FACULTY OF ENGINEERING
AND ARCHITECTURE

How opinion leaders can manipulate opinion dynamics in hierarchical social networks

Xavier Claerhoudt
Student number: 01505080

Supervisors: Prof. dr. ir. Pieter Simoens, Dr. Yara Khaluf
Counsellor: Ilja Rausch

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Computer Science Engineering

Academic year 2019-2020

GHENT
UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

How opinion leaders can manipulate opinion dynamics in hierarchical social networks

Xavier Claerhoudt
Student number: 01505080

Supervisors: Prof. dr. ir. Pieter Simoens, Dr. Yara Khaluf
Counsellor: Ilja Rausch

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Computer Science Engineering

Academic year 2019-2020

GHENT
UNIVERSITY

Preface

Writing a thesis is like forming an opinion, it is rarely done by yourself. A special thank you goes out to Ilja Rausch as the supervisor for this thesis. Throughout the year he was always available for questions or feedback and him sharing his perspective during our regular meetings was an invaluable contribution to this study. I would also like to express my gratitude towards my promotors prof. dr. ir. Pieter Simoens and dr. Yara Khaluf for giving me the opportunity to shape this dissertation to my liking and make this truly a project of my own.

Furthermore I would like to thank my parents, who have supported me along the entire way. Their persistent trust in me has played a big role in me getting to where I am today and I hope they can see their efforts reflected in this work. Last but not least a thank you is in place for my friends and in particular Robin, Alec, Jens-Joris, Michiel and Ward. Because of the extraordinary circumstances with the coronavirus we were all forced to stay at home instead of working together in the library as usual. Nonetheless we all encouraged and followed up on each other, making this last semester more enjoyable in the process.

Xavier Claerhoudt

Permission of use

How opinion leaders can influence the opinion dynamics in hierarchical social networks

by

Xavier Claerhoudt

Academic year 2019-2020
Ghent University
Faculty of Engineering and Architecture

Promotors: prof. dr. ir. Pieter Simoens and dr. Yara Khaluf
Counsellors: Ilja Rausch

Abstract

In this dissertation, the DeGroot model, a long-standing and generally well-accepted model for opinion formation, is used to evaluate the influence of opinion leaders with respect to the opinion dynamics in social networks. To be able to identify these leaders, multiple measures from graph theory are compared with the results of an experiment that evaluates the influence of each node. Special attention is paid to how well these measures correlate with the results of this experiment in both hierarchical and non-hierarchical environments. By using the measure that showed the highest correlation with the experiment outcomes for leadership identification, a set of experiments is conducted to test out how well a combination of influential people can affect the opinion formation process of the rest of the network by agreeing on an initial opinion. The results show that opinion leaders have a non-negligible effect on the rest of the individuals in the network. These observations are further confirmed when running additional experiments in which extra assumptions on the opinion formation process are made, except for the case where nodes pay more attention to individuals that have a similar opinion to their own.

Keywords: Opinion formation, consensus, opinion leadership, group decision making, fusion rules

How opinion leaders can influence the opinion dynamics in hierarchical social networks

Xavier Claerhoudt

Supervisors: prof. dr. ir. Pieter Simoens, dr. Yara Khaluf,
Ilja Rausch

Abstract— **This paper researches the influence of opinion leaders on the opinion dynamics in social networks. Special attention is paid to the case where there is hierarchy present in these networks. To model the process of group opinion formation, different levels of abstractions are used. A study is done to investigate the identification of opinion leaders based on measures from graph theory. Afterwards the DeGroot model is used to study the effect of these opinion leaders on the rest of the network. This paper starts by introducing the current methods used in opinion dynamics modelling, followed by an overview of the setup that was used to simulate these models. After that, the results of several experiments are discussed and the case is made that opinion leaders can have a significant effect on the group opinion.**
Keywords—**Opinion formation, consensus, opinion leadership, group decision making, fusion rules**

I. Introduction

While the concept of public opinion is universally known, the processes at play behind the formation of this opinion are still not fully understood. Group opinion formation is usually studied using abstractions at the level of individuals. Conclusions are drawn by running simulations and observing their development and outcomes. In these simulations there is often no desired outcome.

This paper studies the effectiveness of opinion leaders in steering the opinion of the group towards one side of the argument. To identify these opinion leaders, multiple measures from graph theory are compared against simulation results. Afterwards, using the measure that showed the highest correlation with these simulations, groups of opinion leaders are selected. They are given a strong opinion with the goal to evaluate whether they have the capability to direct the group opinion. This is initially done using the standard version of a common model in opinion formation. Subsequently, the conclusions drawn from these evaluations are tested further when multiple small adaptations to this model are made.

II. Opinion formation

Opinion formation is a complex process that differs from person to person. Because of this, generalized abstractions are required to keep simulations tractable. The current models on opinion dynamics are usually composed of three main building blocks: the opinion expression format, the opinion dynamics environment and the fusion rules [1]. The combination of these three factors determines which interactions will take place and how they are handled.

In reality people generally express their opinions in the form of language. Modelling this would add a large level of complexity and is therefore usually not done [2]. Instead, numeric values [3] or combinations thereof [4] are used to represent opinions. To allow for nuanced opinions, this study uses a real-valued variable between -1 and 1 as an opinion value. People with a negative opinion are considered to be on one side of the argument and those with a positive value on the other side. A value of zero indicates that the person has no opinion on the matter at hand. The reason for using an interval instead of the complete range of real numbers is to prevent certain scenarios, where some individuals have such extreme opinions compared to all others that they completely determine the outcome of the experiment.

After having identified how a person expresses his opinion, the next step is to determine which people can interact with each other. Using a graph to model the opinion dynamics environment is the most logical option, as it makes translation between reality and the model straightforward [5]. In this graph structure the nodes represent individuals and the edges represent connections between two people. Based on the structure of popular online social networks, the decision was made to use a directed graph [6]. Studying the structure and properties of these networks is out of scope for this paper. Therefore different well-studied graphs, which are argued to model society well, were investigated as options. A common structure used as opinion dynamics environments are scale-free graphs [7]. The main characteristic of these networks is their power-law degree distribution. The problem with using them in this study is that there is no way to control their hierarchality during generation and changing the network afterwards could destroy the scale-free property. Therefore the decision was made to go with another option, being triadic graphs. These triadic graphs are constructed from a set of triangular patterns. The networks used in this study consist of a combination of feed-forward and feedback loops. These triangular connections are common in networks of communicating entities [8]. While scale-free graphs did not allow to control the hierarchality, triadic graphs do. This control is exercised by changing the fractions of feed-forward and feedback loops in the network. To verify this, the inherent directionality ξ was calculated for each network, which can be seen as a measure for the hierarchality of a network [9]. Figure 1 presents the result of this calculation. N_s gives an indication of the amount of feed-forward loops that are changed into feedback loops by switching connections in the network. The downward trend shows that changing these connections is a way to control the hierarchy present in the network.

The last element required for the opinion formation model is the fusion rule. This models how the opinions received by a node are used to form an updated opinion. Popular options for fusion rules are the DeGroot model [10], the bounded confidence model [11] and the voter model [12]. The voter model is based on random actions and primarily used in research where a discrete opinion expression format is adopted. The bounded confidence model makes the assumption that people only interact with others when they have similar opinions. The DeGroot model does not make this assumption and is heavily used in research with continuous opinion expression formats [13], for these reasons it is also used here. A node using the DeGroot model will average the opinions of its peers

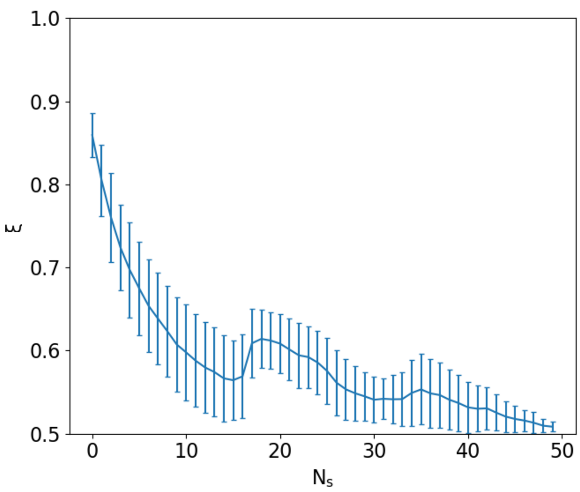


Fig. 1: Effect of switched connections on ξ

at each instant of time and use this average as its updated opinion.

III. Simulation setup

To gather data on the opinion formation process, simulations were done. Because of the fusion rules, agent-based simulations are the most suitable type of simulations. Therefore a simulator had to be chosen that allows using this paradigm. The two explored options were GAMA platform [14] and ARGoS simulator [15]. Since ARGoS simulator places a heavy focus on low-level implementations, so that the code can also be ran on actual robots, and GAMA platform focuses on high-level concepts instead, the decision was made to use GAMA platform.

While the exact simulation to perform may differ, the general setup to run and process it remains largely the same. Therefore Python scripts were used for the pre- and post-processing of simulations. These scripts generate the files and values required for the simulation at hand, run the simulation and process the resulting data afterwards.

IV. Results

Before looking into the influence of opinion leaders, they had to be identified first. In graph theory many measures have been proposed to rank the nodes of a graph. The measures that try to do this based on how well-connected the nodes are, are commonly referred to as centrality measures. A simulation was set up to evaluate the influence of each node in a network. It consists of giving all nodes an opinion value of 0, except for the selected node, which is given an opinion value of 1. This node is also made stubborn in the sense that it does not update its own opinion. Tracking the average final opinion of the network then allows a ranking of the nodes in the network. By checking the Pearson and Spearman correlation coefficients, ρ_p and ρ_s respectively, between these data points and the results of calculating the centrality measures, an indication of which measures are best suited to identify opinion leaders is obtained.

The measures that were tested are the closeness centrality [16], node level (obtained as one of the steps in the calculation of ξ), the PageRank score [17], the HITS score (Hyperlink-Induced Topic Search) [18] and the node out-degree. A sample of 22 networks was used to evaluate their performance. These networks were selected in such a way that different values of the inherent directionality ξ, their main differentiating characteristic, were present in the sample. The

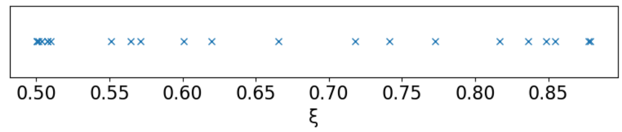


Fig. 2: Division of the selected networks across ξ values

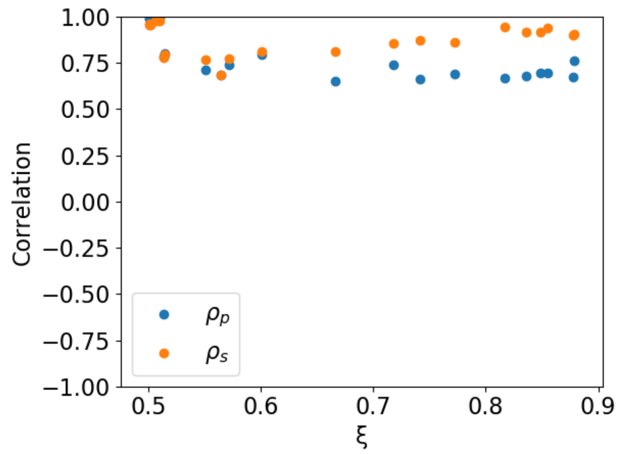


Fig. 3: Correlation between experiment results and node out-degree

division of these networks over the possible ξ values is presented in Figure 2. Evaluating the correlation showed that while most measures performed well for the networks with a low ξ value, their performance significantly degraded when the networks became more hierarchical. However, this was not the case for the HITS score and the node out-degree. Between these two, the measure that scored best when evaluated using the two correlation coefficients, was the node out-degree. The correlation values obtained for this measure are presented in Figure 3.

Having found a measure to identify opinion leaders in the network, the influence of these leaders on the opinion of the entire population could be investigated. Three different scenarios were used to evaluate this, using the original DeGroot model for the fusion rules. In each one, a group of opinion leaders is used to influence the other nodes in the network. The results of these scenarios are than averaged over all 5000 simulations performed, each using a different network. In the first scenario, all nodes were given an initial opinion value of 0, except the leader nodes which were given an opinion of 1. These leader nodes also updated their opinion according to the DeGroot model, similar to the non-leader nodes. Figure 4 presents the results of this experiment. It was conducted for multiple percentages of the nodes selected as leaders, as indicated in the legend. As can be seen from the figure, selecting even a small amount of nodes as leaders has a large impact on the average percentage of positive nodes present in the network.

The same experiment was then conducted again with one major difference, being that all other nodes now had uniformly distributed initial opinions. The orange dots in Figure 5 show the final percentage of positive nodes in this scenario, which was again tested using multiple percentages of nodes selected as opinion leaders. Compared to the scenario where all non-leader nodes had an opinion value of 0 (indicated in blue), a decrease in the average percentage of positive nodes is seen. However, this

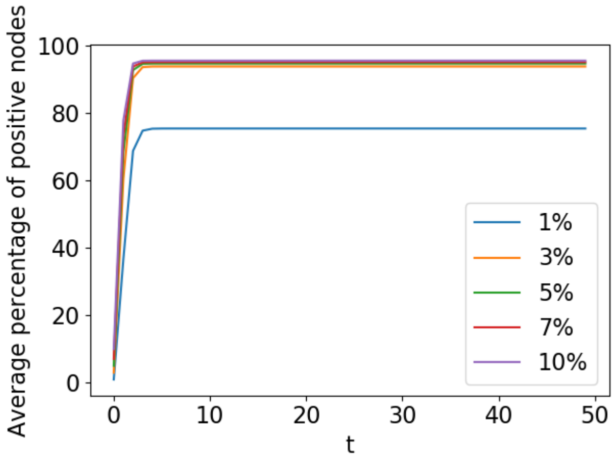


Fig. 4: Average fraction of positive nodes throughout the simulation

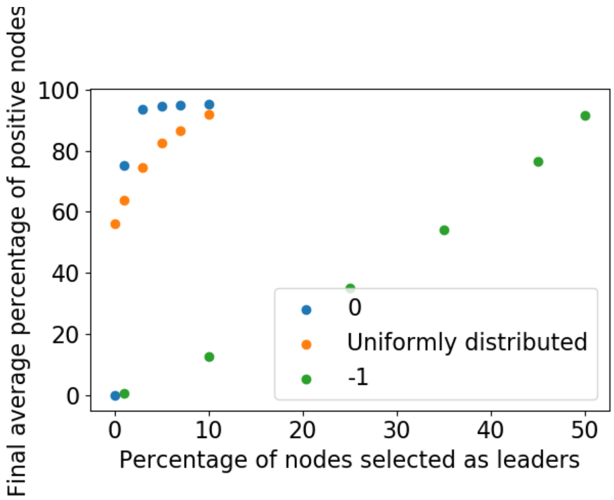


Fig. 5: Effect of number of opinion leaders

decrease becomes smaller when more leaders are added to the selection.

A third scenario where the original DeGroot model was used, is represented by the green dots in Figure 5. In this scenario all non-leader nodes had an opinion value of -1, thus spreading the exact opposite opinion of the leader nodes. This severely affected the ability of the leader nodes to have an impact on the rest of the network. Nonetheless, an increase in the average percentage of positive nodes is seen when increasing the number of opinion leaders, who have an initial opinion value of 1. The fact that carefully selecting 35% of the nodes to counteract the strong negative opinion of the other 65% works in more than 50% of the cases, proves that the opinion leader selection is successful. It also confirms the hypothesis that the opinion leaders can influence the other nodes in the network by agreeing on an initial opinion.

Based on these results, the opinion leaders have shown to be effective in influencing the opinion dynamics of the network when using the standard DeGroot model. However, the DeGroot model is a very general abstraction of the actual opinion formation process. To investigate the robustness of these results, the impact of several small changes to this model was evaluated. Adding self-confidence

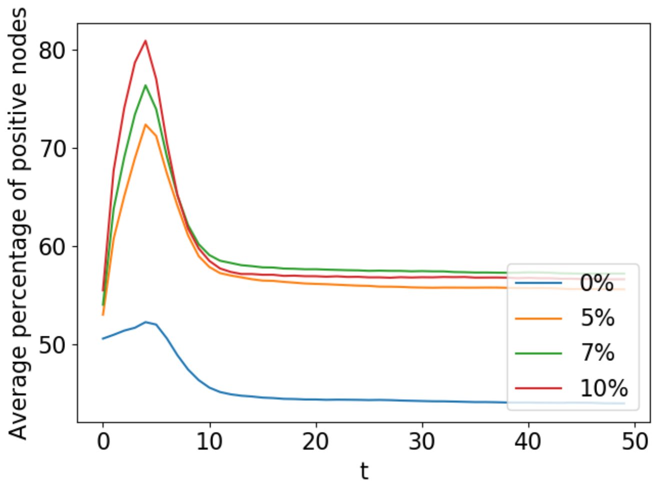


Fig. 6: Average fraction of positive nodes throughout the simulation

to the fusion rules, adding noise to the perceived opinions, placing more weight on nodes higher in the hierarchy or with a higher out-degree and placing more weight on nodes with similar opinions are all scenarios that were simulated. All but one of these scenarios showed similar results and trends when compared to the DeGroot model without the adaptation. This confirmed that the selected leader nodes have an effect on the group opinion formation process. The only case where the opinion leaders were unsuccessful in their attempts to have a significant effect on the network was when nodes give more attention to their neighbors who have similar opinions. The results of that experiment are shown in Figure 6. Closer inspection of the simulations showed that the decrease in the fraction of positive nodes seen in the figure can be explained by the fact that, because nodes value opinions that are close to their own more, a lot of nodes gradually move towards an opinion value of 0.

V. Conclusion

This paper has investigated the effects of opinion leaders on the opinion dynamics in social networks. Research has been done towards measures that can identify these opinion leaders. The result of this evaluation is that, while several measures perform well when the network is not very hierarchical, the out-degree of a node is the best indicator for opinion leadership in the majority of cases. When a group of nodes is selected based on their out-degree and initialized so that they favor one side of the argument, they can have significant effects on the rest of the network. This has been verified with simulations using the DeGroot model and simulations where small adaptations to this model were made. However, there is one scenario where this strategy failed. This happened when the nodes pay more attention to others who have similar opinions. Other strategies to influence the rest of the network should be explored for this scenario in future work.

References

[1] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Information Fusion*, vol. 43, pp. 57–65, 2018.

[2] Y. Dong, X. Chen, H. Liang, and C.-C. Li, "Dynamics of linguistic opinion formation in bounded confidence model," *Information Fusion*, vol. 32, pp. 52–61, 2016.

[3] E. Yildiz, D. Acemoglu, A. E. Ozdaglar, A. Saberi, and A. Scaglione, "Discrete opinion dynamics with stubborn agents," 2011, available at SSRN: https://ssrn.com/abstract=1744113.

[4] M. Laguna, G. Abramson, and D. H. Zanette, "Vector opinion dynamics in a model for social influence," *Physica A: Statistical Mechanics and its Applications*, vol. 329, no. 3-4, pp. 459–472, 2003.

[5] S. Righi and T. Carletti, "The influence of social network topology in a opinion dynamics model," in *Proceeding of the European Conference on Complex systems*, 2009.

[6] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, "Information network or social network? the structure of the twitter follow graph," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 493–498.

[7] S. Fortunato, "Damage spreading and opinion dynamics on scale-free networks," *Physica A: Statistical Mechanics and its Applications*, vol. 348, pp. 683–690, 2005.

[8] I. Rausch, Y. Khaluf, and P. Simoens, "Collective decision-making on triadic graphs," in *Complex Networks XI*. Springer, 2020, pp. 119–130.

