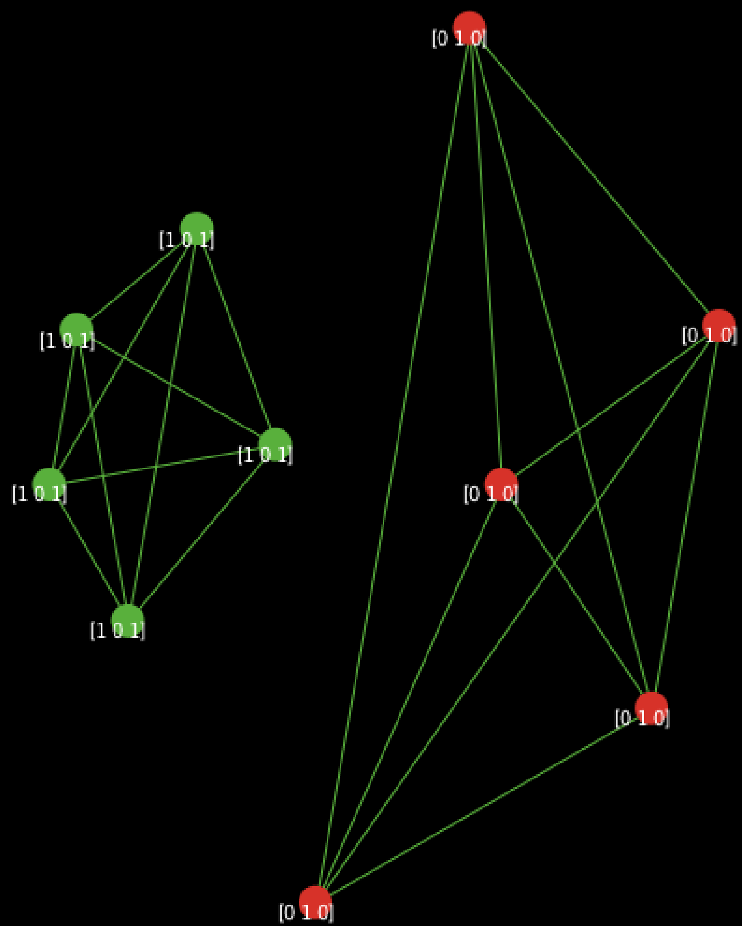
A logical analysis of influencers and polarization in social networks



Djanira dos Santos Gomes

Layout: typeset by the author using LaTeX.
Cover illustration: Polarized network. Created by the author using NetLogo (Wilensky, 1999).

A logical analysis of influencers and polarization in social networks

Djanira dos Santos Gomes
11869674

Bachelor thesis
Credits: 18 EC

Bachelor *Kunstmatige Intelligentie*

University of Amsterdam
Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervisor
Prof. dr. S.J.L. Smets

Institute for Logic, Language and Computation
Faculty of Science
University of Amsterdam
Science Park 907
1098 XG Amsterdam

June 25th, 2021

Abstract

This thesis focuses on modelling the role of influencers and their effect on the two mechanisms that underly the formation of polarized groups: social influence and friendship selection. In contrast to existing quantitative methods that have previously been developed in network theory, we make use of qualitative methods to study the information flow in social networks. We propose a threshold model, which makes use of update rules concerning social influence and friendship selection, in which we make the role of influencers explicit with respect to a logic-based characterization of group polarization. In this thesis we define rigid and flexible influencers to be agents who increase their proportion of friends within one round of social influence and friendship selection. Rigid influencers do so without being influenced to change their opinion, while flexible influencers are allowed to adapt part of their opinion.

The theoretical model is implemented in Python, as well as in the multi-agent modelling environment NetLogo, and used to run a simulation on a number of network topologies with varying threshold values. Through these simulations we examine under which conditions group polarization is more likely to occur and what role influencers play in this process.

The results of this study indicate that the majority of simulations of this model polarize within the first three rounds of friendship selection and social influence. Furthermore, stricter friendship selection thresholds were found to increase the probability of the network developing into two components with opposing opinions. Finally, rigid influencers were found to play a larger role in networks developing full consensus than in networks developing opposing components.

