Ali M. (eds) Advances in Artificial Intelligence: From Theory to Practice. IEA/AIE 2017. Lecture Notes in Computer Science, vol 10351. Springer, Cham.

Bibliography

[1] Robert Axelrod. „The dissemination of culture: A model with local convergence and global polarization“. In: *Journal of conflict resolution* 41.2 (1997), pp. 203–226 (cit. on p. 207).

[2] Rahmatollah Beheshti, Mehdi Jalalpour, and Thomas A. Glass. „Comparing methods of targeting obesity interventions in populations: An agent-based simulation“. In: *SSM - Population Health* 3 (2017), pp. 211–218 (cit. on pp. 207, 211).

[3] Kirsten E Bevelander, Crystal R Smit, Thabo J van Woudenberg, et al. „Youth's social network structures and peer influences: study protocol MyMovez project–Phase I“. In: *BMC public health* 18.1 (2018), p. 504 (cit. on p. 212).

[4] Tibor Bosse, Rob Duell, Zulfiqar A Memon, Jan Treur, and C Natalie van der Wal. „Agent-based modeling of emotion contagion in groups“. In: *Cognitive Computation* 7.1 (2015), pp. 111–136 (cit. on p. 209).

[5] 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 p. 214).

[6] 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. 217).

[7] Damon Centola. „An experimental study of homophily in the adoption of health behavior“. In: *Science* 334.6060 (2011), pp. 1269–1272 (cit. on p. 208).

[8] Damon Centola. „The spread of behavior in an online social network experiment“. In: *science* 329.5996 (2010), pp. 1194–1197 (cit. on p. 208).

[9] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. „Mixing beliefs among interacting agents“. In: *Advances in Complex Systems* 3.01n04 (2000), pp. 87–98 (cit. on p. 207).

[10] Morris H DeGroot. „Reaching a consensus“. In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121 (cit. on p. 206).

[11] Keisuke Ejima, Kazuyuki Aihara, and Hiroshi Nishiura. „Modeling the obesity epidemic: social contagion and its implications for control“. In: *Theoretical Biology and Medical Modelling* 10.1 (2013), p. 17 (cit. on p. 209).

[12] Andreas Flache, Michael Mäs, Thomas Feliciani, et al. „Models of Social Influence: Towards the Next Frontiers." In: *Journal of Artificial Societies & Social Simulation* 20.4 (2017) (cit. on p. 207).

[13] Noah E Friedkin and Eugene C Johnsen. „Social influence and opinions". In: *Journal of Mathematical Sociology* 15.3-4 (1990), pp. 193–206 (cit. on p. 206).

[14] NE Friendkin and EC Johnsen. „Social influence networks and opinion change". In: *Adv Group Proc* 16 (1999), pp. 1–29 (cit. on p. 206).

[15] Philippe J. Giabbanelli, Azadeh Alimadad, Vahid Dabbaghian, and Diane T. Finegood. „Modeling the influence of social networks and environment on energy balance and obesity". In: *Journal of Computational Science* 3.1 (2012), pp. 17–27 (cit. on p. 207).

[16] Philippe J. Giabbanelli, Azadeh Alimadad, Vahid Dabbaghian, and Diane T. Finegood. „Modeling the influence of social networks and environment on energy balance and obesity". In: *Journal of Computational Science* 3.1–2 (2012), pp. 17–27 (cit. on p. 211).

[17] Samuel D Gosling, Peter J Rentfrow, and William B Swann. „A very brief measure of the Big-Five personality domains". In: *Journal of Research in personality* 37.6 (2003), pp. 504–528 (cit. on p. 214).

[18] Rainer Hegselmann, Ulrich Krause, et al. „Opinion dynamics and bounded confidence models, analysis, and simulation". In: *Journal of artificial societies and social simulation* 5.3 (2002) (cit. on pp. 206, 207).

[19] Alison L Hill, David G Rand, Martin A Nowak, and Nicholas A Christakis. „Infectious disease modeling of social contagion in networks". In: *PLOS computational biology* 6.11 (2010), e1000968 (cit. on p. 209).

[20] Ulrich Krause. „A discrete nonlinear and non-autonomous model of consensus formation". In: *Communications in difference equations* 2000 (2000), pp. 227–236 (cit. on p. 207).

[21] Keith Lehrer. „Social consensus and rational agnoiology". In: *Synthese* 31.1 (1975), pp. 141–160 (cit. on p. 206).

[22] Paul Marsden. „Memetics and social contagion: Two sides of the same coin". In: *Journal of Memetics-Evolutionary Models of Information Transmission* 2.2 (1998) (cit. on p. 209).

[23] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal. „Interacting Agents and Continuous Opinions Dynamics". In: *Heterogenous Agents, Interactions and Economic Performance*. Ed. by Robin Cowan and Nicolas Jonard. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 225–242 (cit. on p. 207).