[9] V. Domínguez-García, S. Pigolotti, and M. A. Munoz, "Inherent directionality explains the lack of feedback loops in empirical networks," *Scientific reports*, vol. 4, p. 7497, 2014.

[10] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[11] G. Mckeown and N. Sheehy, "Mass media and polarisation processes in the bounded confidence model of opinion dynamics," *Journal of Artificial Societies and Social Simulation*, vol. 9, no. 1, 2006.

[12] C. M. Schneider-Mizell and L. M. Sander, "A generalized voter model on complex networks," *Journal of Statistical Physics*, vol. 136, no. 1, pp. 59–71, 2009.

[13] H. Han, C. Qiang, C. Wang, and J. Han, "Soft-control for collective opinion of weighted degroot model," *Journal of Systems Science and Complexity*, vol. 30, no. 3, pp. 550–567, 2017.

[14] P. Taillandier, B. Gaudou, A. Grignard, Q.-N. Huynh, N. Marilleau, P. Caillou, D. Philippon, and A. Drogoul, "Building, composing and experimenting complex spatial models with the gama platform," *GeoInformatica*, vol. 23, no. 2, pp. 299–322, 2019.

[15] C. Pinciroli, V. Trianni, R. O'Grady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, M. Birattari, L. M. Gambardella, and M. Dorigo, "ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems," *Swarm Intelligence*, vol. 6, no. 4, pp. 271–295, 2012.

[16] F. A. Rodrigues, "Network centrality: an introduction," in *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. Springer, 2019, pp. 177–196.

[17] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Technical Report 1999-66, November 1999, previous number = SIDL-WP-1999-0120. [Online]. Available: http://ilpubs.stanford.edu:8090/422/

[18] J. M. Kleinberg, "Hubs, authorities, and communities," *ACM computing surveys (CSUR)*, vol. 31, no. 4es, pp. 5–es, 1999.

Contents

Preface i

Permission of use ii

Abstract iii

Table of contents viii

List of figures x

1 Introduction 1

2 Opinion formation 3

2.1 Opinion expression format 4

2.2 Opinion dynamics environment 6

2.2.1 Inherent directionality as a hierarchy measure 10

2.3 Fusion rules 13

2.3.1 DeGroot model of social influencing 16

3 Setup of experiments 18

3.1 Simulator 20

3.2 Pre- and post-processing 22

3.2.1 Preprocessing 22

3.2.2 Postprocessing 24

4 Influencing the network using opinion leaders **25**

4.1 Opinion leaders . . . 25

4.1.1 Translating opinion leadership to networks . . . 26

4.1.2 Centrality measures . . . 27

4.2 Influence using opinion leaders . . . 39

4.2.1 Unopinionated environment . . . 40

4.2.2 Environment with uniformly distributed opinions . . . 42

4.2.3 Environment with negative opinions . . . 47

5 Special cases **49**

5.1 Self-confidence . . . 49

5.2 Increasing importance of leaders . . . 52

5.2.1 Weight based on hierarchy . . . 53

5.2.2 Weight based on out-degree . . . 54

5.3 Weight based on similarity of opinion . . . 56

5.4 Noisy opinions . . . 60

6 Future work **62**

7 Conclusion **65**

Bibliography **68**

List of Figures

2.1 Line- and star-shaped patterns 9
2.2 Triagonal base patterns 12
2.3 Effect of switched connections on inherent directionality ξ 13

3.1 Experiment pipeline 19

4.1 Division of selected networks across ξ values 31
4.2 Correlation between average final opinion and closeness centrality 32
4.3 Correlation between simulation results and node level 33
4.4 Correlation between simulation results and PageRank 35
4.5 Correlation between simulation results and HITS 36
4.6 Correlation between simulation results and node out-degree 37
4.7 Number of zealots in the networks 39
4.8 Evolution of the fraction of positive opinions in an initially unopinionated network 40
4.9 Effect of number of leaders on final opinion in the network 43
4.10 Evolution of the fraction of positive opinions in an initially uniformly distributed network 44
4.11 Relation between ξ and simulation success 45
4.12 Simulation outcomes when 10% of nodes are selected as leaders 46
4.13 Evolution of positive opinions in an initially very negative network 48

5.1 Effect of self-confidence on opinion dynamics 51
5.2 Effect of number of opinion leaders 51

5.3 Simulation outcomes when 10% of nodes are selected as leaders 52
5.4 Effect of hierarchy-based weighting on opinion dynamics 55
5.5 Effect of out-degree-based weighting on opinion dynamics 56
5.6 Effect of weighting based on similarity of opinion on opinion dynamics . . . 58
5.7 Simulation outcomes versus ξ 59
5.8 Effect of noise on opinion dynamics 61

Chapter 1

Introduction

Public opinion is a well-known concept used to indicate what the majority of people think. However, the way in which each individual comes to its conclusion, called the opinion formation process, is still not fully understood. This process varies from person to person, making it difficult to study. Most research therefore focuses on using generally accepted abstractions on the level of the individuals to evaluate hypotheses on the group opinion formation. This evaluation happens through simulations. In these simulations there is usually no desired outcome in mind. Instead, they are run and afterwards the outcome is observed. In this study the goal is to identify global trends in the population when individuals with a large influence are used to try to steer the opinions of the other people.

Current opinion dynamics models usually consist of a set of standard elements. In chapter 2 these elements are explained in detail and the different options for each one are evaluated. To gather data on how the selected options affect the opinion formation process, simulations are done. The setup used to run and process these simulations is explained in chapter 3. After explaining how the experimental setup is implemented, it is used to evaluate multiple measures from graph theory on their effectiveness to identify key individuals in the network, as explained in chapter 4. During this identification, special attention is paid to their performance when there is a significant level of hierarchy present. Evaluation of these measures will lead to a decision to use one of them to rank the nodes in the network, with the goal of identifying opinion leaders. These leaders will then be used in the rest of chapter 4 to evaluate whether they have the ability to direct the rest of the network by agreeing on an initial opinion. This is done for both initially unbiased networks as well as networks where there is a preference for the non-goal opinion. Attention is also paid to the correlation between the number of leaders used to influence the network and the eventual outcome. Afterwards, in chapter 5, these results are compared to scenarios where extra assumptions are made on the aspects playing a role in opinion formation. These assumptions are translated into small adaptations to the original opinion dynamics model used and their effect is evaluated. Based on the outcomes of the simulations, some suggestions for future work are made. The results from all these experiments are then bundled into the final conclusion.

Chapter 2

Opinion formation

The dictionary defines an opinion as a view or judgment formed about something, not necessarily based on fact or knowledge [11]. How this opinion is formed, is however not further disclosed, the most likely reason being that it is a very complex process involving multiple factors [54]. Nonetheless it has been the topic of numerous studies throughout the years. In today's society, interaction with other people is a major reason for changes in opinions for most individuals [43]. This cause for a change in opinion is usually referred to as social influence and plays a big role in many phenomena like the spreading of fear during epidemics or the spreading of innovative ideas [43]. Factors playing a role in social influence include group structures and personal characteristics of the individuals in those groups [48, 6]. Because of the vital role it plays in fields like marketing and sociology, multiple models for social influence have been proposed and studied in earlier work [2, 10, 18].

The current models on opinion dynamics are usually composed of three main elements: the opinion expression format, the fusion rules and the opinion dynamics environment [15]. The combination of these three factors affects the results of the simulations and should thus be chosen carefully. The opinion expression format determines how each agent expresses his opinion. This can be either a discrete variable, continuous variable or even a vector depending on the goal of the research [60, 51, 55, 35]. The fusion rules determine how the exchange of opinions is handled. Again, there is a lot of variability between models, as some allow agents to share dishonest opinions or account for memory of an agent while others do not [3, 12, 10]. It is important to choose the fusion rules and opinion expression format with respect to the research goal. The opinion dynamics environment also plays a key role in the model as it determines between which individuals interaction is possible. In the current scope, opinion dynamics environment refers to the network structure that fully describes the relations between individuals. Unfortunately, there is a lack of empirical evidence regarding the structure of real world influence networks [59]. The size and structure of the opinion dynamics environment also depends on the circumstances under investigation. While some research focuses mainly on very hierarchical networks of agents, which might resemble a workplace environment for example, others may have multiple levels of networks taking into account both informal and formal connections between peers [55, 37].

In the current state-of-the-art the focus is mainly centered around the three possible end scenarios: consensus, polarization and fragmentation. A consensus is reached when all individuals in the network agree. The network is in a polarized state when the individuals are divided into opposing subgroups. When more than two opinions are present, the network is said to be fragmented. Because of the focus on the end scenarios, the simulations are usually done without intervention or desired outcome in mind. The recent past has shown that it might be interesting to examine how this opinion formation process can be influenced and manipulated in order to prevent for example the spread of fake news. In contrast, numerous preventive measures were taken in order to slow down the spread of diseases based on epidemic modelling, which draws many parallels to opinion formation models [16, 17]. For these reasons, the goal of this thesis is to look into ways to promote or inhibit the opinion formation in a hierarchical social network. This chapter gives an overview of opinion formation modelling with the intent of providing a solid base for the further exploration of the topic in the next chapters.

2.1 Opinion expression format

While expressing an opinion comes naturally for most individuals in the form of language, this is hard to model. Although there are models that use language to express opinions, it is quite uncommon [14]. Nonetheless, when simulating opinion formation there is a need for a model for opinion expression. To overcome this obstacle, a level of abstraction is added in the form of an opinion expression format. This format determines how each agent expresses his opinion. In order to enable complex simulations using mathematical models this expression of opinion is usually done with a numeric value [15]. As mentioned in the introduction of this chapter, this numeric value can be discrete or continuous depending the model and research goal. Discrete opinions have the advantage that the opinion of every agent is very clearly expressed. The downside that comes with this, is that they do not allow much room for nuance. They can also add a level of complexity to the fusion rules, as the outcome of those rules needs to be one of the discrete options. The opposite is true for continuous opinion expression formats. These allow individuals to express nuanced opinions. The nuance can even be so large that an individual is no longer expressing an opinion but instead expressing a doubt between the options. In some cases the opinion expression format is chosen to be a vector [35]. This vector should then be interpreted as a combination of discrete or continuous opinion values on different topics. This format is usually only chosen when the goal of the study is to specifically explore the correlation between opinion values in the opinion formation process.

Considering the fact that the goal of this thesis is to see how opinion leaders can influence the opinion dynamics, there is no need to simulate the effects of combinations of opinions on different topics. Introducing this does not only make simulating the opinion formation more computationally intensive, but it also introduces another level of control parameters for the simulation, as the correlation between these opinions needs to be taken into account. While this could be an interesting expansion, it is considered out of scope in this study. Keeping in mind the possibility of expressing nuance, it was decided to use a continuous variable for the opinion expression format. Without further consideration of the opinion expression format, individuals could have opinion values ranging from $-\infty$ to ∞. It is not hard to imagine the possibility that these extreme opinion values could significantly influence the opinion formation process. Taking for instance a population of individuals in which all but one individual have a low absolute opinion value and the only other person has an extremely high opinion value. In that case, this one person would have a powerful effect on all others and the result of the simulation would not be very useful.

One could compare this scenario with a group of people having a civilized conversation while one person is shouting above everyone else, thereby making it impossible for the others to have any meaningful interaction. This problem can be mitigated by setting an upper and lower limit on the possible opinion values. This is called a bounded continuous opinion [34]. This type of opinion expression format has the advantage that, while it leaves room for nuanced opinions and gradual change in opinion, there is a limit avoiding that a single agent dominates the rest of the network.

2.2 Opinion dynamics environment

Another important role in the opinion formation process is played by the environment [15]. The environment dictates the possible interactions between individuals. As it tries to model real networks of people, not every person is directly connected with every other person. When trying to simulate social networks, the choice for a graph to model the environment is the most logical as translation between reality and the model is straightforward [51, 59, 53, 36]. The graph structure represents the network of individuals and their connections. Each node in the graph represents a person and each edge a connection between two people. Direct interaction is only possible between nodes that are connected by an edge. Here, it is important to make the distinction between undirected and directed networks. In an undirected network edges do not have a direction, which implies that when two persons are connected by an edge communication can happen in both directions. In directed networks on the other hand, communication is only possible in one direction, which is determined by the direction of the edge. When looking at popular online social networks, this directed nature of influence is present in most cases and therefore directed networks will be used during this thesis [44, 38]. Using Twitter as an example, it is clear that when one user follows another this does not necessarily mean that the inverse is true [28, 21]. The study of the structure of these networks encompasses a whole area of research by itself. Using well-studied networks allows to focus on the underlying dynamics.

Unfortunately, as mentioned in the introduction, empirical evidence regarding the structure of real world influence networks is limited [59]. This suggests that an evaluation of different options is needed. Based on the goal of this study, there are several requirements for the opinion dynamics environment. First and foremost the goal is to model social networks. This implies that the chosen structure should exhibit characteristics of these social networks. So an option has to be chosen from the ones that are argued to model society well. Another requirement for the environment is that there is some form of hierarchy present. Since this thesis investigates ways in which opinion leaders can influence the opinion dynamics in hierarchical social networks, there needs to be a way to set up and control this hierarchy. While some studies pay attention to the possible combination of multiple levels of influence networks in the opinion formation process, this is not done here. The main reason why this is sometimes done, is to model and evaluate the differences between the different social circles an individual is part of and how these interact, which is not in the scope of this thesis.

Keeping this decision in mind, there is still a multitude of options to choose from. Over the years, multiple types of networks have been said to model society well. The common theme in these networks is that they try to model the statistical properties of society as a whole by manipulating the properties at the level of individual nodes. A type of network commonly found in the literature is a so-called scale-free network [19]. The defining characteristic of this type of network is the power-law degree distribution [45]. The degree distribution is the probability distribution over the number of connections that a node in the network has. The number of connections of a node is also called the degree of that node. A power-law degree distribution thus means that when the degree of the nodes is plotted versus the amount of nodes with this degree the curve follows a power function and the distribution can be written in the following form

$$P(k) \sim k^{-\gamma}$$

where γ is a constant and k is the node degree [45]. A network following this power-law degree distribution consists of a minority of very well-connected nodes with a high degree and a large number of nodes with few connections. A popular example of a network that exhibits this property is the internet, where there are a few websites that are pointed towards by a large number of others, while the majority of websites only has a very small number of other websites linking towards them. It is sometimes argued that this type of network is also a good model for social contacts in society [23]. Looking at social media again for an explanation, examining the number of followers or friends users have, shows a similar trend, with a small number of users having up to millions of peers while most have close to none in comparison. One of the algorithms that can be used to generate scale-free networks is the Barabasi-Albert algorithm [1, 57]. This algorithm works in two phases:

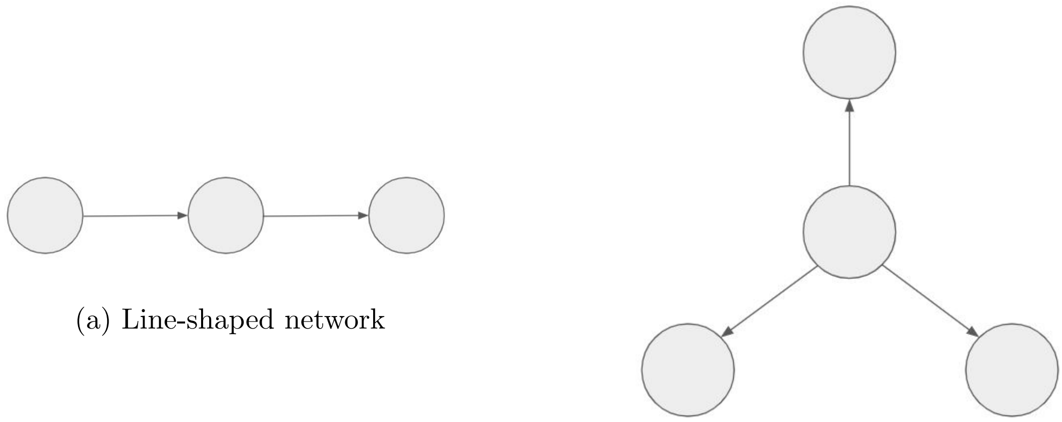
1. Initialization phase: A set of m_0 nodes is created that are fully interconnected by adding one node at a time to the network and adding a connection from this node to all other nodes already present in the network.

2. Growth phase: Every time a new node is added, it will have m ($\leq m_0$) edges to existing nodes. Choosing the nodes to connect with is done based on the degree of these nodes such that the probability p_i of connecting to node i with degree k_i is

$$p_i = \frac{k_i}{\sum_j k_j}$$

Because of the use of this particular equation for p_i, which determines the nodes to which these new nodes are connected, the network ends up with the scale-free characteristic. This way of adding connections is also called preferential attachment. The problem with using scale-free networks here, lies in the fact that this thesis focuses on hierarchical social networks. Note that scale-free networks can exhibit signs of the presence of a hierarchical structure [50]. In this algorithm there is however no way to control the hierarchical nature of the resulting network. Making changes after the generation could destroy the scale-free property of the network and is therefore not advisable.

Because of this problem with the control of the hierarchality, the decision was made to look for alternatives to scale-free networks. Specifically, the search focused on a type of network that allows for controlling the hierarchy. An obvious answer to this problem would be the use of networks based on line- or star-shaped patterns. These networks allow for hierarchy to be clearly present as can be seen in Figure 2.1. In Figure 2.1a the leftmost node is the highest in the hierarchy, followed by the middle and right node. In Figure 2.1b the



(a) Line-shaped network

(b) Star-shaped network

Figure 2.1: Line- and star-shaped patterns

middle node is the leader, based on the structure of the network with all other nodes being on the same but lower level in the hierarchy. The question here is whether or not these networks are still a good representation of social networks. Even when expanding these networks by adding more of the same patterns, they remain very simple. This simplicity of course contradicts the presence of complex communication patterns in real social networks. These are greatly reduced when compared with the scale-free networks inspected earlier. The risk of using these networks is that, next to the opinion dynamics, the structure of the network strongly determines the experiment outcome. To find a middle ground between the possibility to control the hierarchy present in the network and how well they represent social networks, triadic graphs were considered.

Similar to the networks evaluated in the previous paragraph, triadic graphs are constructed from a set of basic patterns [49]. These basic patterns, used to construct more complex graphs, are called motifs. In the case of triadic graphs, these motifs are triangular. These triangular connections are prevalent in networks of communicating entities [49]. They also allow to control the hierarchy when creating the network. For these reasons it was decided to use these triadic graphs for the opinion dynamics environment in this study. To be able to verify the control on the level of hierarchy present in the network, a measure is needed which indicates how hierarchical a network is. The next section deals with this measure and examines whether the statement that the level of hierarchy can be controlled in triadic graphs is valid.

2.2.1 Inherent directionality as a hierarchy measure

To be able to judge whether hierarchy is present in a social network, a standardized way of comparing graphs is needed. Ideally, an analytical measure should be set up which takes the network as an input and returns a numeric output, indicating how hierarchical the network is. Based on this output, a comparison between different graphs is done and a well-grounded decision can be made on whether or not the network is considered hierarchical. Research on these measures has already been conducted with different strategies and outcomes [42, 25, 13]. According to literature, a mathematically sound way of measuring hierarchy, is by calculating the inherent directionality ξ of the network [13]. ξ gives an indication of how well all nodes could be ordered along a one-dimensional axis, such that the links existing between these nodes align as much as possible with respect to their pointing direction. In this way, the existence of this inherent directionality is related to the existence of a hierarchical structure in the network. A higher ξ value indicates a more hierarchical network structure.