Contents

1 Introduction

The modelling of *social influence* in social networks has become increasingly popular in recent years in network theory, logic and in AI (see e.g. Baltag et al. (2018); Smets and Velázquez-Quesada (2020); Baltag et al. (2013); Tsugawa and Kimura (2018); Easley and Kleinberg (2019)). Social influence describes the process of agents changing their behaviour or opinions based on those of agents to whom they are socially related (Smets and Velázquez-Quesada, 2020). The mechanism of social influence is one aspect that underlies the creation of different social phenomena such as *group polarization* (Alvim et al., 2019; Young Pedersen et al., 2020). More specifically, the occurrence of group polarization can be triggered by *homophily*: the tendency of humans to make social connections with people who have similar interests (*friendship selection*) and to adapt their beliefs and opinions to those of their social relations (*social influence*) (Smets and Velázquez-Quesada, 2020; Easley and Kleinberg, 2019). Due to the formation of these like-minded groups of people, *echo chambers* can form: extreme opinions representing similar views are more likely to be heard than those representing opposing ones (Young Pedersen et al., 2019). Furthermore, one is likely to object more strongly to opposing views (Alvim et al., 2019). This causes people to develop radicalized opinions, thus dividing social networks into groups with two strongly opposing views, rather than a distribution of nuanced opinions (Alvim et al., 2019; Young Pedersen et al., 2020).
To date, in logic only a few studies have investigated the interplay between friendship selection and social influence (Smets and Velázquez-Quesada, 2020). Furthermore, even though both friendship selection and social influence play a substantial role in the process of homophily (Smets and Velázquez-Quesada, 2020; Easley and Kleinberg, 2019), existing models of group polarization are mainly focused on the dynamics of social influence. Several models have been constructed to simulate both these dynamic processes separately, of which in logic a large number makes use of the techniques of Dynamic Epistemic Logics (DEL) (see e.g. Baltag et al. (2018, 2013); Young Pedersen et al. (2019); Christoff and Hansen (2015); Smets and Velázquez-Quesada (2017); Ditmarsch et al. (2008)). Smets and Velázquez-Quesada (2020) used these techniques to combine the dynamics of friendship selection and social influence into one threshold model and investigate the intertwining dynamics.
Other than polarization, influencers are of interest in social network research (see e.g. Tsugawa and Kimura (2018); Alvim et al. (2019); Chen et al. (2012)) and are expected to play an important part in the process of group polarization (Tsugawa and Kimura, 2018). The ability of influencers to sway the opinions of their followers has been reflected in the rapid development of influencer marketing during the past decade (Geyser, 2021). There is increasing concern about negative implications of this ability: for example, in 2020 Dutch influencers were recruited

by an organization called *Viruswaarheid* ("Virus Truth") to promote a conspiracy theory concerning the Covid-19 virus (Holligan, 2020). This occurred in the heat of the pandemic when the Dutch were strongly divided over the restrictions placed by the government and it was feared that the influencers would undermine the support of the public health campaign.

In network theory, the term *influencer* indicates an influential node in the network: that is, a node that spreads information to a larger part of the network than other nodes (Tsugawa and Kimura, 2018). In social network analysis focused on influencers, two notions of *influencers* can be distinguished, which in this thesis will be referred to as the *natural influencer* and the *designated influencer*, respectively. Whereas in the former case, influential power is dependent on the current position of the agent in the network, in the latter it is regarded as a property of the agent. This thesis will focus on modelling the role of natural influencers and their effect on the two mechanisms that underly the formation of polarized groups. With focus on the logic-based literature on social influence and polarization, the role of influencers will be defined in a logic-based setting. When doing so, different parameters will play a role, such as the number of agents (*nodes*), the structure of the network, the valid propositions at each node and the dynamics of information in the network.

In contrast to existing quantitative methods that have previously been developed in network theory, we make use of qualitative methods to study the information flow in social networks. We propose a threshold model in which we make the role of natural influencers explicit with respect to a logic-based characterization of group polarization, as defined by Easley and Kleinberg (2019). This model takes the model provided by Smets and Velázquez-Quesada (2020) as a starting point.

Furthermore, the theoretical model is implemented in Python, as well as in the multi-agent modelling environment NetLogo (Wilensky, 1999), and used to run a simulation on a number of network topologies described by Zollman (2013) with varying threshold values. Through these simulations we examine under which conditions group polarization is more likely to occur and what role influencers play in this process. The full implementation of this research can be found at https://github.com/DdosSantosGomes/Thesis.