Calculating ξ comes down to measuring the fraction of edges going in each direction of the imaginary axis on which the nodes have been ordered. By taking the maximum of these two fractions, a percentage between 50 and 100 is obtained. A ξ value of 50% indicates that there is no apparent direction present for the communication in the network and thus a relatively unhierarchical graph. A value of 100% indicates that all communication in the network goes in the same direction. Evaluating ξ for the networks presented in Figure 2.1 gives ξ=100% for example, as all edges go from a higher level in the hierarchy to a lower level. To evaluate whether an edge points up or down the hierarchy, the levels of the nodes connected by this edge need to be known. By comparing these levels and the direction of the edge, a decision is then made on whether it goes up or down the hierarchy. In order to calculate the level l_j of each node j a system of linear equations is set up. To understand the equations, some definitions from graph theory need to be introduced first.

Definition. In-degree and out-degree of a node
The in-degree and the out-degree of a node are, respectively, the number of incoming and the number of outgoing connections at that node.

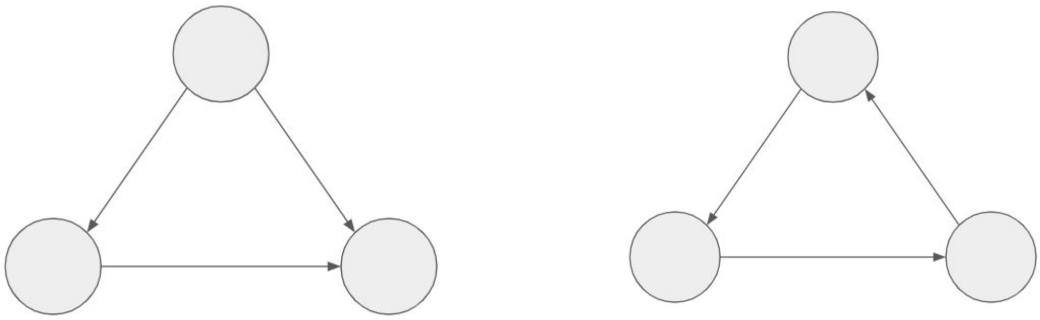
Definition. Adjacency matrix
The adjacency matrix of a graph with n nodes is an n-by-n matrix for which the entry at row i and column j is 1 if there is a connection from node i to node j or 0 if there is not.

Definition. Basal node
A basal node is defined as a node with 0 in-degree and thus has no link pointing to it. These nodes are also referred to as zealots.

Given these definitions, the set of equations used to calculate the level of each node in the hierarchy can be formulated as [13]:

$$\begin{cases} l_j = 1 + \frac{1}{k_j} * \sum_{i=0, i \neq j}^{n} A_{ij} * l_i & \text{, non-basal node} \\ l_j = 0 & \text{, basal node} \end{cases} \tag{2.1}$$

In equation 2.1 k_j is the in-degree of node j, A_{ij} is the entry in the adjacency matrix in row i and column j and l_j is the level of node j. Not every network necessarily has basal nodes in it. This can cause the matrix describing this system of equations to be singular, making the equations unsolvable because of the degrees of freedom present. To address this issue, the least-squares method of solving linear systems of equations is used, which results in each node getting a floating point number as a level. When the levels of all nodes in the system have been identified using these equations, ξ can then be calculated easily by iterating over the edges and checking their direction. To automate this calculation a script was created in the Python programming language which takes a list of the adjacency lists for each node as input. The adjacency list of a node is a list containing the numbers of all the other nodes for which there is an edge pointing from the current node towards this other node.



(a) Feed-forward loop (b) Feedback loop

Figure 2.2: Triagonal base patterns

Having identified a measure for checking whether a network structure contains hierarchy or not, the next question is whether the hierarchality can be controlled by the topology of triadic graphs. As mentioned before, these triadic graphs consist of a combination of triagonal patterns. The graphs used in this study are constructed from two specific kinds of these triagonal patterns: the feed-forward loop and the feedback loop. These patterns are shown in Figure 2.2. Calculating ξ on the basic patterns shows that the feed-forward loop shown in Figure 2.2a is highly hierarchical as all edges go down the hierarchy while the feedback loop shown in Figure 2.2b is not as there is no clear hierarchy present. Based on these observations, the assumption is made that by combining a number of these patterns, ξ values between 50% and 100% can be obtained. To evaluate this, a set of 5000 networks consisting of 343 nodes was used. This set was constructed using both networks consisting of 343 and 686 of these triagonal patterns. For each of these two types, 50 seeds for the random number generator were used to generate graphs. These initial 100 networks consisted only of feed-forward loops. By reversing some connections in these networks and thereby changing the feed-forward loops into feedback loops, additional networks were generated. For the networks consisting of 343 feed-forward loops, this reversal of connections happened in steps of 7. For those having 686 initial feed-forward loops, a step-size of 14 was used. In this way 50 networks are generated for every seed, resulting in 5000 networks. Revisiting the earlier hypothesis, the expected outcome of calculating ξ on each of these networks should show that when the number of reversed connections increases, the value

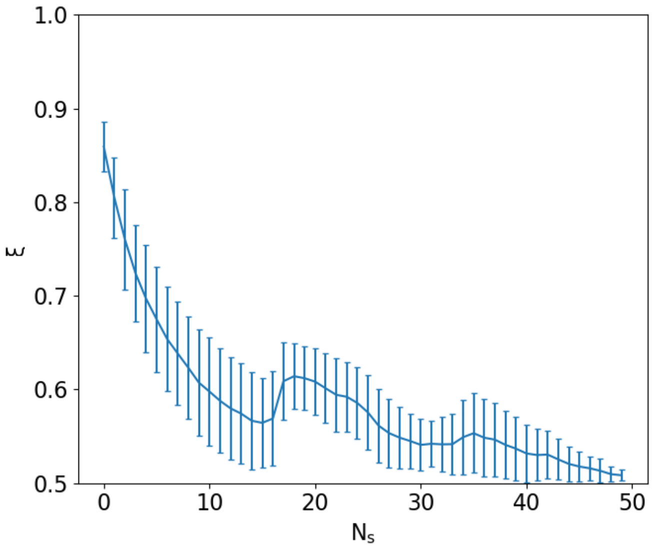


Figure 2.3: Effect of switched connections on inherent directionality ξ

for ξ decreases. Figure 2.3 shows that this indeed happens and thus confirms the hypothesis. On the x-axis the number of switching events N_s is shown, being either 7 or 14 reversed connections per step depending on the amount of triagonal patterns present in the graph. On the y-axis the inherent directionality ξ is visualized. For each of these switching events, the average ξ value and standard deviation were calculated across all 100 networks. The figure shows a downward trend when N_s increases. Based on this observation, the choice for triadic graphs as opinion dynamics environment is preferred as it allows controlling the hierarchy and resembles realistic social structures. The slight increase between N_s=15 and N_s=20 is due to the fact that around this point the occurrence of nodes with zero in-degree disappears. This makes the calculation for ξ slightly imprecise.

2.3 Fusion rules

Within the context of opinion dynamics, individuals need to be able to share and change their opinions. In reality, this is an extremely complex process for which there is no consensus on how it works to date [39]. To enable simulation, levels of abstraction have been created to model the different factors playing a role as shown in sections 2.1 and 2.2 for the opinion expression format and opinion dynamics environment respectively. While these are important parameters, the real complexity lies in the fusion rules. The fusion rules determine how the exchange in opinions is handled for each agent. The term fusion rules comes from the fact that each agent changes his opinion by fusion his current opinion with those of his neighbors, using a set of mathematical rules. The fact that these rules are the same for each agent in the simulation, is what makes agent-based simulations a popular strategy for examining the opinion dynamics. Because the fusion rules deal with the incorporation of the opinion of others into your own, they depend on the opinion expression format and the opinion dynamics environment. The fact that there is a plethora of reasonable possibilities to choose from for each of these three factors, is one of the reasons there are many different possibilities being explored [15].

The opinion expression format has implications for the set of rules that determine how the individual incorporates the opinions of his peers into his own because the mathematics used for one format can not always be used for all formats. For instance, the dynamics used for a discrete format can not in all cases be applied to a continuous format and vice versa. On top of this, some models allow agents to share fake opinions and thus lie to others or express uncertainty. There are also models which account for an agent's memory with respect to its own opinion. Each of these options calls for a change in the fusion rules. In the current research however, the assumption is usually made that agents share honest opinions and there is only one single form of opinion expression format used [15]. Because of this immense option space, it is important to choose the fusion rules and opinion expression format with respect to the research goal. As mentioned in section 2.1, the opinion expression format used here is a bounded continuous variable. Agents are also assumed to share their opinion honestly. The opinion dynamics environment has an impact on the fusion rules, because it determines which and how many other opinions need to be taken into account when forming an updated opinion.

After deciding how each agent expresses its belief in the topic at hand and what the network of agents looks like, it is important to delve deeper into the mathematical models of social learning. Does an agent take into account its own previous opinion by having a certain degree of self-confidence? Are the opinions of all neighbors equally as important or is there a way of classifying which neighbors have more importance? The question studied in this thesis is in which way the answers to these questions impact the result of the simulations and which conclusions can be drawn from this. In most studies one of three basic fusion rules is chosen or an extension of these is used [15]. These three basic models are the DeGroot model, the bounded confidence model and the voter model [10, 9, 8]. In the voter model, all nodes are supposed to be ordered along a regular lattice. Each node updates its opinion by choosing a direct neighbor randomly and adopting that neighbors opinion. The opinions are usually binary variables indicating two possibilities to choose from. This is where the name voter model originated from. Because of this model's limits regarding the possible interaction patterns, this raises the question of how representative it is of real opinion dynamics. The bounded confidence model tries to be more representative by taking into account psychological factors when an individual updates his opinion. The name stems from the fact that an individual only pays attention to other opinions when they are not too different from its own opinion. This is supposed to simulate the behavior of individuals when the assumption is made that people are more likely to believe others when these others confirm their beliefs. This is however a strong assumption to make. The DeGroot model is a more generalized version of these models. In the DeGroot model, the opinions of all peers of the node are taken into account when updating its opinion. Different weights can be assigned to these opinions but in the most general version of the model all weights are equal. These weights typically do not change over time. Because it is a very general model and it has been studied for a long time, it is considered to be the classic model in opinion dynamics [15]. Based on this information, it was decided to use the DeGroot model throughout this thesis, starting with the most general version as a baseline for further experiments. In these further experiments, assumptions are then made on the distribution of the weights assigned to the opinions of neighbors and the effects of these assumptions are evaluated. The following section presents a more detailed explanation of this model.

2.3.1 DeGroot model of social influencing

The DeGroot model of social influencing, named after Morris H. DeGroot, who first proposed it in 1974, has been around for several decades [10]. Although it is not a very sophisticated model at its core, that does not make it any less relevant as is proven by the fact that after 46 years it is still actively used [15]. Especially in cases where the goal is to analyze the theory behind network dynamics, its simplicity is helpful in gathering insights in the processes going on in the networks. The key idea behind the model can be summarized with the famous sentence "you are the average of the 5 people you spend the most time with" [24]. Translating this to a more formal description of the model means that at every instant each node takes an average of the opinions of its peers. In the most generic form of the model this is all there is to it. Since it has been around for such a long time however, this version has been studied thoroughly and therefore in some studies, instead of taking the normal average of the opinions of the peers, a weighted average is used [10, 36, 26]. The weights assigned to the opinions of the peers can be changed depending on the characteristics of the neighbor. This way, the weights represent the amount of social influence between agents and each node can be assigned a certain level of importance. Without loss of generality, the version using weighted averages can be used to write down the equations used by node j to calculate its new opinion o_j, obtaining the following results:

$$\begin{cases} o_{j,t+1} = \sum_{i=1,i\neq j}^{n} W_{ij} * A_{ij} * o_{i,t} & \text{, in-degree} > 0 \\ o_{j,t+1} = o_{j,t} & \text{, in-degree} = 0 \end{cases} \tag{2.2}$$

In equation 2.2 W_{ij} is the importance node j attaches to the opinion of node i and A_{ij} is the entry in the adjacency matrix indicating whether there is a connection from node i to node j. To obtain the normal version where there is no difference in weight between neighbors, the non-zero entries in column i of the weight matrix W should equal $\frac{1}{k_i}$ where k_i is the in-degree of the node i. This equation is sometimes presented without the adjacency matrix A. In this case, the adjacency is still implicitly present because an entry of 0 in W indicates that there is no connection between these two nodes.

As mentioned in the previous paragraph, changing the weights can be interpreted as changing the social influence of a node. When making changes to these weights, special attention is required to make sure that the sum of the weights assigned to all neighbors of a node equals 1. If the sum of the weights is higher or lower, certain values for the opinion are more likely to occur. When the sum is higher than 1, the fusion rules will favor a higher absolute value of opinion, when the sum is lower the fusion rules will cause the opinion to converge to 0. This causes the results to differ from the outcome of the opinion dynamics and initialization and makes them unusable. As the goal is to study ways to influence the opinion dynamics in a network using these interactions and initial opinions, special attention needs to be paid to these types of small implementation errors, which could determine the outcome of an experiment before it has even started. The specifics of how and why weights should be changed between neighbors or simulation steps are dependent on the specific scenario under evaluation and will be presented in chapter 5.

Chapter 3

Setup of experiments

The goal of this study is to investigate the role opinion leaders play in opinion dynamics. Before any statements can be made about the influence they have, data needs to be available from which these conclusions can be drawn. To generate this data, a set of scenarios to be simulated is established. Based on these scenarios, a setup is constructed to run the experiments as well as gather and process data in an efficient manner. Throughout the different scenarios, the general setup remains largely the same across all simulations. The goal of this chapter is to present an overview of this setup, explaining how simulations are run, starting from the scenario that needs testing and concluding with the processing of the simulation results.

The general outline of the setup is presented in Figure 3.1. As indicated, the first step consists of defining the scenario that needs to be simulated. Once a scenario has been worked out, the next step is to determine which input is required and generate this necessary input for each of the cases to be run. For the implementation of the opinion spreading simulations, the GAMA platform simulator was chosen [56]. GAMA is short for GIS & Agent-based Modelling Architecture. The reasoning behind this decision, alongside with the alternative that was considered, are explained in section 3.1. This simulator takes an XML file as input, which defines the experiment-specific parameters. Apart from this XML file, another file describing the opinion dynamics environment is also required in the current context. The environment is described using a CSV file that contains a list of all

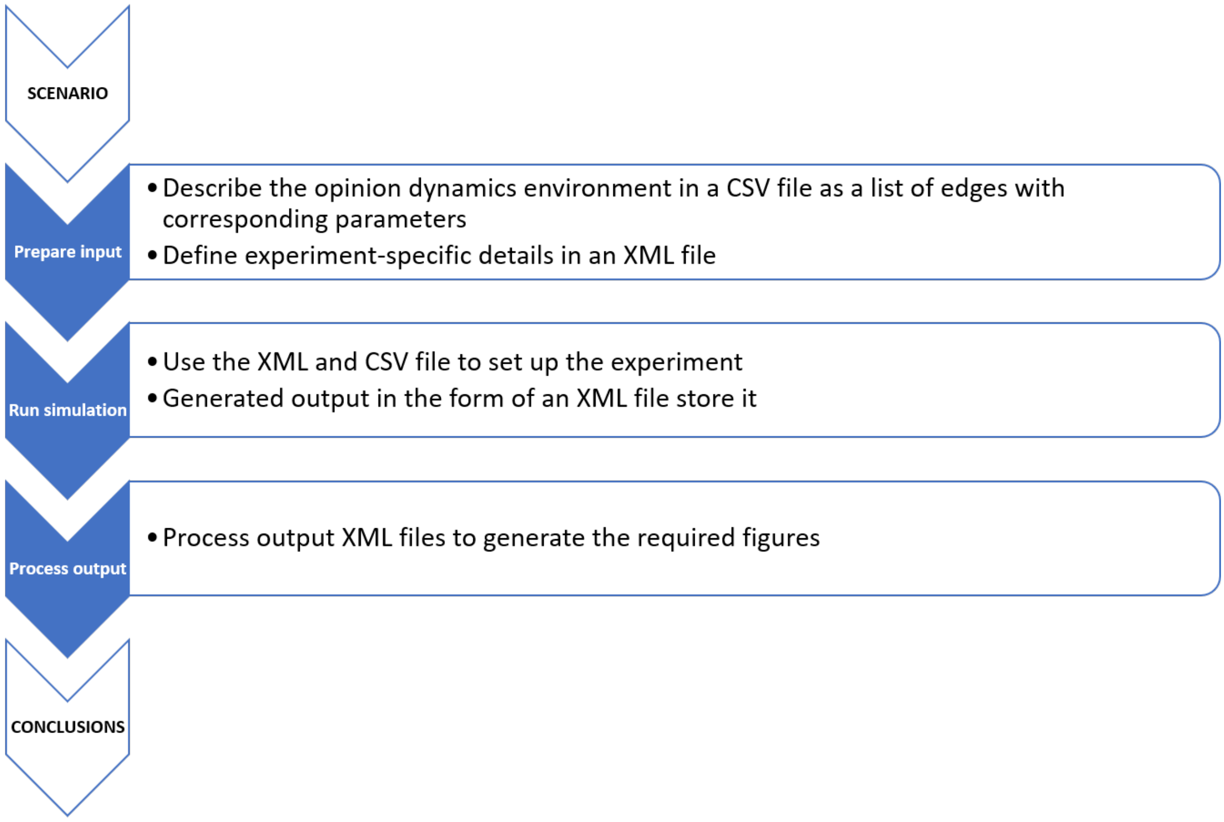


Figure 3.1: Experiment pipeline

the edges in the social network. Each row in the file stores one edge as a combination of the source and destination node number. In case additional edge parameters are required, these can be stored in the same row. In these files the node numbers start at 1 and go up to 343. In section 3.2 a more detailed explanation of how these files are built up and generated automatically is presented. Given these two files, the experiments can be carried out using the command line interface available for the simulator. The output of the experiments is stored in a folder on the file system, again using the XML format. This output contains the values for certain variables at each step of the simulation. These values can be used for further processing and analysis of the results to then draw conclusions. In the following sections the different stages of the simulation pipeline are explained in greater detail, starting off with the central part, the simulator, as it determines the pre- and post-processing steps.

3.1 Simulator

As mentioned in the introduction, GAMA platform was chosen to run the simulations. Before the decision was made to go with this simulator, other options were also explored. An important factor to consider was the fact that the simulator needed to lend itself for agent-based simulations, as multi-agent systems are more suitable to investigate this type of social behavior [40]. Based on available experience, the current state of the art and trends found in other research, there were two options: GAMA platform and ARGoS simulator [47]. Both of these simulators focus on agent-based simulations. Each of these simulators therefore allows to write the simulation code in line with this paradigm, meaning that the code specifies the exact behavior of each agent but not the behavior of the group. For GAMA platform this is done using a custom language while the ARGoS simulator uses C++. The reason why C++ is used is that the code written for ARGoS simulations is meant to be run on a specific kind of physical robots. Although this can be an advantage in some studies, it does not benefit the current goal. GAMA platform on the other hand focuses on more abstract, high-level concepts in the simulations. The consequence of this different focus is that the custom language used in this simulator provides many features for modelling these high-level concepts out-of-the-box. In fact, it has built-in support for network structures defining the environment in which the agents are placed. GAMA platform also has extensive documentation publicly available on its website. For these reasons, GAMA platform was used as the simulation tool within the framework of this thesis.

The custom modelling language used in GAMA platform is called GAML, an abbreviation of GAma Modelling Language. It is easy to learn for someone with previous programming experience and there is a full tutorial available on their website. In this tutorial the capabilities of the simulator are introduced, alongside the syntax of GAML, implementing increasingly complex experiment definitions along the way. As indicated previously, the simulator also comes with an implementation of certain high-level concepts from graph theory such as nodes and edges, which enabled writing clean code for the simulations used in this study. Graph structures are available as a standard datatype in GAML. This datatype comes with a set of operations which greatly reduce the time needed to program experiments involving graphs. Especially useful in this regard was the use of the procedure to find neighbors of nodes. Because of the DeGroot model, this is a very common operation in the simulations. The simulator has a separate view that is used when an experiment is running. In this view, figures can be constructed to allow tracking of important experiment variables in real time and visual debugging of simulation setups. This makes the task of setting up new experiments a lot simpler as there is no need to go through the remainder of the pipeline when trying to find small errors in only one part of it. Furthermore, this also allows for quick prototyping of new simulations to test out hypotheses.

Although the graphing and data-gathering tools present in the simulator are adequate for quick analysis and debugging in separate instances of an experiment, this is inconvenient for larger experiment definitions possibly involving different networks. When the goal is to run a single experiment on thousands of graphs, a systematic way to set everything up and store the resulting data to be processed later is required. GAMA platform allows this because it has a command line interface. This interface permits running experiments in a headless mode, meaning that no window has to be opened when a simulation is started. This makes the process of running multiple simulations significantly faster as it greatly reduces the overhead of starting and stopping the program every time a new one is started. When running experiments from the command line, XML files, which describe the experiment, are needed as input. In non-headless mode, all variables can be declared in the program itself using the GAML file. When using the headless mode however, the new values need to be defined in the XML input file so that the simulator knows how to change these between different instances of the simulations. As mentioned before, a CSV file is also used to define the opinion dynamics environment. The location of this file on the system is an example of the variables that need to be changed in between simulations and is thus present in the XML input file. One of the first steps in all the simulations is to use this CSV file to set up the network of agents, as described in chapter 2. Once the agents are created, an iteration over the CSV file with all the edges is performed, adding them one by one using the standard operators present in the GAML language. Simulations generate XML files which contain the values of a set of tracked variables at every step of the simulation. If required, the simulator still provides a way to capture images of the simulations when it is used in headless mode. This allows closer inspection of the simulation in instances where visual confirmation of the results might be required, while still enabling the efficient running of multiple simulation instances in a row.

3.2 Pre- and post-processing

As indicated in section 3.1, a lot of files need to be created and processed for execution and processing of the simulations. While reuse of some of these is possible, in a lot of cases an automated mechanism is required to take care of the others. Fortunately, these files are highly structured, so scripts can be written to carry out this task. Because of the immense library of publicly available modules and based on previous experience, it was opted to use Python as the programming language in which to write these scripts. Another advantage of using Python is that it enables the execution of command line tasks from within the Python code. This means that running simulations in headless mode can be started within a script.

3.2.1 Preprocessing

For every simulation two files are required: one describing the opinion dynamics environment (the network of agents) and another containing the values for the variables that differ between simulation instances. Initially, each graph describing a network was defined in a separate text file. However, iterating over CSV files is more convenient in the simulator compared to text files. Because of this, the initial files had to be converted. To do this conversion in an efficient way, a script was written which generates a CSV file for each network, containing a list of edges labeled by the number of their source and destination nodes, based on the text files. This CSV file with the network structure can be reused and only had to be converted once. The information defining the network as a whole is encoded in the file name so it can still be used when needed and no information is lost in the process. Depending on the scenario under test, it is possible that extra parameters for the edges are required. In these cases, an iteration over all the network files is required to calculate and add these parameters.

The second file that is required in order to run an experiment in headless mode, is the XML file with the specific values for simulation variables. One of these variables is the location of the CSV file defining the opinion dynamics environment. Because this differs for every instance, this implies that for every scenario to be tested 5000 different simulation definitions have to created. Using the Python core XML processing module an experiment definition template file is copied, changing the values of the required parameters to the ones to be used in the simulation. These simulation definitions are then combined in one file, which is then passed to the program as a command line argument. The reason why they were all combined into one file is that the simulator allows this, with the goal of running them one after the other without having to start and stop the program every time. Apart from the location of the CSV file containing the network structure, another important variable to be changed between instances of the experiment is the seed for the random number generator used by the simulator. Since certain aspects of the simulations, such as the initial opinions of nodes, are generated using the built-in pseudorandom number generator, it is important to change the seed of this generator between experiments so that the results are not dependent on its value, keeping in mind that results need to be replicable later. If the same seed would be used for all instances of the experiment, the observations obtained from the experiments could be dominated by random fluctuations, hiding the more insightful fundamental system behavior. To ensure both replicability and the use of different seeds, the same distribution of random seeds was used for the original generation of the networks. The seeds used for the generation of these graphs were encoded in their filenames and thus available to be used during the setup of the experiment. In this way 50 different seeds are used across the experiments, ensuring that any conclusions stem from the actual implementation of an experiment instead of random chance. The last variable that was important in every single one of the scenarios was the experiment ID. This was of no importance for the experiment itself. Instead, it was important out of practical considerations, since this variable determines the name of the output file generated by the simulator. Being able to trace results back to the network and experiment from which they originate, allows for investigation of possible inconsistencies and debugging of simulations. Further variables depended on the experiment at hand, but were added to the definition template and script using the same process.

3.2.2 Postprocessing

When using the simulator in headless mode, the output of simulations is stored in XML format. For every simulation, a new file is created in the same folder that has the simulation ID in the filename. This folder is passed to the simulator as a command line argument. These files contain the values for certain variables at every step of the simulation and are structured based on these time steps. Indicating that a variable should be tracked in the output is done by changing the GAML file in which the simulation is defined. Since not all variables are important for the analysis, this limits the size of the output files and allows for clean processing code. A benefit of tracking each of these variables step by step, is the fact that no experiment has to be repeated in case the value at earlier instants in time is required. As the values are grouped per time instant, processing is intuitive. Using the Python module for processing XML files, these values could be translated into variables in the script. Based on the goal of the experiment, these are then used for further analysis and the generation of graphs.

Chapter 4

Influencing the network using opinion leaders

Now that an overview has been given on the current state of the art in opinion formation research, the base model for this study has been chosen and a system for running simulations has been set up and discussed, the following step is to define the experiment scenarios. Based on these scenarios, the influence of opinion leaders on the rest of the network can be investigated. The eventual goal is to derive a general set of rules, formed from the experiment results, that determine how the effectiveness of these opinion leaders in changing the opinion dynamics can be estimated. In section 4.1 an evaluation of the term opinion leadership is given, explaining how it is interpreted and calculated in this study. Afterwards in section 4.2, the opinion leaders, identified using the results from section 4.1, are put to the test. This is done by placing them in a series of scenarios where the other nodes are progressively more biased against the opinion leaders.

4.1 Opinion leaders

What makes someone an opinion leader? The term opinion leadership stems from the two-step flow of communication model [30]. As the name suggests, this model assumes that communication and thus influence happens in two steps. In the original study, the origin of all information is assumed to be mass-media. The hypothesis is then that this information flows through opinion leaders, who interpret it and put it into context. The general public uses this interpretation and forms its opinion based on how the opinion leaders have interpreted the information and in which context it has been placed by them. In today's world however, one could make the argument that opinion formation would follow a one-step flow of communication. This is because a large percentage of people get their information from one source, being social media [31]. With the rise of big data analysis, more and more of this information is tailored to each person separately, seemingly cutting out a step in the two-step flow of communication model. Nonetheless, when looking at the structure of popular social networks, it is clear that a lot of people still follow so-called influencers. Studies examining whether this model is still valid today have found that, when analyzing the nature of communication flows via the digital platform Twitter, long-standing communication theories, like the two-step flow model, are still valid while direct one-step flows and more complex network flows are also present [27, 7]. Taking this into account, an opinion leader is defined as follows:

Definition. Opinion leader
An opinion leader can be characterized as an individual who presents a version of information in the form of an opinion to a number of peers. Those peers pay attention to this information and use it in their own opinion formation process. By doing so, the opinion leader has a non-negligible impact on the opinion formation process of the network as a whole.

4.1.1 Translating opinion leadership to networks

The definition of the term opinion leadership implies that there are both nodes that spread opinions and nodes that receive these opinions. In this study, all edges in the network are directed with the underlying implication that influence does not necessarily have to go both ways, as explained earlier in section 2.2. Based on this, the definition of an opinion spreader and an opinion follower as used in this thesis can be given as follows:

Definition. Opinion spreader
An opinion spreader is a node in the network that has an out-degree higher than 0.

Definition. Opinion follower
An opinion follower is a node in the network that has an in-degree higher than 0.

It is clear that these are not mutually exclusive. In fact, most nodes will be both an opinion spreader and an opinion follower, as they have both outgoing and incoming connections. The distinction between an opinion spreader and an opinion leader lies in the fact that even though a node might spread its opinion to a peer, this does not necessarily mean it does so in a way that has a non-negligible impact on the network as a whole. While the influence of opinion spreaders plays a big part in the result of the opinion dynamics in the network, the importance of opinion followers should also not be disregarded. Studies have shown that for influential individuals to have an effect on the network there needs to be a critical mass of opinion followers present [59].

Judging by the density of the nodes and the direction of the connections, a person could intuitively sense which nodes might be more important than others just by looking at the graph. However, since computers lack this visual intuition, the question remains how to translate this to an approach that can be followed by a computer. In graph theory there are many measures to calculate the influence of nodes, e.g. by considering so called "centrality measures". In section 4.1.2 a selection of these measures is presented. An evaluation of these measures is done to check which give a good indication of opinion leadership. For each measure, an experiment is performed to gather data to investigate its effectiveness as an indicator for opinion leadership.

4.1.2 Centrality measures

It is important to know which nodes play a critical role in the network. Could there be a way to rank all the nodes in the network, with the first being the most important and the last being the least important? This question of course largely depends on what is meant with importance. Multiple approaches with the goal of ranking the nodes in a network have already been explored to great detail in different contexts [46, 33, 20]. In graph theory there are multiple metrics to measure the centrality of the nodes for example [45]. The centrality gives an indication of the distance to all other nodes in the network, making it a good start for identifying the important nodes in the current context. The core question is thus which existing measures could be used to assess the influence of a node in the simulations carried out in this study. The goal of this leadership identification is to discover a set of opinion leaders. Using this set of nodes, the influence of opinion leaders on the network as a whole can then be explored in further simulations.

As the goal is to identify which nodes carry importance during the simulations, it makes sense to use the simulations to find a reference point, using the opinion dynamics described in chapter 2. Having this reference, a comparison between the measures can be carried out to eventually find the measure with the highest correlation with the influence of a node. The goal of these reference simulations would be to measure how much each node can influence the rest of the network by itself. Because the communication between all nodes in a network can be complex, the interactions between all nodes not involving or caused by the node under investigation should be limited as much as possible. This is the only way to obtain an objective indication of the influence of this single node. To accomplish these goals, an experiment has been set up with the specific goal being to determine the influence each single node has on the network as a whole. In this experiment every node in the network has an opinion value of 0 initially, except for the node under examination, which has an opinion value of 1. As all other nodes are initially unopinionated, meaningful interactions can only be caused by the one node that has a different initial opinion. Because of the overwhelming amount of initially unopinionated nodes, there is a risk that the node under investigation has no chance to spread its opinion, but instead immediately gets influenced by its peers instead. To prevent this and actually gather useful data, this one node is also made stubborn in the sense that it does not update its own opinion like the rest of the nodes do. Instead it just keeps spreading its opinion value of 1 at every step. Nodes with this kind of behavior are called zealots in literature and posses the potential of having a big influence on the opinion dynamics in the network [58, 36]. By letting the simulation run for a predetermined amount of steps, the rest of the nodes in the network, using the DeGroot model, will determine their new opinion at each time instance. At first, only small changes will occur on the network level because only one node actually has "something to share". As time progresses however, and based on how well connected this initial node is, more and more nodes will start receiving meaningful updates from their neighbors and have opinion values ranging between 0 and 1. Thus, by tracking the average opinion of the entire network after this predetermined amount of steps, a data point is obtained to compare the other measures with. A starting node with more influence will result in a higher average opinion than a starting node with zero to no influence. This should also be reflected in the results of the calculation of the measures.

In theory it would be possible to run these experiments for every node in every network used. The data obtained from all these simulations could then be used to construct the set of opinion leaders and further explore how these nodes can influence the rest of the network in more realistic scenarios where the entire network is opinionated for example. A big downside to this approach is the amount of time it would take to run all these experiments. In order to have a faster way of identifying opinion leaders, an exploration into existing, analytical measures was done. Using these measures, an explanation of how an opinion leader can be identified using the network characteristics can also be derived. The outcomes from running the experiment described above on the nodes of a selected set of networks can be compared to the results obtained by calculating existing measures on these nodes with the hopes of finding similarity. Once such measures are found, the leadership identification can then happen by calculation of this measure instead of having to run a large number of simulations.

To identify opinion leaders, a measure is needed on the node level that approximates the results of the simulations. Both in graph theory and sociology, a lot of effort was invested into determining ways to identify the most important nodes in a network [20]. The wording used may differ but the core question remains the same: which nodes are more important than others in relation to the experiment outcome of the network as a whole? The importance of nodes in these scenarios stems from their ability to distribute information or opinions. The existing measures that fall into this category are commonly grouped under the name centrality measures. The next paragraphs provide a definition and explanation of the tested measures. The results of the comparison between these measures and the average final opinion of the network is also presented. The evaluation of these centrality measures is done by putting them side by side with the results of the simulations with a single, stubborn, opinionated node. To measure the relation between the simulation results and the calculated measure, the correlation between the two is calculated for a subset of the networks. A sufficiently high correlation indicates that the simulations could be replaced by calculation of the measure under investigation and produce a similar set of opinion leaders. The specific correlation coefficients used to compare the two results for each node are the Pearson correlation coefficient ρ_p and the Spearman rank correlation coefficient ρ_s [5, 61]. ρ_p is defined as

$$\rho_p = \frac{\sum_i (x_i - \bar{x}) * (y - \bar{y})}{(n-1) * s(x) * s(y)}$$

where $\bar{x}$ indicates the average of all x values, $s(x)$ their standard deviation and n the amount of data points. x_i and y_i are the results of both measures calculated on the same data point i. The reason why these two correlation coefficients are used is that, while the Pearson coefficient gives an indication of correlation, the results can be affected by the presence of outliers [32]. To prevent possible outliers from impacting the results too much, it was decided to add another correlation measure which is not dependent on the actual values: the Spearman rank correlation coefficient ρ_s. To calculate ρ_s the results are first sorted and the resulting ranks are then compared for each node, resulting in the formula

$$\rho_s = 1 - \frac{6 * \sum_i (r_i - s_i)^2}{n * (n^2 - 1)}$$

where n stands for the number of data points again and r_i and s_i are the rank of the result when the measure is calculated on data point i. Each centrality measure is evaluated against the average final opinion of the network on a set of 22 networks. These networks are selected such that different values of ξ -their main differentiating characteristic- are present in the test set, so that the results are more representative for the complete set of networks. Recall that ξ is the inherent directionality of the network, as introduced in section 2.2.1. In Figure 4.1 this division of the selected networks over the possible values for ξ can be seen.

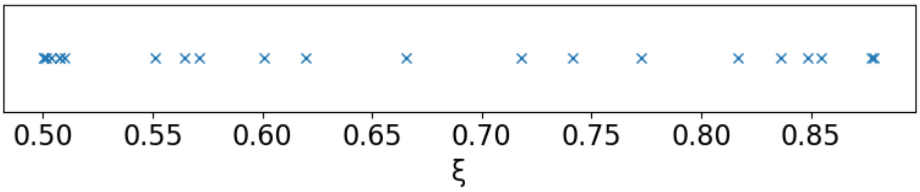


Figure 4.1: Division of selected networks across ξ values

Closeness centrality

Closeness centrality is a measure originating from graph theory. It is designed to provide an indication of how far the other nodes in the network are on average from the node for which it is calculated. This makes it a popular way of identifying central, important nodes in a graph and also a good first candidate to identify opinion leaders in the scenarios examined in this thesis. It is defined as follows in the literature:

Definition. Closeness centrality
In a connected graph, closeness centrality (or closeness) of a node is a measure of centrality based on the connections of this node. It is high when the node has a short average distance to all other nodes in the network. [52, 4].

Based on the way the closeness centrality of a node is calculated, a higher closeness centrality should indicate that the node has a more central position in the network as the distance to other nodes is shorter than for a node with a lower closeness centrality. Since the closeness centrality of a node is generally dependent on the amount of nodes present in the network, the value is usually normalized, leading to the following formula

$$C_i = \frac{N}{\sum_{j=1, j\neq i}^{N} d(i,j)}$$

for each node i [52]. In this formula, N stands for the total amount of nodes in the network and $d(i,j)$ for the shortest distance between node i and j. This normalization enables comparison of the values across different networks. Trying to extend this interpretation into opinion dynamics, the assumption could be made that a node with high closeness centrality should influence the network as a whole more quickly. The main questions are thus if the

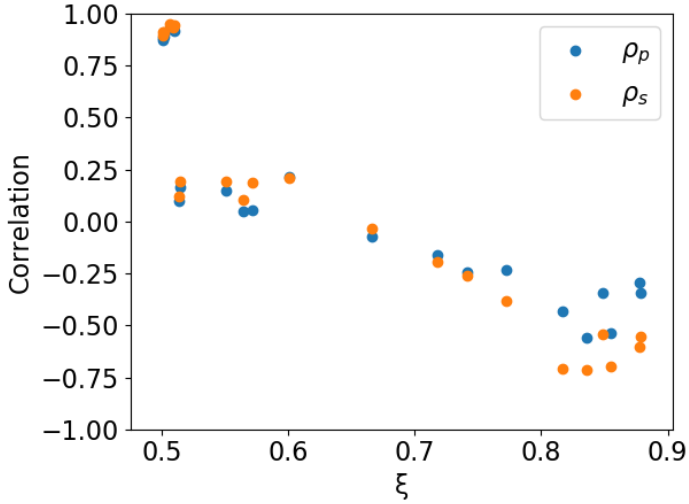


Figure 4.2: Correlation between average final opinion and closeness centrality

assumption that influencing all other nodes happens through the shortest distance paths is valid and if the speed of this influencing is the only factor needed to determine whether a node will have a high or low influence on its peers.

In Figure 4.2 the results of the evaluation of the closeness centrality as a replacement for the simulations are plotted. The inherent directionality ξ is plotted on the x-axis to be able to assess whether the correlation is high enough for the selected set of test networks, as indicated in Figure 4.1. The correlation is plotted on the y-axis. In the figure, the values obtained for ρ_p are shown. It is clear that while the correlation is quite positive for networks with a low directionality, this does not hold when checking networks with a higher ξ. Plotting the rank correlation as well, shows that the results are not heavily influenced by outliers. This leads to the conclusion that this measure could hold up as an approximation of the influence of a node in networks where no significant hierarchy is present. In more hierarchical networks, it cannot be used to replace the simulation results. When looking for an explanation as to why this measure does not correspond as well with the simulation results in highly directional networks, a possible answer could be found in the way these networks are built up. Because these networks are built up from a combination of mainly feed-forward loops, there is a chance that certain nodes are not

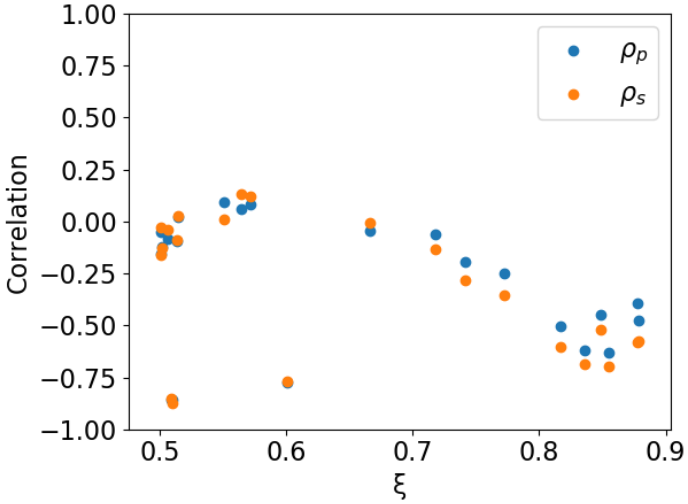


Figure 4.3: Correlation between simulation results and node level

able to reach each other. In the closeness centrality calculation, these combinations are left out of the equation and therefore they do not really have an effect on the result of the calculation. This is in sharp contrast with the result for the average opinion in the network when running the simulations, as the nodes that are not reachable will keep their initial opinion of 0 and lower the average final opinion in the network.