2 Literature Review

2.1 Dynamic Epistemic Logics

Dynamic Epistemic Logics (DEL) are extensions of modal logic that include dynamic operators (Baltag et al., 2018) and thus enable the modelling of transformations of multi-agent systems. These dynamic features prove to be particularly

useful for the study of information flow in social networks (Smets and Velázquez-Quesada, 2020).
Transformations of a social network model are defined by update procedures. These procedures may affect the epistemic properties or features of agents, as is evident in the case of epistemic or doxastic updates (Baltag et al., 2018; Smets and Velázquez-Quesada, 2020), or updates concerning social influence (Baltag et al., 2018; Smets and Velázquez-Quesada, 2020; Baltag et al., 2013; Christoff and Hansen, 2015). Others, such as friendship selection updates (Smets and Velázquez-Quesada, 2017), affect the social structure of the model.
Although the update procedure is defined globally in these models, the updates are in fact local procedures: an act of social influence, for instance, may not affect the beliefs of every agent in the model.
In particular, threshold network models make use of threshold-limited updates: the individual agents in such models update their positions or relations solely if a certain threshold is reached (Baltag et al., 2018; Smets and Velázquez-Quesada, 2020, 2017).
Smets and Velázquez-Quesada (2020) used DEL techniques to describe the processes of friendship selection and social influence in a threshold model. The corresponding update rules can be considered duals of one another: whereas the friendship selection update affects the nearest neighbours of an agent based on their positions on certain topics, the social influence update impacts the agent's positions on those topics based on those of their nearest neighbours. Alternately applying these updates on a social network model thus results in a simulation of the dynamics of a social network.

2.2 Influencers in Social Networks

The concept of *power* in social networks has been widely researched in both network theory and sociology. A main challenge in this field of research is to analyse to what extent power results from the position of an agent in the network, and to what extent it should be regarded as a property of the agent itself (Easley and Kleinberg, 2019).
The same challenge exists in research into influential power. Two approaches to the problem of defining *influencers* have been proposed: the *natural influencer* and the *designated influencer*. Whereas definitions describing the natural influencer consider influential power a result of the position of an agent, those describing the designated influencer consider it a property.
In network theory, influential nodes in social networks have been defined according to either of these definitions. To date, only qualitative definitions of designated influencers have been constructed and only quantitative definitions of natural influencers have been constructed.

2.2.1 Natural Influencers

Natural influencers obtain their power through their position in the network. As a result, agents may acquire or lose their influencer status in a changing network. Social media influencers, for instance, can be considered natural influencers.
In social network models, natural influencers are identified quantitatively through influence measures. These measures compare the positions of single nodes in the network in order to identify influential nodes. Examples of quantitative measures of influence are *centrality* measures and *PageRank* (Tsugawa and Kimura, 2018). The accuracy of these measures is tested through investigation of the actual influential power of identified influential nodes in existing social networks. The influential power of these nodes is empirically measured by the size of the *information cascades* initiated by these nodes: the number of nodes influenced by the node concerned (Tsugawa and Kimura, 2018).
The most straightforward, though least accurate, measure is *degree centrality*, which considers the nodes with the largest numbers of nearest neighbours in a social network to be influencers (Chen et al., 2012). This measure tends to overvalue the influential power of nodes with a high number of nearest neighbours, but a relatively low number of next nearest neighbours. An alternative, proposed by Chen et al. (2012), is the *local centrality measure*, which incorporates both nearest and next nearest neighbours and was found to provide a more adequate representation of influence in a social network. Measures with a higher accuracy are of higher computational complexity.

2.2.2 Designated Influencers

Designated influencers have the property of being influential. They have higher influential power in a social network by definition and thus cannot lose their influencer status. Authorities and experts, for instance, can be considered designated influencers.
In social network models, designated influencers are defined qualitatively. Liu et al. (2014), for example, proposed a model of influence through *believed reliability* as an alternative to threshold influence. In their model, nodes that are considered more reliable have higher influential power.
Alvim et al. (2019) took a similar approach: their model includes an *influence graph*, describing the influence of every node in the model on each of its nearest neighbours. Using DEL, belief updates of an agent are then defined as a function of the influences from their nearest neighbours.
In their graph configurations, Alvim et al. (2019) mention two common configurations of influencers in a social network: an *unrelenting influencers* graph and a *malleable influencers* graph. Both describe situations where two influencers both

have a strong influence on all other agents. However, unrelenting influencers are not influenced by other nodes, whereas malleable influencers are slightly influenced by others.

2.3 Measuring Group Polarization

Both qualitative and quantitative methods of measurement of polarization have been defined in previous research. Alvim et al. (2019), for instance, made use of a quantitative measure: the Esteban-Ray polarization measure computes the degree of polarization among a group of agents from a distribution of these agents over their beliefs. According to this measure, polarization is characterized by a few large groups of agents each having respectively different beliefs.

As the Esteban-Ray measure focuses solely on the distribution of beliefs over the entire network, it does not take into account the aspect of homophily. As a consequence, a social network where consensus has been reached within clusters does not necessarily have a higher degree of polarization than one where agents within clusters disagree.

One qualitative polarization measure, which takes homophily into consideration, was described by Easley and Kleinberg (Easley and Kleinberg, 2019), as well as Young Pedersen et al. (2020), and makes use of the *Structural Balance Property*. The measure is based on the phenomenon of *structural balance*: a triangle of nodes is structurally balanced if either all three nodes are friends, or two nodes are friends and the third node is their mutual enemy. With respect to *signed*[1]graphs, a balanced triangle either consists solely of positive edges or contains exactly one positive edge.

Generalized to *complete*[2] graphs, a network is balanced if either every two nodes are friends, or a balanced division of the network exists. In the latter case, the network can be divided into groups X and Y, such that every two agents $i, j \in X$ are friends, every two agents $i, j \in Y$ are friends, and every two agents $i \in X$ and $j \in Y$ are enemies. Equivalently, a balanced network does not contain any cycles containing an odd number of negative edges.

For a non-complete graph to be structurally balanced, all existing edges within the same group should be positive, while those connecting opposite groups should be negative.

Lastly, Easley and Kleinberg (2019) provide *weak structural balance* as an alternative to structural balance, which disregards the existence of triangles consisting

[1]A *signed*, or *labeled*, graph is a graph in which each edge connecting two nodes is either annotated with a positive or negative sign (Easley and Kleinberg, 2019).

[2]A *complete*, or *fully connected*, graph is a graph in which each pair of nodes is connected by an edge (Easley and Kleinberg, 2019). A signed complete graph is a complete graph in which each edge connecting two nodes is either annotated with a positive or negative sign.

solely of negative edges.

2.4 Network Topology

In order to study the development of polarization in relation to influencers, we restrict our attention to a subset of network structures. More specifically, we focus on random network structures as well as three well-known network topologies mentioned by Zollman (2013). As this thesis makes use solely of symmetric relations, the graphs in question are undirected.

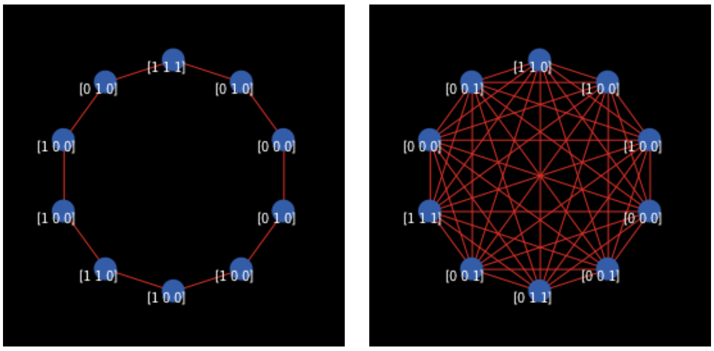
The multi-agent modelling environment NetLogo provides an algorithm (see Wilensky and Rand (2008)) for an Erdös-Rényi (ER) random network, which is a randomly built network consisting of a specified number of nodes and number of links. The algorithm randomly selects pairs of nodes to be connected by an edge until either the requested number of links, or the maximum number of possible links between the given nodes, has been reached. Depending on the numbers of nodes and links, the algorithm might result in a *disconnected*[3] graph.

Based on Zollman (2013), the *ring, star*, and *fully connected topology* are considered. Furthermore, we define the *double star* as a variation of the star. Each of the topologies describes a *connected*[4] graph. The *ring topology* (see figure 1a) describes a ring-shaped graph: the graph consists of one single cycle, where every node is directly related to exactly two other nodes. The *fully connected topology* (see figure 1b) describes a *fully connected* graph, in which every node is directly related to every other node. The *star topology* (see figure 2a) describes a star-shaped graph, where one centered node is related to every other node in the network, and the other nodes are not related to one another. Lastly, the *double star* variant (see figure 2b) describes a network composed of two equally large star networks, such that their center nodes are connected to one another.

Figures 1 and 2 provide visualizations of the network topologies, created in the NetLogo modelling environment (Wilensky, 1999). The networks in figures 1a, 1b, 2a and 2b each consist of ten nodes and an according number of links, while the ER random network in figure 2c consists of twenty nodes and ten links.