Node level

A logical example of a measure that has values that are strongly correlated to ξ is the node level. The calculation of this level has been presented in section 2.2.1 as a step in the calculation of the inherent directionality of the network. When the network is more directional in nature, it would seem reasonable to assume that a node that has a higher level can influence more nodes than a node further down in the hierarchy. The question here is thus how high this correlation between the node level and the simulation output -being the average final opinion- is.

As Figure 4.3 shows, this question seems to have a clear answer: the level of a node does not have a high correlation with the results from the simulations. Both the Pearson and Spearman correlation coefficients show that in most cases, especially the ones with a high ξ, the correlation is centered around 0 and whenever it does have a higher absolute value it is negative. Judging by these results, it is clear that the node level cannot be used as an approximation of the influence a node has on the rest of the network. A reason for this result could be that, because the used triadic graphs are constructed from feedback and feed-forward loops, the distance between most nodes is not that high. Because of this, the node levels can lie very close together, while the simulation results may show more variation. A possible explanation for the negative values is found in the algorithm that calculates ξ. The algorithm gives basal nodes, characterized by an in-degree of 0, a level of 0 [13]. The other nodes are then given a higher value for their level. When looking at Figure 2.1a as an example, this means that the left node would get a level of 0 and the node levels increase when moving to the right in the graph. Levels calculated in this way are also called trophic levels. This name stems from the study of animal food chains [13]. This result for node level might seem counter-intuitive, because the nodes that are highest in the hierarchy get the lowest level. However, this does not make a difference for the calculation of ξ, as only the fractions of edges going in either direction matter. Reversing the node levels to make them more intuitive would therefore give the exact same result. However, it does give an explanation for the anti-correlation seen in Figure 4.3.

PageRank

Because the main characteristic of a node related to the hierarchy present in the network does not offer a solution to the problem that closeness centrality does not perform well in networks with high directionality, the search continued for a measure that performs well in all cases. Another popular way of ranking the nodes in a network was devised in 1998 in the form of the PageRank algorithm [46]. The original purpose of this algorithm was to web pages for search engines. Because of its success, the algorithm has gained a lot of popularity. Since its inception, it has been applied in other areas as well, including opinion formation [22, 29]. The definition of the algorithm goes as follows:

Definition. PageRank
PageRank is a centrality measure that calculates a score for the nodes in a graph using the structure of the incoming links. The underlying assumption is that more important nodes

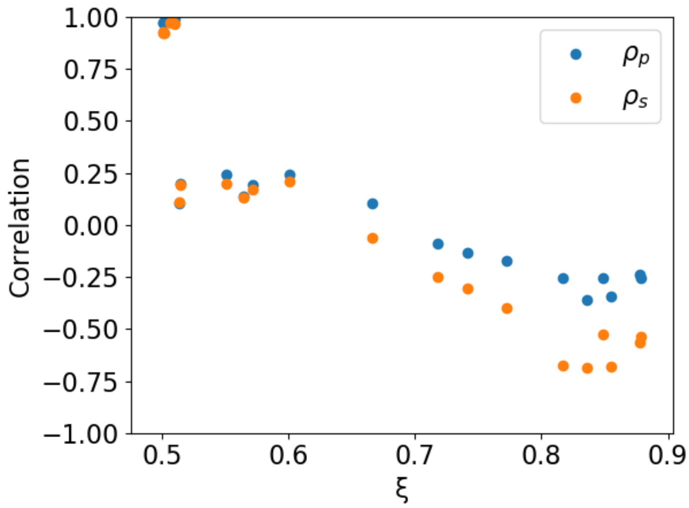


Figure 4.4: Correlation between simulation results and PageRank

are likely to have more incoming links [46].

Basing the ranking of the nodes on the incoming links makes sense when ranking web pages, the question is whether this also proves to be a good strategy for opinion leadership. Keeping in mind the fact that more hierarchical networks are built up from a higher share of feed-forward loops, the expectation is that for these networks PageRank might have trouble ranking the nodes correctly. The reason behind this statement is that in feed-forward loops the nodes with a higher in-degree reach fewer nodes than the others. As can be seen in Figure 4.4, this assumption appears to be confirmed by the results. Similar to the closeness centrality, PageRank performs very well in graphs with a low ξ value, but falls off when the inherent directionality gets higher. When looking at the networks with a high value for ξ, the correlation is definitely an indication that interchanging the simulation results with the results from the PageRank calculation is not a useful approach.

Hyperlink-Induced Topic Search (HITS)

Having studied a measure based on the structure of the incoming links, the next logical step was to study one based on the outgoing links. HITS analysis, or hubs and authorities analysis as it is often referred to, is another algorithm used to rank the nodes in a

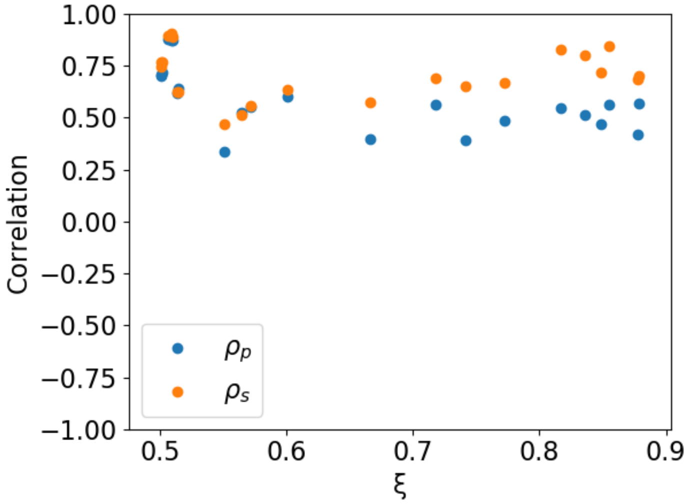


Figure 4.5: Correlation between simulation results and HITS

graph [33]. Similar to PageRank, it was originally constructed as a way to rank web pages. As the name suggests, it considers two ways in which a node can be important: as a hub or as an authority. Authorities are nodes that are assumed to have some form of inherent importance and hubs are nodes that have importance because they have information on where these authorities can be found [45]. Thus each node gets not one but two centrality scores during the analysis. A node with a high authority centrality can be characterized by the fact that it is pointed to by a lot of nodes with a high hub centrality. A high hub centrality is achieved by pointing to a lot of nodes with a high authority centrality on the other hand. The calculation of these centrality values is done using only the links of the network as input. Similar to PageRank, it works for directional networks only. Because the current goal is to identify a measure that correlates well with the influence a node has on the network, the centrality score used here is the hub centrality. As a hub is characterized by having connections pointing towards high influence nodes, the conjecture is made that it could correlate well with the values obtained trough the simulations.

When calculating and comparing these hub centrality scores to the output of the simulations, it can be seen that the correlation is indeed high. As Figure 4.5 shows, the effectiveness of this measure also holds up for networks with a higher ξ, which is in sharp

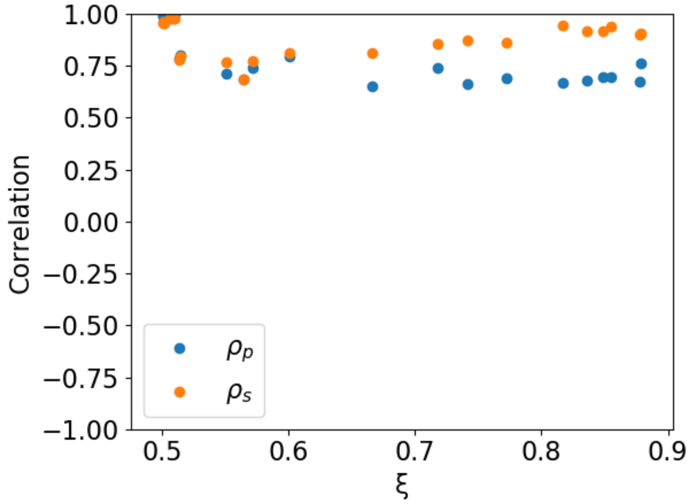


Figure 4.6: Correlation between simulation results and node out-degree

contrast with the measures tested out so far.

Node out-degree

Based on how well the hub centrality score performed when comparing it with the simulation results, another measure based on the structure of the outgoing links was tested. This measure is the out-degree of the nodes. The hub centrality of a node increases with increasing out-degree, but the difference is that the out-degree does not make an extra distinction based on which node is pointed towards. This makes it easy to understand conceptually and also very fast to calculate. Generally a higher out-degree should mean that the node is a better at spreading its own opinion, as it influences a larger amount of nodes faster. The question is whether this simple characteristic of the node is able to give a good indication of opinion leadership, as measured through the simulations.

Figure 4.6 shows that this characteristic indeed seems to be a very promising measure. Node out-degree correlates well with the simulation results across all values for the inherent directionality, similar to the hubs and authorities analysis. Therefore, it can be concluded that the influence of a node increases together with its out-degree.

Selection of a centrality measure

Based on the results obtained from the simulations and the calculation of the metrics on the networks, two conclusions on node leadership can be drawn:

1. A single node is usually not enough to have a significant impact on the network if it is not stubborn.
2. Opinion leadership is strongly correlated with the number of connections a node has to others. The fewer steps it takes to reach a high number of nodes, the higher the influence.

The first conclusion was derived from the observation that when a focal node in is not explicitly made stubborn, it has little to no effect on the group opinion. Instead the opinion of this node converges to 0 very fast, leading to all nodes having an opinion of 0. This suggests that zealots have a higher impact on an initially unbiased network when using the DeGroot model, compared to nodes that update their own beliefs.

The second conclusion is drawn from the results of the comparison between the different measures and the output of the simulations. An interesting remark here is that the assumption that the ρ_p values would be affected by outliers, seems to be unfounded in this case. ρ_p and ρ_s show similar results for each of the measures investigated. Of these measures, the node level, PageRank score and closeness centrality have a correlation value close to zero in a large portion of the test cases. For the node level, this indicates that, although there may be a strong hierarchy present in the network, the node level by itself is not a sufficient characteristic for opinion leadership in all networks. Compared to the node level, the PageRank algorithm and closeness centrality both scored similar on average. These measures have a high correlation with the results of the simulations that decreases very quickly when ξ increases. This problem is not encountered when evaluating the HITS analysis and node out-degree as possible indicators for node leadership. Both of these show similar correlation values across all tested networks. Because the results are comparable and the performance of node out-degree as a benchmark for opinion influence is intuitively more straightforward, it was decided to use this as an indicator of opinion leadership.

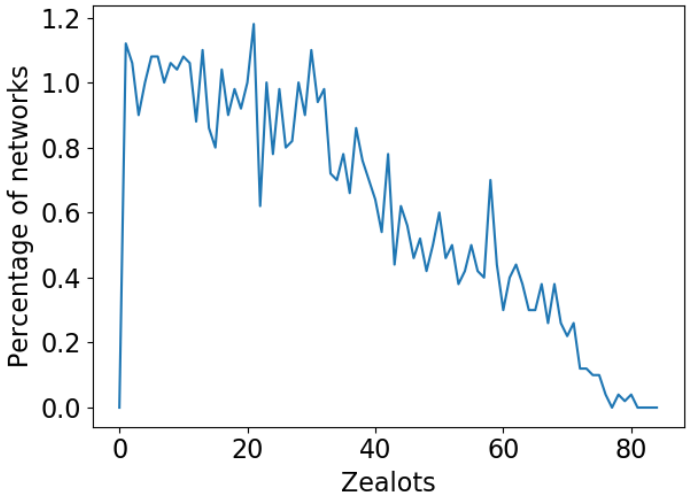
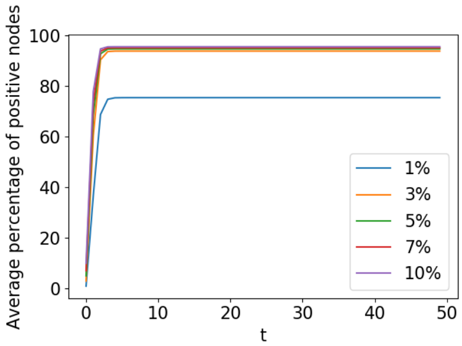


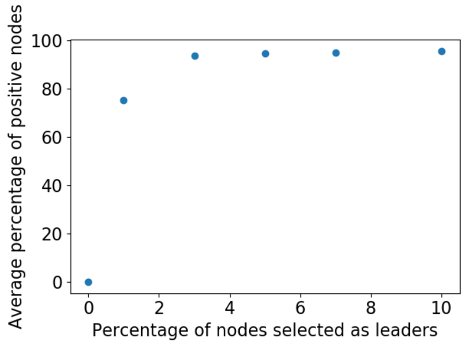
Figure 4.7: Number of zealots in the networks

4.2 Influence using opinion leaders

Now that a measure has been found to identify the opinion leaders in the network, the next step is to use these leaders to try to change the opinion dynamics in the network as a whole. The way the opinion leaders are used to achieve this goal, is by giving them an opinion value of 1 during the initialization of the experiment. The opinion leaders also update their opinion according to the DeGroot model, similar to all other nodes in the network. The goal of each experiment is to have the network reach a consensus on the positive side of the opinion space, which is limited by -1 on one side and by 1 on the other side. The experiment is considered to be a success if by the end 90% of the nodes are on the same side as the leader nodes were in the beginning. The reason behind using a percentage of the nodes instead of all of them, is that in a subset of the networks, there are zealots present in the form of nodes with zero in-degree. These nodes do not change their opinion. It is therefore not in every case possible to have the leader nodes influence all other nodes in the network. In Figure 4.7 the number of zealots is plotted versus the percentage of networks where this occurs. As can be seen in Figure 4.7, the number of zealots is usually small, except in a tiny percentage of the networks.



(a) Number of positive nodes versus time



(b) Effect of number of opinion leaders

Figure 4.8: Evolution of the fraction of positive opinions in an initially unopinionated network

4.2.1 Unopinionated environment

The first conducted experiment had the goal to see whether a small selection of nodes actually has the capacity to get all other nodes on their side of the argument when this set of nodes is carefully selected. The idea for this experiment came from the conclusion that a single node does not have the power to change the outcome of the opinion dynamics process in the network when it is not stubborn. As mentioned above, the selected nodes are given an opinion value of 1. In this experiment, the nodes that are not selected as leaders have an opinion value of 0 and are thus considered to be unopinionated. This experiment has many similarities with the one used to measure a node's influence presented in section 4.1.2. The main difference between these experiments is the fact that there are more nodes that are given an opinion value of 1 and that they are selected based on the out-degree (which is indicative of opinion leadership, as discussed in section 4.1.2). These opinion leaders are also no longer explicitly made stubborn, but instead update their own opinion using the DeGroot model based on the structure of the network. The question is thus if this group of nodes with a high opinion leadership value is able to influence the rest of the network or whether they are influenced by the rest of the network instead.

To investigate this, the simulations described in the previous paragraph are run. To check whether the rest of the nodes in the network update their opinion values to a positive value, the average number of nodes with a positive opinion is plotted versus the simulation step. This number is averaged across all 5000 networks and thus also across all values of the inherent directionality ξ. The plot in question is shown in Figure 4.8a and shows that carefully selecting the nodes does enable the combination of them to convince the rest of the network to make a change to their opinion value. On average, almost 80% of the nodes have a positive opinion by the end and this in the most prudent case. This influencing also happens very fast, as can be seen by the very steep initial rise in nodes with a positive opinion. The experiment has been carried out with different numbers of nodes selected as leaders. Because this selection is based on the out-degree, all nodes that are among the 1% of nodes with highest out-degree, are also within the highest 3% and so on. For each of these percentages, a different line is drawn with a different color, as indicated in the legend. In all cases, it is clear that the group opinion converges very quickly. When 1% of nodes are selected as leaders, the number of cases where the majority of nodes ends up with a positive opinion is significantly lower than for the scenarios where more nodes are selected as leaders. Figure 4.8b plots the final percentage of nodes with a positive opinion versus the percentage of nodes that have been selected as leaders. Judging from this figure, increasing the amount of nodes selected as opinion leaders also comes with an increase of the average number of nodes with a positive opinion in the final step of the simulation. The fact that there seems to be a limit on the percentage of positively opinionated nodes that is lower than 100%, can be explained by the presence of zealots in some networks. These do not adopt a positive opinion when they are not selected as a leader node. This could explain why the average percentage of positive nodes is lower in the case where only 1% of the population is selected as leaders. This 1% corresponds to 3 of the 343 nodes. When increasing the percentage of nodes selected as leaders, it is assumed that more of the networks zealots will be present in the selection. This is because the data shows that the correlation between the node in-degree and out-degree is very weak. This has a double effect on the result. They help spreading the positive opinions, similar to the other leader nodes, but adding a zealot to the selection also means that one fewer zealot will be constantly counteracting the efforts of the leader nodes in influencing the rest of the network.

While Figure 4.8 shows that the number of nodes with a positive value for their opinion generally increases fast and to a high percentage of the total nodes, it does not give an indication on how positive the resulting value actually is. As the DeGroot model constantly takes averages of opinions and in this experiment a fairly low amount of nodes is selected as opinion leader, the expectation was that the resulting average opinion would not be very high. Especially in the case where only 3 nodes are given an initially positive opinion, this is expected. It would however be interesting to see how this average final opinion moves when the number of nodes selected as leaders increases. Figure 4.9 gives an indication of this movement. As expected, the average final opinion is low when only 1% of nodes are selected as leaders. There does appear to be a correlation between the two percentages. When the percentage of nodes used as opinion leaders increases, the average final opinion tends to increase as well. Keeping in mind that opinion values are bounded by 1, these averages are not negligible compared to the fraction of nodes that is opinionated. The standard deviation also increases with an increase in the amount of leader nodes, indicating that there is also a higher variation in the final opinion. Aside from this increase in standard deviation, the trend of the mean value is still upwards. The conclusion can thus be drawn that when the rest of the network is initially unopinionated, increasing the amount of nodes that agree on an initial opinion generally leads to both an increase in the number of nodes that move towards this opinion and an increase in the extend to which they move their opinion.

4.2.2 Environment with uniformly distributed opinions

While testing the effect the opinion leaders have on an unopinionated network is a good starting point, this scenario does not represent the majority of real cases. Usually individuals already have an opinion on the matter, which is unlikely to be as extreme as -1 or 1. The next experiment was designed with this in mind. In this scenario, every node is given a random opinion distributed uniformly between -1 and 1, except the nodes selected as leaders which still get an opinion value of 1. The goal of this scenario is to investigate whether the influence of the leaders shows similar trends as in the scenario where the rest of the network is unopinionated. In the latter case, nodes only had the option of either staying unopinionated or having a positive opinion. However, now there is the possibil-

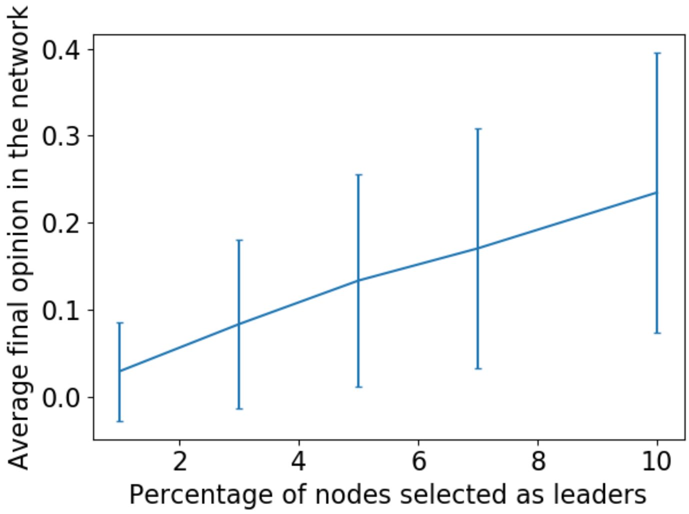
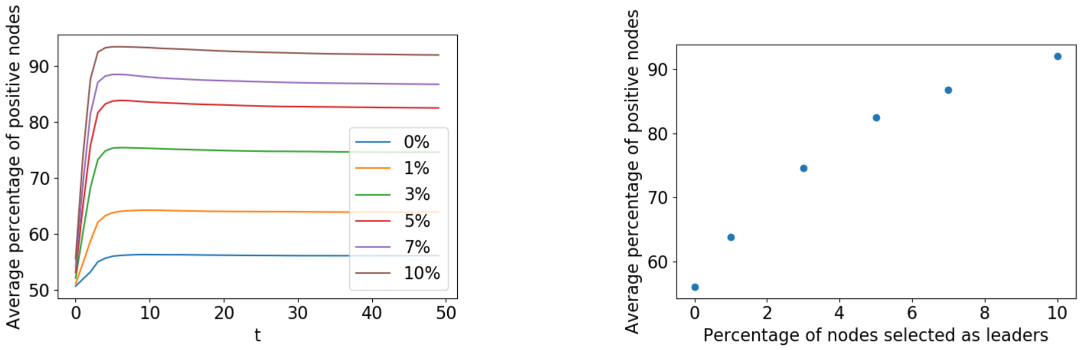


Figure 4.9: Effect of number of leaders on final opinion in the network

ity that the network reaches a consensus on a negative opinion. Since the other nodes have both positive and negative values for their opinion initially, it remained to be seen whether the non-leader nodes cancel out each other's opinion or whether the combination of leader nodes with a very positive opinion and other nodes that randomly have a positive opinion actually provides a boost to the leaders, helping them to convince other nodes faster. Based on the results of the experiment done in section 4.2.1, the expectation is that increasing the number of leader nodes will again have a positive effect on the average number of positive nodes in the network. However, this difference is presumed to be less extreme, as the leaders now have to cope with the fact that on average half the network is initially spreading a negative opinion.

Figure 4.10 presents the average percentage of positive nodes throughout the simulation and the effect the number of opinion leaders has on this percentage, as it was done for the previous scenario. Because of the uniformly distributed initial opinions, the network starts with approximately 50% of the nodes having a positive opinion on average, not counting the leader nodes. The experiment is also conducted with no nodes selected as leaders to have a reference case. When the mean outcome of this reference case is calculated over all 5000 networks, the result is that on average a little more than half the nodes have



(a) Number of positive nodes versus time

(b) Effect of number of opinion leaders

Figure 4.10: Evolution of the fraction of positive opinions in an initially uniformly distributed network

a positive opinion. This should be interpreted as the network reaching a consensus on a negative opinion 50% of the time and a consensus on a positive opinion in the other cases. The fact that it is not exactly 50% is attributed to random chance. As can be seen in Figure 4.10a, when nodes are selected as leaders, this percentage increases with the percentage of nodes selected as leaders, similar to the scenario where non-leader nodes were initially unopinionated. Figure 4.10b plots the density of positive final opinion. While the increase in average percentage of nodes with a positive opinion was very large in Figure 4.8b, it is more moderate here. The graph also shows that while adding more nodes to the selection of leaders will increase the average percentage of positive nodes, the effectiveness drops with the amount of leader nodes. As there are more cases where no consensus on a positive opinion is reached, it might be interesting to investigate in greater detail any cases where this does not happen.

Evaluation of unsuccessful experiments

As mentioned in the introduction of section 4.2, an experiment is considered to be unsuccessful if in the end less than 90% of the individuals have a positive opinion value. In the more realistic case of starting out with a uniformly distributed opinion space, there are much more simulations where this occurs, as indicated by the difference in percentages between Figure 4.8b and 4.10b. Therefore, further investigation of the experiments is needed

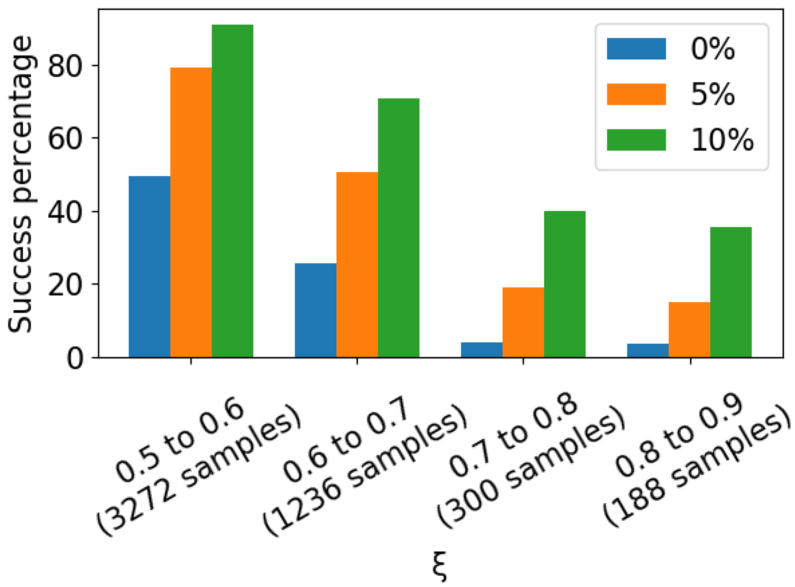
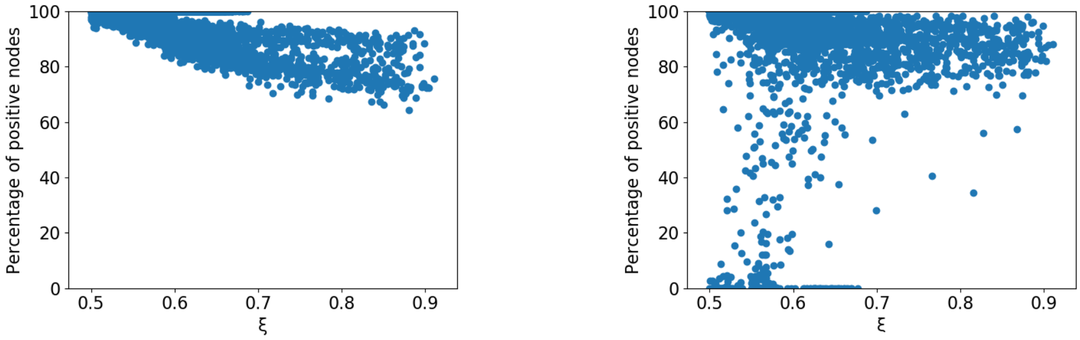


Figure 4.11: Relation between ξ and simulation success

to find out what might happen in these unsuccessful simulations.

In Figure 4.11 the success rate of the simulations is plotted against the inherent directionality ξ for multiple percentages of nodes selected as leaders. In each of the cases, there is a clear downward trend. On the x-axis the number of samples over which the average is taken is indicated. Even though there are more network samples with a low ξ value, there are still 188 instances with an inherent directionality value between 0.8 and 0.9 which allows to draw statistically relevant conclusions. The results for networks with a ξ value above 0.9 have been omitted, as their number is too low to get meaningful results. These outcomes indicate that networks in which a significant level of hierarchy is present are harder to fully convince. A possible reason might be the combination of the fact that these networks have a higher chance of containing zealots and the fact that the majority of the edges go along the same direction in the hierarchy. When a zealot is present at some level in the hierarchy, the fact that most nodes point along the same direction will cause that all nodes beneath this zealot are harder to influence. The conclusion can thus be drawn that the presence of zealots, who can keep counteracting the influence of the selected opin-



(a) Network initially unopinionated

(b) Uniformly distributed initial opinions

Figure 4.12: Simulation outcomes when 10% of nodes are selected as leaders

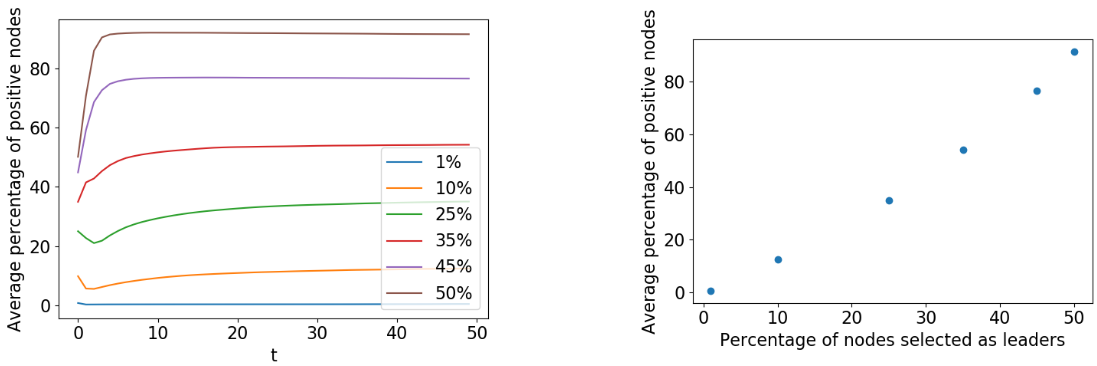
ion leaders, presents an obstacle for changing the opinion dynamics using this set of leaders.

To further analyze the outcome of these simulations, the individual results of the simulations with 10% of the nodes selected as leaders are given in Figure 4.12 for both the case where the nodes are initially unopinionated and the case where the opinions are distributed uniformly. As Figure 4.12a shows, in the unopinionated scenario the majority of the nodes has a positive opinion value in all the networks. There is however still a decline in the percentage with increasing ξ value. When the inherent directionality is low, a fraction of the simulations reach a state where 100% of the nodes have a positive opinion. This state is never reached for networks with a higher ξ. This confirms the decline in the percentage for networks with a higher ξ value, as seen in Figure 4.11. This decline is still present in the graph on the right. While the bulk of the cases still shows a clear majority of nodes having a positive opinion, there are also other outcomes present. A percentage of the simulations reaches a consensus on a negative opinion value, as is indicated by the number of dots at the bottom of the graph. This is to be expected because of the randomization of the opinions. In some cases, the majority of nodes has a negative opinion from the start which makes convincing the rest of the network significantly harder for the opinion leaders. This scenario is further explored in section 4.2.3. More interesting are the cases where the final percentage of positive nodes is between 25% and 75%. In these cases, the number of nodes not having a positive opinion cannot be solely caused by zealots or a large number of nodes with an initially negative opinion. Therefore, a different explanation may be plausible. Inspecting the data more closely shows that, for the experiment which is plotted in Figure 4.12b, 82 simulations end with both more than a 100 nodes with a positive opinion and more than a 100 nodes with a negative opinion. These simulations thus reach a polarized final state. In this polarized state there is a large concentration of nodes on the two opposing sides. These concentrations are also called echo chambers, suggesting that the nodes within an echo chamber reinforce their own opinions, while not being influenced by the others.

4.2.3 Environment with negative opinions

Now that two scenarios have been explored in which the rest of the network is more or less neutral initially, either because the individuals are unopinionated or because the opinions were uniformly distributed, another interesting case would be to see what happens when the rest of the network is strongly inclined towards a certain opinion. Suppose the network already has a consensus on a positive opinion nothing too interesting would happen. Selecting opinion leaders and giving them an extreme positive opinion would only reinforce this consensus. Having the network start with a consensus on a negative value instead, is more interesting. As a test of the effectiveness of identification of opinion leaders and their influence, in this scenario each node in the network is given an opinion value of -1 except the leaders, who still get an opinion of 1. The question then becomes whether careful selection of these leaders, using the measure found in section 4.1, can lead to a convergence on a positive value when less than 50% of the nodes have been selected as leaders.

Figure 4.13 shows the results of this experiment. As expected, choosing a low amount of nodes as leaders does not have a significant effect on the simulation. When 1% of individuals gets an opinion value of 1 and the other 99% have an opinion of -1, these 3 selected nodes are unable to convince the rest of the network, because there are both so many nodes to convince and they are on the complete opposite side of the opinion value spectrum. Shifting this ratio brings considerable changes to this outcome however. When 10% of the nodes are selected as opinion leaders, the experiment is expected to still be



(a) Number of positive nodes versus time (b) Effect of number of opinion leaders

Figure 4.13: Evolution of positive opinions in an initially very negative network

very likely to terminate with all nodes negative. Looking at the average over all 5000 results, over 10% of the nodes end up with a positive opinion in this case. So on average an increase in the amount of positive nodes is obtained already. Adding more leaders to the selection makes this increase higher and higher. When the network is initially perfectly balanced, with half of the nodes having a positive opinion and half of the nodes having a negative opinion, on average over 80% of the nodes end up with a positive opinion. This is an indication that the approach of carefully assessing the importance of the nodes as opinion leaders and selecting them using the results is working.

Chapter 5

Special cases

In chapter 4, a measure was identified to assess opinion leadership and using this measure different sized groups of leaders were set up. Using these groups, the influence of the opinion leaders on the rest of the network was evaluated for different scenarios. In each of these scenarios, the standard version of the DeGroot model was employed as the fusion rule. In this chapter some special scenarios are explored. Each of these cases makes a small modification to the general DeGroot model. The goal of these adaptations is to explore certain extra parameters or assumptions that could be argued to have an influence on the opinion dynamics in a social network setting. In each of these experiments, the rest of the network gets uniformly distributed random opinions between -1 and 1, except for the scenario described in section 5.4.

5.1 Self-confidence

In the original DeGroot model, each node uses the average opinion of its neighbors as its updated opinion. As this version has not been refuted, this is not a bad assumption to make. Even though individuals can value the opinion of others greatly, they usually also value their own opinion. This notion led to the idea of testing whether the opinion leaders still have a similar effect as in chapter 4, when self-confidence is added to the fusion rules.

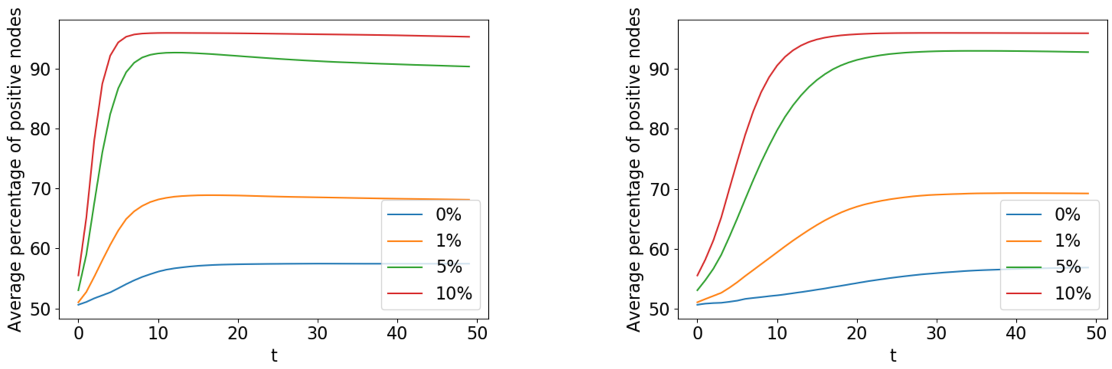
Based on the structure of the network there are already nodes that actually have self-confidence. Zealots could be seen as nodes with 100% self-confidence instead of as nodes which do not receive updates. It has already been shown during the process of trying to identify the measure for opinion leadership, that these nodes can have an impact on the network. The question is thus whether more nodes with a certain level of self-confidence will have a significant impact on the ability of opinion leaders to convince the rest of the network. In this experiment the nodes not selected as leaders have uniformly distributed random opinions between -1 and 1.

To model this adaptation of the standard DeGroot model, a new mathematical description of the fusion rules is needed. Similar to the normal version, an average of the opinions of the neighbors of a node is calculated. Subsequently, a weighted average of this average and the current opinion of the node is computed and this weighted average is then used as a new opinion. This leads to the following formula:

$$\begin{cases} o_{j,t+1} = s * o_{j,t} + (1-s) * \sum_{i=1, i\neq j}^{n} W_{ij} * A_{ij} * o_{i,t} & \text{, in-degree} > 0 \\ o_{j,t+1} = o_{j,t} & \text{, in-degree} = 0 \end{cases} \tag{5.1}$$

for the new opinion. The parameter s is a floating point number between 0 and 1 indicating the self-confidence level of the nodes. The other parameters are the same as in equation 2.2. One expected change in the results is a delay in the evolution of the system. Having self-confidence means that more importance will be given to the initialization of the system and it will therefore take longer for the effects of this initialization to fade away.

Figure 5.1 shows the evolution of the fraction of positive nodes through time. In the left graph, this evolution is shown when the nodes have a self-confidence s of 0.5, in the right graph the self-confidence is 0.8. The experiment is again done for multiple percentages of nodes selected as leaders. When comparing the two graphs, it is clear that the self-confidence adds a delay to the opinion formation in the network. What is also apparent, is that there is little difference in the end results between the two cases. This is confirmed by Figure 5.2, which plots the final average number of positive nodes versus the number of nodes selected as opinion leaders. In this graph, the results for the same experiment with no explicit self-confidence are also added. The figure shows that there is indeed little



(a) 50% self-confidence

(b) 80% self-confidence

Figure 5.1: Effect of self-confidence on opinion dynamics

to no difference between the case with s equal to 0.5 and the case with 0.8. Judging from the graph, adding self-confidence actually increases the percentage of positive nodes in the network. This seems surprising at first, because adding self-confidence should mean that nodes are harder to convince, as is confirmed by the slowdown in changes. However, self-confidence also implies that the leader nodes keep their strong opinions longer. However, because these initial opinions of leader nodes are so strong compared to the random opinions of the nodes the leader nodes have more time to convey their strong opinions and

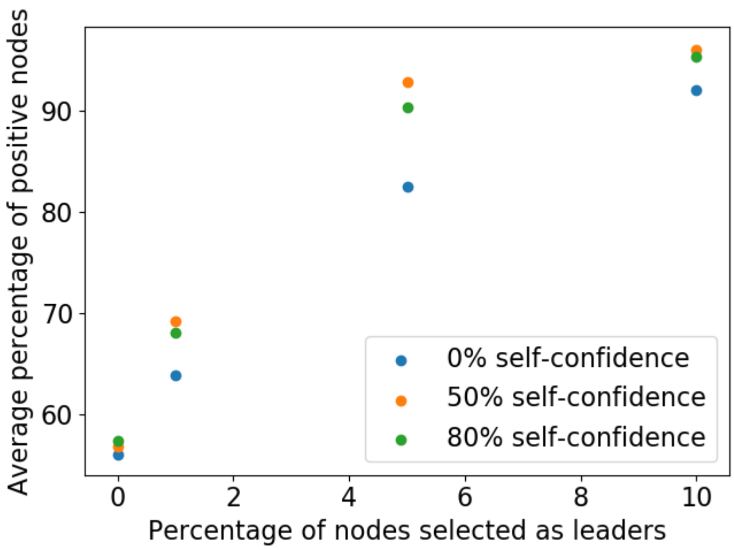
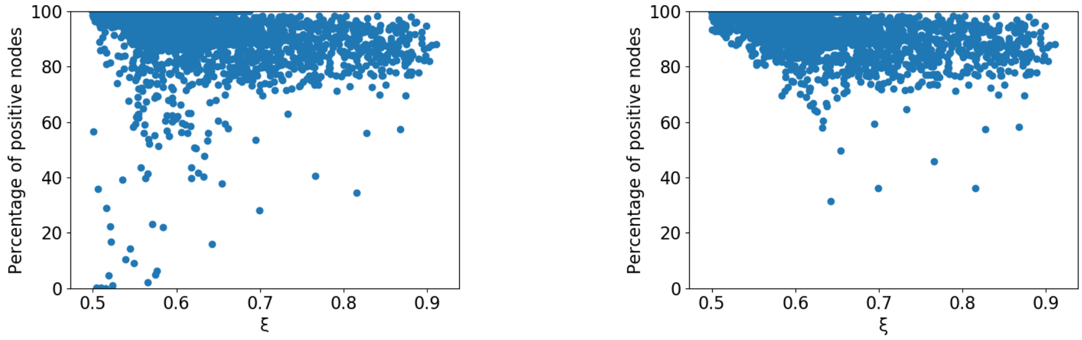


Figure 5.2: Effect of number of opinion leaders



(a) 50% self-confidence

(b) 80% self-confidence

Figure 5.3: Simulation outcomes when 10% of nodes are selected as leaders

thus have an increased effect on the network.