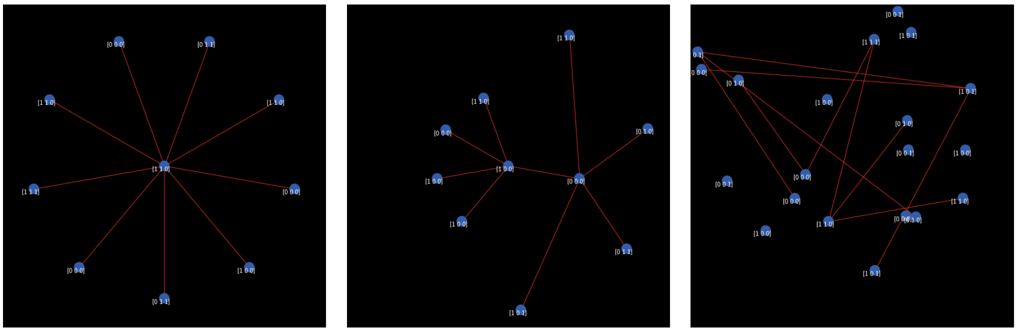
[3] A *disconnected* graph is a graph consisting of at least two *components* (Easley and Kleinberg, 2019). A *component* (sometimes referred to as *connected component*) is a subgraph of a graph which is connected when considered as a graph in isolation, for which holds that the nodes forming the component do not form a subset of a larger set that is connected. In other words, there exists no edge linking a component to any other node in the graph.

[4] A *connected* graph is a graph in which there exists a path between each pair of nodes (Easley and Kleinberg, 2019).



(a) Ring (b) Fully connected

Figure 1: Ring (a) and fully connected network (b). Both networks consist of 10 nodes.



(a) Star (b) Double star (c) ER random network

Figure 2: Star (a) and double star (b), both consisting of 10 nodes, and ER random network example (c). This network consists of 20 nodes and 10 links.

3 Method and Approach

3.1 Social Network Model

Our social network model takes the model provided by Smets and Velázquez-Quesada (2020) as a starting point. The domain of their model represents a set of agents who form beliefs concerning different topics and form friendship relations with one another. Different from the model provided by Smets and Velázquez-Quesada, this model consists of only one social network. Furthermore, our model extends the model defined by Smets and Velázquez-Quesada to a signed graph consisting of positive and negative relations. A positive relation between two agents will be interpreted as a friendship, whereas a negative relation implies mere acquaintanceship. The negative relation in this model is not necessarily hostile,

though antagonistic and characterized by disagreement, as described by Easley and Kleinberg (2019). The valuation function in our model is as defined by Smets and Velázquez-Quesada.
Following Smets and Velázquez-Quesada (2020), let $\mathcal{A} \neq \emptyset$ be a finite set of agents. Let $\mathcal{T} \neq \emptyset$, with $\mathcal{A} \cap \mathcal{T} = \emptyset$, be a finite set of topics, with $\{\mathcal{R}_T\}_{T \in \mathcal{T}}$ a pairwise disjoint collection such that for each $T \in \mathcal{T}$, $\mathcal{R}_T$ is a finite non-empty set of positions on topic $\mathcal{T}$. Define $\mathcal{R} := \bigcup_{T \in \mathcal{T}} \mathcal{R}_T$.
In this thesis we work with the case study in which we define the set of topics $\mathcal{T} = \{P, Q, R\}$ with positions $\mathcal{R}_P = \{p\}$, $\mathcal{R}_Q = \{q\}$, and $\mathcal{R}_R = \{r\}$. We have for each topic that agents may or may not adopt the position concerned with that topic. Consider, for instance, the interpretation of topic P as "vaccination" and the interpretation of position p as the opinion "vaccination should be compulsory". This means that an agent who does not take position p on topic P does not believe that vaccination should be compulsory.

Definition 3.1 (Social Network Model)**.** A social network model (SNM) for a given $\mathcal{A}, \mathcal{T}$ and $\{\mathcal{R}_T\}_{T \in \mathcal{T}}$ is a tuple $\langle N_P, N_N, V \rangle$ where

- The positive relations in the social network are given by $N_P \subseteq \mathcal{A} \times \mathcal{A}$, with $N_P ij$ indicating that agents $i \in \mathcal{A}$ and $j \in \mathcal{A}$ are friends;
- The negative relations in the social network are given by $N_N \subseteq \mathcal{A} \times \mathcal{A}$, with $N_N ij$ indicating that agents $i \in \mathcal{A}$ and $j \in \mathcal{A}$ are acquaintances, but are not friends;
- $V : \mathcal{T} \times \mathcal{R} \to \wp(\mathcal{A})$ is the valuation function, with each $V_T(r) \subseteq \mathcal{A}$ (defined only for $r \in \mathcal{R}_T$) indicating the set of agents adopting position $r \in \mathcal{R}_T$ on topic $T \in \mathcal{T}$.