Figure 5.3 shows the exact outcomes of the simulations for the instances where 10% of the nodes have been selected as opinion leaders. Comparing these outcomes to the ones shown in Figure 4.12b, shows that there are fewer cases where the network has no nodes with a positive opinion in the end. A noteworthy difference between the left and the right graph is that while the self-confidence increases there are fewer networks that end up in a polarized state. For the scenario where nodes have a self-confidence s of 0.8, only 24 of the 5000 simulations end up with more than 100 nodes on either side. This is a 71% decrease from the 82 simulations where the nodes did not have an explicit level of self-confidence and is again an indication that the self-confidence of the leaders helps them convince the rest of the network.

5.2 Increasing importance of leaders

A second adaptation made to the original model is based on the assumption that people pay more attention to the opinion of leader figures. In many cases these leaders speak from a position of authority they attained because of their credibility in the matter. In this network-based approach for the simulations, there are two possible ways to change the DeGroot model with the goal of giving more importance to the opinion of the leaders:

1. Base weights on hierarchy
2. Base weights on leadership measure

In sections 5.2.1 and 5.2.2 these scenarios are further explored.

5.2.1 Weight based on hierarchy

In this scenario, the weights used in the DeGroot model are based on the direction of the edges in the hierarchy. When a node has an incoming edge that points downward in the hierarchy, this means that the source node has a higher position in this hierarchy. Therefore more attention should be given to his opinion. Inversely, when the incoming connection is pointing upwards in the hierarchy, less importance should be attached to the opinion of the source node. Enforcing these different weights in the fusion rules is done as follows:

1. Increase the number of perceived incoming edges by adding a number of non-existing edges f
2. The weight w_d used for the connections coming from nodes lower in the hierarchy is then calculated as $w_d=\frac{1}{d+u+f}$. Where d stands for the number of connections coming from a node downwards in the hierarchy and u for the number of connections for which the source node is higher in the hierarchy.
3. The weight w_u used for the connections coming from nodes that have a higher position in the hierarchy equals $w_u=w_d + \frac{f}{u} * w_d$

Using these formulas, more importance will be attached to opinions of nodes that are higher in the hierarchy. It can also be confirmed that the sum of all weights is one by calculating it:

$$\begin{aligned} u * w_u + d * w_d &= u * (w_d + \frac{f}{u} * w_d) + d * w_d \\ &= w_d * (u + f + d) \\ &= \frac{1}{d+u+f} * (u + f + d) \\ &= 1 \end{aligned}$$

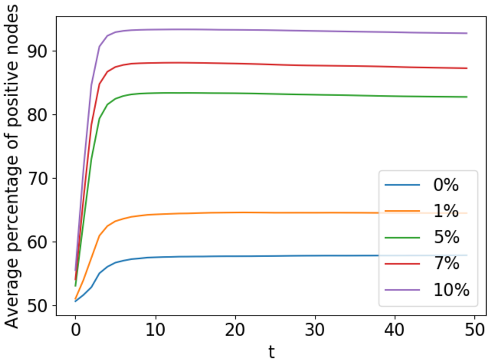
How much more importance is given to nodes more upwards in the hierarchy, can be controlled by changing the parameter f. In the experiments done here, f is set as $\lceil \frac{d}{u} \rceil$. The reasoning behind using this instead of a constant number, is to make sure that a node pays extra attention to its neighbors with a higher level when this node has few of those neighbors. If all incoming connections have the same direction, no difference in weighting is applied. The assumption is that, because more attention is paid to nodes higher in the hierarchy, the network will converge faster.

Figure 5.4 shows the development of the average fraction of positive nodes throughout the simulations. Adjusting the weights based on the direction of the connection in the hierarchy seems to not have a big influence on the final outcome of the experiment compared to the case where no difference in weighting is applied shown in Figure 4.10. Also the speed at which nodes are convinced is not significantly affected, judging by the similar initial increase in percentage. Checking whether the inherent directionality affects the success rate also shows similar results as in the case where equal weights are used, because there is still a decline in success rate with increasing ξ. The reason for these minor differences could be that the leader selection is not based on the position in the hierarchy. Changing the leader selection to take this into account, with the goal of improving the success percentage in more hierarchical networks, is unlikely to work. This is because the results in section 4.1.2 did not show a high correlation between node level and influence.

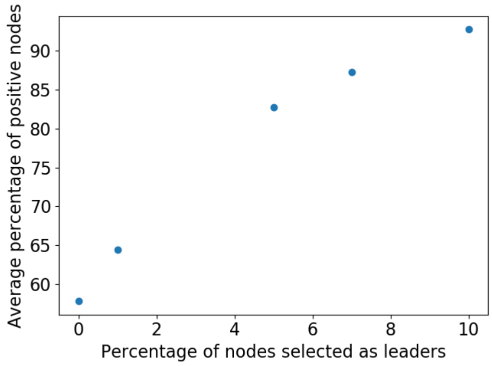
5.2.2 Weight based on out-degree

Another way to increase the importance of possible leadership figures is to use the leadership measure identified in section 4.1.2 as a reference to base weight on. In this scenario, each node checks the out-degree of all its neighbors when initializing the weights. Because the network is static, this only has to be done once. To make sure the sum of the weights remains one, these out-degrees are normalized leading to the following formula

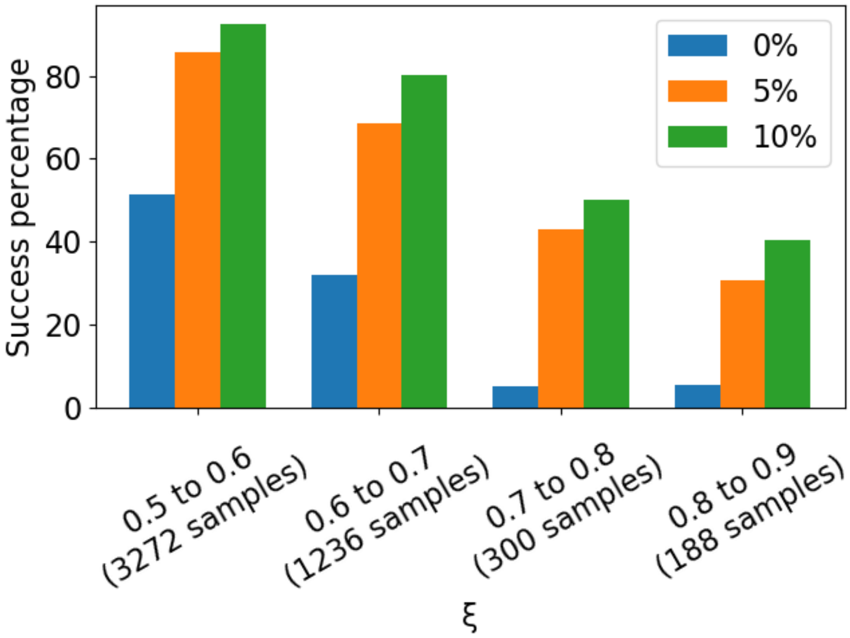
$$w_i = \frac{d_i}{\sum_j d_j}$$



(a) Number of positive nodes versus time



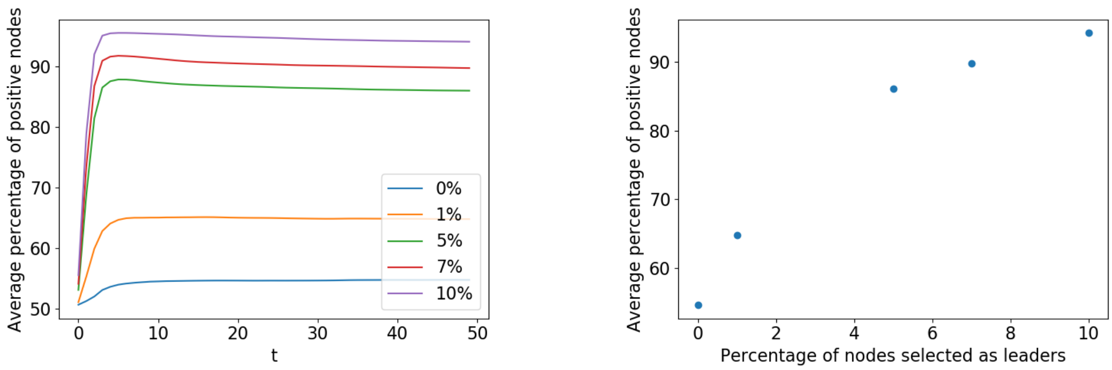
(b) Effect of number of opinion leaders



(c) Relation between ξ and simulation success

Figure 5.4: Effect of hierarchy-based weighting on opinion dynamics

for the weight used for the opinion of node j's neighbor i. In this formula d_i indicates the out-degree of neighbor i and the sum is taken over all neighbors of node j. The expected result of using this weighted average as a new opinion, is an increase in the speed at which the leader nodes convince the rest of the network. Because the weight is based on the same measure as the one used to select leader nodes, more attention will be given to them. Therefore they should have a stronger ability to convince other nodes.



(a) Number of positive nodes versus time (b) Effect of number of opinion leaders

Figure 5.5: Effect of out-degree-based weighting on opinion dynamics

Analyzing the results of the experiment shown in Figure 5.5 shows that the speed at which the network changes opinion is indeed very high. An interesting behavior that can be seen in Figure 5.5a, is the slight drop in the average number of positive nodes after the first few steps. A cause for this might be the presence of zealots in the networks. After the first couple of steps, the effect of the increased weight put on the opinion of the leader nodes decreases. This is because these leader nodes also take part in the opinion formation process and will start spreading opinions closer to the average opinion of the network. Because the weights are based on the out-degree, they are however more effective in the first steps of the simulation. Therefore, their initial influence might spread further than in the case where all weights are uniform. Nodes that change opinion because of this extra reach will not have large absolute values for their opinion and can be converted again more easily as well. In the end, the average number of positive nodes in the simulations is slightly higher than in the scenario where equal weights are used across neighbors. This indicates that even though this strategy is effective in influencing the opinion dynamics, its effects are only minor compared to adding extra leader nodes to the selection.

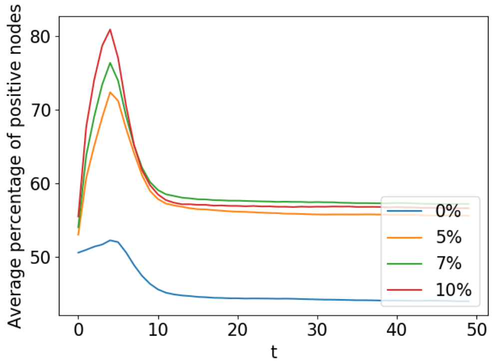
5.3 Weight based on similarity of opinion

Another adaptation is based on the intuition that many people prefer to hear their own beliefs get confirmed by others. Based on this fact, the idea was to adjust the weights used in the DeGroot model based on the opinion of the neighbors of the node. This is not a new idea, as variations of it have already been studied in the past. A well-known implementation is called the bounded confidence model, introduced earlier in section 2.3. The main difference between this model and the DeGroot model, is the fact that not all the opinions of the neighbors are taken into account [41]. Because this study uses the DeGroot model for the fusion rules, a middle ground between these two models is implemented in which all opinions are taken into account. However, the weights vary depending on how similar the opinions are between the two neighboring nodes. Because the weight needs to increase when the difference in opinion becomes smaller, the inverse of this difference is the basis on which the weights are distributed. To make sure the sum of all weights remains one, a normalization factor is needed. Taking this into account, the following formula is obtained for the weight w_i used for neighbor i of node k:

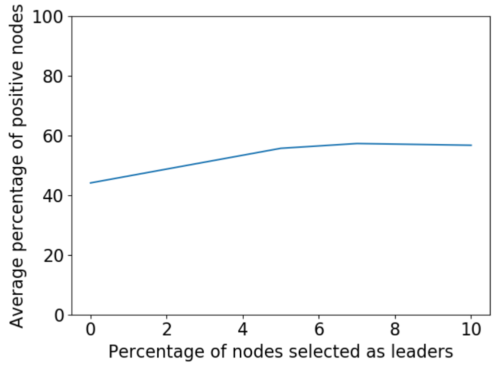
$$w_i = \frac{1}{d_i} * \frac{\prod_j d_j}{\sum_j d_j}, \; d_i = \max(0.05, |o_k - o_i|)$$

where d_i indicates the difference in opinion value between the current node and its neighbor i and o_j indicates the opinion of node j. The summation and product are taken over all neighbors of the node. To make sure that no weights of infinity are obtained by dividing by zero in case of extremely similar opinions, a minimum of 0.05 is placed on the value of d_i. This way, all individuals in the network will keep paying attention to all their neighbors, possibly using different weights.

This appears to be a significant adjustment to the standard version of the model which takes unweighted averages at every step of the simulation. In this new version, all the weights have to be calculated again at every step, before evaluating the new opinion value. As a consequence of the way this change is implemented, there is also the possibility of having large differences in weights. This can severely affect the effectiveness of the opinion leaders in changing the opinion dynamics of the network. The big problem is that these opinion leaders are given a very high initial opinion to spread their belief on that side of the argument. However, the fact that these opinions are so high might work against them in this case as this could have the effect that most nodes pay more attention to other non-leader nodes.



(a) Number of positive nodes versus time



(b) Effect of number of opinion leaders

Figure 5.6: Effect of weighting based on similarity of opinion on opinion dynamics

The results of the simulations seem to confirm that many nodes pay less attention to the opinion leaders as Figure 5.6b shows that the final average percentage of positive nodes lies around 50%. This remains the case when the number of leaders in the network is increased. The conclusion to draw from this result would be that the majority of nodes pays almost no attention to the selected nodes. This is however contradicted by Figure 5.6a. In this figure, it is clearly indicated that at some point in the simulation, there is on average a clear majority of nodes on the positive side of the argument. The question is thus what causes the fact that the majority of the nodes does not end up with a positive opinion in the end. Figure 5.7 shows the individual end results of the simulations where 10% of the nodes have been chosen as opinion leaders. It shows that in most networks there is still quite a large percentage of positive nodes present at the final step. All cases where more than 80% of the nodes in the network ended up with a positive opinion value are shown in green. The simulations where more than 80% of the nodes have a negative opinion are shown in red. In case less than 25% of the nodes have an opinion value different from zero they are plotted in orange. All cases where no 80% majority is attained on either side, but more than a quarter of the nodes have an opinion are indicated in blue. The amount of blue points is an indication of the formation of echo chambers. The fact that these echo chambers form is not necessarily a surprise, as this scenario facilitates their formation. Especially at higher inherent directionality values, the formation of these echo chambers is

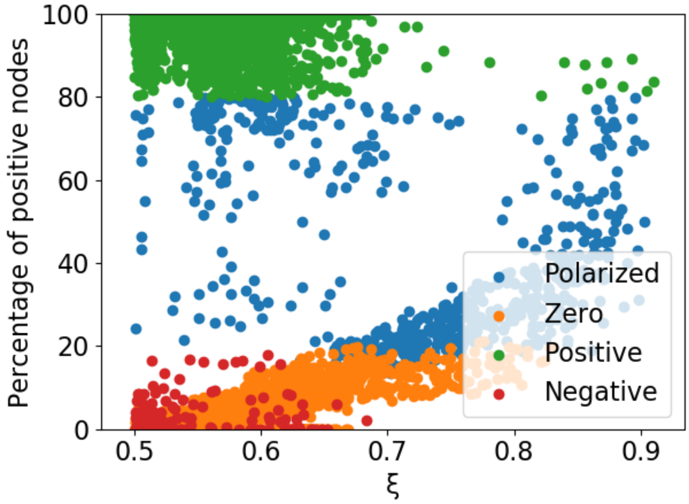


Figure 5.7: Simulation outcomes versus ξ

present. An explanation for this can be found in the fact that when ξ increases, the network resembles a tree structure more. If one node in the tree has a negative opinion, it will keep sharing this negative opinion with all nodes that are lower in the tree. Because there are very few edges going in the opposite direction, a disagreement in an upper level can cause the formation of these echo chambers. Closer inspection of the individual simulation results shows that, apart from reaching a consensus and the formation of echo chambers, the opposite extreme is also present. In this extreme, practically all nodes converge to a opinion value of approximately zero. There are almost no opinionated nodes left in the simulation and the network thus ends up in an unopinionated state. This could also be caused by the fact that this adaptation of the DeGroot model makes the nodes more inclined to agree with their neighbors who have similar opinions. When a node has an average opinion close to zero, neighbors on both sides of the argument will value this opinion more than opinions that are contradictory to their own. This increased importance of nodes which are unopinionated enables the simulations to converge on an opinion value of 0. Based on these results and the scenario itself, an arguably more effective strategy to influence the network as an opinion leader would be to have a more moderate initial opinion and slowly increase the opinion value as the simulation goes on.

5.4 Noisy opinions

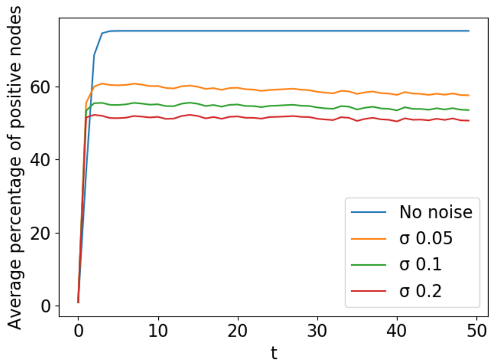
A last adaptation made to the standard DeGroot model is based on the consideration that people do not always clearly state their opinion on certain matters. The opinion expression format is an abstraction because individuals do not express their opinion as a floating point number in real life. Instead they usually express these opinions in the form of a verbal or written statement. There is not always a clear scale on which these statements can be scored to indicate the underlying opinion. Because of this vagueness, there is often room for interpretation and discussion. This observation is the inspiration for this scenario, where random noise is added to the opinions received from neighbors. As mentioned in the introduction, in this scenario the non-leader nodes are initially unopinionated. The goal is to check whether the selected set of nodes still has a similar capacity of influencing the rest of the network when there is a random factor affecting the expressed opinions.

The noise factor added to the opinion values received from the neighboring nodes is sampled from a normally distributed variable centered around zero. To evaluate the impact of this noise, the experiment is run with different values for the standard deviation σ on the noise. It is also ran with both 1% of nodes and 5% of nodes selected as opinion leaders. This way a comparison can be made on the effect of the level of noise on the opinions and the number of leader nodes present in the network. Adding noise changes the fusion rules so that the following equation is obtained:

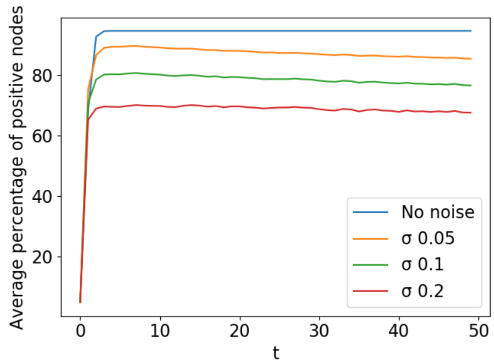
$$\begin{cases} o_{j,t+1} = \sum_{i=1,i\neq j}^{n} W_{ij} * A_{ij} * (o_{i,t} + \mathcal{N}(0,\, \sigma^2)) & \text{, in-degree} > 0 \\ o_{j,t+1} = o_{j,t} & \text{, in-degree} = 0 \end{cases} \tag{5.2}$$

Considering how effective selecting even a very small percentage of the nodes as opinion leaders was in an unopinionated environment without noise, adding noise is expected to not have a significant effect on the final results as on average this noise should cancel out. A certain slowdown in the process is expected however, because the noise introduces variation in the received opinions, which makes it possible that the leader nodes spread less extreme opinions, thus having a weaker effect.

Figure 5.8 shows the outcome of these experiments. As can be seen from the graph, the



(a) 1% of nodes selected as leaders



(b) 5% of nodes selected as leaders

Figure 5.8: Effect of noise on opinion dynamics

results do not agree with the expectations. There is no perceivable slowdown in the process. Similar to the other scenarios, adding leader nodes increases the fraction of positive nodes present in the network. A result that is clear from both figures, is that adding noise lowers the average number of positive nodes. In the case where 1% of the nodes have been selected as opinion leaders, it lowers the percentage from around 80% to around 60%. An explanation for this phenomenon can be obtained by comparing this graph to Figure 4.10a. When looking at the results with noise, a similar final result is achieved as in section 4.2.2 where there was no noise, but opinions were initially uniformly distributed. A similar comparison can be made for the graph in Figure 5.8b. Because it takes time for the opinion of the leader node to spread throughout the network, in the beginning only the noisy opinions are being spread between most nodes. This might cause the network to be close to uniformly distributed in opinion by the time the opinion of the leader node reaches most nodes, therefore increasing the difficulty for the leader nodes to influence other nodes.

Chapter 6

Future work

Although this thesis has investigated multiple scenarios in the field of opinion formation in social networks, there is a lot left to explore. Studies are still being conducted on both the sociological aspects as well as on simulation and modelling of these aspects. Because the processes impacting opinion formation are very complex, levels of abstraction need to be implemented to keep the simulations tractable. These abstractions result in generally accepted frameworks, which combine theories on opinion dynamics into mathematical equations. Since generalization of the results is an important factor in most studies, most focus is usually placed on identifying large trends in the data, as has been done in this thesis.

Based on the results of this study, one possible next step could be to investigate the leader identification measures further. Here, only one measure was deemed most correct in characterizing opinion leadership, while in reality a combination of factors could be at play. On top of this, all nodes were made stubborn in the reference simulations to ensure their impact could be measured. In the other simulations this was usually not the case. Investigating the relationship between zealotry and opinion leadership could therefore provide a finer understanding of the characteristics found in influential individuals. Another possible next step is making more assumptions on opinion formation and adapting the DeGroot model accordingly. One example of such an adaptation could be to have heterogeneous levels of self-confidence across the nodes in the network. Another possibility is investigating how the combination of several of these assumptions affects the simulation outcomes and draw conclusions based on these results. One adaptation of the DeGroot model impacted the effectiveness of the leader nodes, who try to force the network to change opinions with their extreme opinions, heavily. This adaptation is when individuals pay more attention to others with similar beliefs. Leaving the resulting model unchanged, other strategies could be explored to steer the global opinion of the network in this scenario.

Another possible route to take is making more severe changes to the fusion rules or opinion dynamics environment. In this study, the focus has been on scenarios where all individuals spread their opinion constantly and are connected only to other people. Two interesting changes to these scenarios could be to change the fusion rules so that not every node in the network makes its opinion known at every point in time or adding very well-connected mass media agents. The first possible change is inspired by the SIR model used to simulate the spread of diseases in animal populations. In this model, a node can be in one of three states being susceptible, infected or recovered. These states are also where the name of the model comes from. Translating this to opinion dynamics, it would mean that a node is either:

- not spreading an opinion but paying attention to the opinion of others
- spreading an opinion while also paying some attention to the network
- convinced of an opinion and is no longer spreading nor paying attention to others

This third option would be a direct translation of the recovered state of the original SIR model. It is however not intuitive that nodes stop both sharing and taking in opinions. Therefore an existing different version of this model could be used instead. In this version, called the SIS model, no state of immunity exists. For the case of opinion dynamics, this would mean that all nodes still pay attention to their neighbors who are spreading their opinions, but not always have to spread their own. Making a change to the opinion dynamics environment is also a possibility. With the advance of mass media in recent years, an interesting addition to this environment might be to add a mass media agent that is connected to the majority of the nodes. The effectiveness in influencing the network using this agent could then be compared to the case where such an agent is not present, but a collection of opinion leaders is used instead.

Chapter 7

Conclusion

In chapter 2 the different parts of opinion formation models have been described. For each of these parts, a decision has been made on which option to use after explaining the differences and analyzing which one fits the current goal best. Subsequently, an illustration of the experimental setup used to simulate the opinion dynamics was given in chapter 3. This setup was used to implement different scenarios with the goal of studying the influence of opinion leaders on the rest of the network. In section 4.1 an investigation into characterizing opinion leaders was discussed. From this investigation multiple conclusions were drawn. First of all, it was determined that a single node does not have the potential to make significant changes to the opinion dynamics in the network when this node participates in the opinion formation process himself using the DeGroot model. Instead, nodes had to be made stubborn, so that their effect on the remaining, unopinionated, network could be measured and compared to the results of several mathematical measures originating from graph theory. Apart from this, comparing the different measures with the simulation results showed that the most effective measure to characterize opinion leaders was the out-degree of the individual. While several other properties performed well for leadership identification, their correlation to the final average opinion usually falls off when there is a higher level of hierarchy present in the network.

Using the out-degree as a leadership measure, groups of nodes were identified and used to try to achieve a consensus on a chosen side of the argument when these nodes all partake in the opinion formation process. The results from these experiments showed that the initial opinion of the non-leader nodes in the network plays an important role in whether or not the selected set of opinion leaders will be capable to convince them. When all individuals that have not been selected have no opinion initially, only a small selection of nodes is needed to pull the rest of the network to one side. In the case where the other nodes in the network have uniformly distributed random opinions, the network as a whole becomes harder to convince and a higher number of nodes has to be selected as opinion leaders to achieve comparable results. As a definitive test of the success in selecting important nodes, an extra experiment was done. In this experiment, every node that is not selected as an opinion leader starts with a negative opinion, while the leader nodes try to spread a positive opinion. In this scenario, the out-degree proved to be an effective measure for opinion leadership. Normally, one would expect the outcome of the experiment to be equally divided between converging on a positive or a negative opinion when the initial distribution of these opinions is also equal. This is however already the case when only 35% of the nodes are selected using their out-degree and given a positive opinion. The data also confirms the intuition that having more nodes agree on an opinion initially, makes the network more likely to converge on this opinion and also influences how strongly the individuals in the network adopt this opinion.

Having identified a leadership measure and a general testing strategy with reference cases, the conclusions drawn in chapter 4 are put to the test. This test is in the form of the adaptations made to the DeGroot model chapter 5. Adding uniform levels of self-confidence to the model makes it harder to influence nodes in the beginning, but does not affect the number of cases in which a consensus is reached. Having the individuals pay more attention to leader figures also has no significant impact on these results. The leader nodes still have a significant impact on the network, but are not capable of convincing the remaining nodes. These results show the robustness of the DeGroot model. However, when basing the importance nodes give to their peers on the opinions of these peers or when adding noise to the opinions of neighbors, notably different results are perceived. This form of weighting has many similarities with the bounded confidence model for opinion dynamics. Because nodes pay more attention to nodes that confirm their own beliefs, the leader nodes, which are initially given an extreme opinion, are less effective in convincing the rest of the network. In this case, a better strategy would be to have the leaders start out with a more moderate opinion and increment their opinion values throughout time. This way, their peers would value their input more on average. The last adaptation made to the DeGroot model, was adding noise to the opinion values a node receives from its neighbors. This change to the model was tested with all non-leader nodes having an initially unbiased opinion. Compared to the same experiment without the added noise, the leader nodes are less successful in convincing the rest of the network to move towards the goal opinion. In fact, adding a little bit of normally distributed noise leads to the same outcome as if the other nodes in the network would have had uniformly distributed opinions.

Based on the results of these experiments, the case is made that opinion leaders have the potential of changing the opinion dynamics, simulated using the DeGroot model, in both hierarchical as well as non-hierarchical social networks. This remains true as long as there is a certain threshold of them present. Another condition is that the weights placed on their opinion by their neighbors is not disproportionately small compared to the weights used for other peers. The results also show that this influence is less strong in networks where there is a significant level of hierarchy present.

Bibliography

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. Topic-aware social influence propagation models. *Knowledge and information systems*, 37(3):555–584, 2013.

[3] Rafael A Barrio, Tzipe Govezensky, Robin Dunbar, Gerardo Iniguez, and Kimmo Kaski. Dynamics of deceptive interactions in social networks. *Journal of The Royal Society Interface*, 12(112), 2015.

[4] Alex Bavelas. Communication patterns in task-oriented groups. *The journal of the acoustical society of America*, 22(6):725–730, 1950.

[5] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[6] Linda L Carli. Gender and social influence. *Journal of Social issues*, 57(4):725–741, 2001.

[7] Sujin Choi. The two-step flow of communication in twitter-based public forums. *Social Science Computer Review*, 33(6):696–711, 2015.

[8] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3): 581–588, 1973.

[9] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98, 2000.

[10] Morris H DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[11] Oxford dictionary. Opinion, 2020. URL `https://www.lexico.com/en/definition/opinion`.

[12] Peter Sheridan Dodds. Slightly generalized contagion: Unifying simple models of biological and social spreading. In *Complex Spreading Phenomena in Social Systems*, pages 67–80. Springer, 2018.

[13] Virginia Domínguez-García, Simone Pigolotti, and Miguel A Munoz. Inherent directionality explains the lack of feedback loops in empirical networks. *Scientific reports*, 4:7497, 2014.

[14] Yucheng Dong, Xia Chen, Haiming Liang, and Cong-Cong Li. Dynamics of linguistic opinion formation in bounded confidence model. *Information Fusion*, 32:52–61, 2016.

[15] Yucheng Dong, Min Zhan, Gang Kou, Zhaogang Ding, and Haiming Liang. A survey on the fusion process in opinion dynamics. *Information Fusion*, 43:57–65, 2018.

[16] Bartłomiej Dybiec, A Kleczkowski, and CA Gilligan. Controlling disease spread on networks with incomplete knowledge. *Physical Review E*, 70(6):066145, 2004.

[17] Joshua M Epstein. Modelling to contain pandemics. *Nature*, 460(7256):687–687, 2009.

[18] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2017.

[19] Santo Fortunato. Damage spreading and opinion dynamics on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 348:683–690, 2005.

[20] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

[21] Maksym Gabielkov and Arnaud Legout. The complete picture of the twitter social graph. In *Proceedings of the 2012 ACM conference on CoNEXT student workshop*, pages 19–20, 2012.

[22] David F Gleich. Pagerank beyond the web. *SIAM Review*, 57(3):321–363, 2015.

[23] A Grabowski and RA Kosiński. Epidemic spreading in a hierarchical social network. *Physical Review E*, 70(3):031908, 2004.

[24] Aimee Groth. You're the average of the five people you spend the most time with, 2012. URL `https://www.businessinsider.com/jim-rohn-youre-the-average-of-the-five-people-you-spend-the-most-time-with-2012-7?r=US&IR=T`.

[25] Mangesh Gupte, Pravin Shankar, Jing Li, Shanmugauelayut Muthukrishnan, and Liviu Iftode. Finding hierarchy in directed online social networks. In *Proceedings of the 20th international conference on World wide web*, pages 557–566, 2011.

[26] Huawei Han, Chengcang Qiang, Caiyun Wang, and Jing Han. Soft-control for collective opinion of weighted degroot model. *Journal of Systems Science and Complexity*, 30(3):550–567, 2017.

[27] Martin Hilbert, Javier Vásquez, Daniel Halpern, Sebastián Valenzuela, and Eduardo Arriagada. One step, two step, network step? complementary perspectives on communication flows in twittered citizen protests. *Social Science Computer Review*, 35 (4):444–461, 2017.

[28] Bernardo A Huberman, Daniel M Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *arXiv preprint arXiv:0812.1045*, 2008.

[29] Vivek Kandiah and Dima L Shepelyansky. Pagerank model of opinion formation on social networks. *Physica A: Statistical Mechanics and its Applications*, 391(22):5779–5793, 2012.

[30] Elihu Katz. The two-step flow of communication: An up-to-date report on an hypothesis. *Public opinion quarterly*, 21(1):61–78, 1957.

[31] Kyung-Sun Kim, Sei-Ching Joanna Sin, and Eun Young Yoo-Lee. Undergraduates' use of social media as information sources. 2014.

[32] Yunmi Kim, Tae-Hwan Kim, and Tolga Ergün. The instability of the pearson correlation coefficient in the presence of coincidental outliers. *Finance Research Letters*, 13: 243–257, 2015.

[33] Jon M Kleinberg. Hubs, authorities, and communities. *ACM computing surveys (CSUR)*, 31(4es):5–es, 1999.

[34] Gang Kou, Yiyi Zhao, Yi Peng, and Yong Shi. Multi-level opinion dynamics under bounded confidence. *PloS one*, 7(9), 2012.

[35] MF Laguna, Guillermo Abramson, and Damián H Zanette. Vector opinion dynamics in a model for social influence. *Physica A: Statistical Mechanics and its Applications*, 329(3-4):459–472, 2003.

[36] Zhihong Liu, Jianfeng Ma, Yong Zeng, Li Yang, Qiping Huang, and Hongliang Wu. On the control of opinion dynamics in social networks. *Physica A: Statistical Mechanics and its Applications*, 409:183–198, 2014.

[37] Luis López, Jose FF Mendes, and Miguel AF Sanjuán. Hierarchical social networks and information flow. *Physica A: Statistical Mechanics and its Applications*, 316(1-4): 695–708, 2002.

[38] Lydia Manikonda, Yuheng Hu, and Subbarao Kambhampati. Analyzing user activities, demographics, social network structure and user-generated content on instagram. *arXiv preprint arXiv:1410.8099*, 2014.

[39] Loretta Mastroeni, Pierluigi Vellucci, and Maurizio Naldi. Agent-based models for opinion formation: A bibliographic survey. *IEEE Access*, 7:58836–58848, 2019.

[40] Mattia Mazzoli, Tullio Re, Roberto Bertilone, Marco Maggiora, and Jacopo Pellegrino. Agent based rumor spreading in a scale-free network. *arXiv preprint arXiv:1805.05999*, 2018.

[41] Gary Mckeown and Noel Sheehy. Mass media and polarisation processes in the bounded confidence model of opinion dynamics. *Journal of Artificial Societies and Social Simulation*, 9(1), 2006.

[42] Enys Mones, Lilla Vicsek, and Tamás Vicsek. Hierarchy measure for complex networks. *PloS one*, 7(3), 2012.

[43] Mehdi Moussaïd, Juliane E Kämmer, Pantelis P Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PloS one*, 8(11), 2013.

[44] Seth A Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 493–498, 2014.

[45] Mark Newman. *Networks*. Oxford university press, 2018.

[46] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 11 1999. URL `http://ilpubs.stanford.edu:8090/422/`. Previous number = SIDL-WP-1999-0120.

[47] Carlo Pinciroli, Vito Trianni, Rehan O'Grady, Giovanni Pini, Arne Brutschy, Manuele Brambilla, Nithin Mathews, Eliseo Ferrante, Gianni Di Caro, Frederick Ducatelle, Mauro Birattari, Luca Maria Gambardella, and Marco Dorigo. ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems. *Swarm Intelligence*, 6(4): 271–295, 2012.

[48] Lisa Rashotte. Social influence. *The Blackwell encyclopedia of sociology*, 2007.

[49] Ilja Rausch, Yara Khaluf, and Pieter Simoens. Collective decision-making on triadic graphs. In *Complex Networks XI*, pages 119–130. Springer, 2020.

[50] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical review E*, 67(2):026112, 2003.

[51] Simone Righi and Timoteo Carletti. The influence of social network topology in a opinion dynamics model. In *Proceeding of the European Conference on Complex systems*, 2009.

[52] Francisco Aparecido Rodrigues. Network centrality: an introduction. In *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*, pages 177–196. Springer, 2019.

[53] Casey M Schneider-Mizell and Leonard M Sander. A generalized voter model on complex networks. *Journal of Statistical Physics*, 136(1):59–71, 2009.

[54] Pawel Sobkowicz. Modelling opinion formation with physics tools: Call for closer link with reality. *Journal of Artificial Societies and Social Simulation*, 12(1):11, 2009.

[55] Xiao Song, Wen Shi, Yaofei Ma, and Chen Yang. Impact of informal networks on opinion dynamics in hierarchically formal organization. *Physica A: Statistical Mechanics and its Applications*, 436:916–924, 2015.

[56] Patrick Taillandier, Benoit Gaudou, Arnaud Grignard, Quang-Nghi Huynh, Nicolas Marilleau, Philippe Caillou, Damien Philippon, and Alexis Drogoul. Building, composing and experimenting complex spatial models with the gama platform. *GeoInformatica*, 23(2):299–322, 2019.

[57] Wai M Tam, Francis CM Lau, and Chi K Tse. Construction of scale-free networks with adjustable clustering. In *Proceedings of 2008 International Symposium on Nonlinear Theory and its Applications*, pages 257–260, 2008.

[58] Gunjan Verma, Ananthram Swami, and Kevin Chan. The impact of competing zealots on opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 395:310–331, 2014.

[59] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[60] Ercan Yildiz, Daron Acemoglu, Asuman E Ozdaglar, Amin Saberi, and Anna Scaglione. Discrete opinion dynamics with stubborn agents. Available at SSRN: https://ssrn.com/abstract=1744113, 2011.

[61] Jerrold H Zar. Spearman rank correlation. *Encyclopedia of Biostatistics*, 7, 2005.

FACULTY OF ENGINEERING
AND ARCHITECTURE

How opinion leaders can manipulate opinion dynamics in hierarchical social networks

Xavier Claerhoudt
Student number: 01505080

Supervisors: Prof. dr. ir. Pieter Simoens, Dr. Yara Khaluf
Counsellor: Ilja Rausch

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Computer Science Engineering

Academic year 2019-2020

GHENT
UNIVERSITY