The dual function of V returns the set of positions $\Lambda_T(i) \subseteq \mathcal{R}_T$ that each agent $i \in \mathcal{A}$ holds on topic $T \in \mathcal{T}$ and is given by $\Lambda_T(i) := \{r \in \mathcal{R}_T \mid i \in V_T(r)\}$ (Smets and Velázquez-Quesada, 2020). We define $\Lambda(i) := \bigcup_{T \in \mathcal{T}} \Lambda_T(i)$.

The propositional language describing an SNM is as defined by Smets and Velázquez-Quesada (2020), though adapted to allow for the description of positive and negative relations.

Definition 3.2 (Language $\mathcal{L}$)**.** With $i, j \in \mathcal{A}, T \in \mathcal{T}$ and $r \in \mathcal{R}_T$, formulas in $\mathcal{L}$ are given by

$$\varphi, \psi ::= i_r^T \mid i \rhd^P j \mid i \rhd^N j \mid \neg\varphi \mid \varphi \wedge \psi$$

where formulas of the form i_r^T denote that agent i has taken position r on topic T; those of the form $i \rhd^P j$ (resp. $i \rhd^N j$) denote that agent i has a positive (resp. negative) social relation with agent j. Other Boolean operators, such as the disjunction and material implication, are defined as usual. For the sake of simplicity in notation, formulas of the form i_r^T will hereinafter be referred to as i_r, as the topic T follows from the position r with $\{\mathcal{R}_T\}_{T \in \mathcal{T}}$ being pairwise disjoint. Given a SNM $M = \langle N_P, N_N, V \rangle$, formulas in $\mathcal{L}$ are semantically interpreted as follows:

$$
\begin{array}{rr}
M \Vdash i_r \text{ iff}_{\text{def}}\, i \in V_T(r) & M \Vdash \neg\varphi \text{ iff}_{\text{def}}\, M \nVdash \varphi \\
M \Vdash i \rhd^P j \text{ iff}_{\text{def}}\, N_P ij & M \Vdash \varphi \wedge \psi \text{ iff}_{\text{def}}\, M \Vdash \varphi \text{ and } M \Vdash \psi \\
M \Vdash i \rhd^N j \text{ iff}_{\text{def}}\, N_N ij &
\end{array}
$$

A formula $\varphi \in \mathcal{L}$ is valid (notation: $\Vdash \varphi$) when $M \Vdash \varphi$ holds for every SNM M.

3.2 Social Network Dynamics

3.2.1 Social Influence

Following Smets and Velázquez-Quesada (2020), an agent in the network is influenced to take a certain position on a topic if the proportion of his relations having this position is greater than or equal to a certain *threshold*. This process is described by the *social influence* operation. For our model, we extend the definition of this operation provided by Smets and Velázquez-Quesada to include negative relations, such that an agent is influenced by friends as well as acquaintances.

Definition 3.3 (Social Influence)**.** Let τ be a real number in [0, 1]. Let $M = \langle N_P, N_N, V \rangle$ be a SNM and, for $i \in \mathcal{A}$, define

$$
\begin{aligned}
N_P[i] &:= \{j \in \mathcal{A} \mid N_P ij\}, \\
N_N[i] &:= \{j \in \mathcal{A} \mid N_N ij\}, \\
N[i] &:= N_P[i] \cup N_N[i],
\end{aligned}
$$

such that $N[i]$ is the set of agents socially connected to agent i; $N_P[i]$ is the set of friends of i; and $N_N[i]$ is the set of mere acquaintances of i.
The *social influence* operation is a model transformation which acts on a given SNM $M = \langle N_P, N_N, V \rangle$, and given $\mathcal{A}, \mathcal{T}$ and $\mathcal{R}_T$, and results in an updated SNM $M^{\dagger(\tau)} = \langle N_P, N_N, V' \rangle$, where for any $i \in \mathcal{A}, T \in \mathcal{T}$ and $r \in \mathcal{R}_T$ in M, the new valuation function is given by

$$
i \in V'_T(r) \text{ iff}_{\text{def}} \begin{cases} \frac{|N[i] \cap V_T(r)|}{|N[i]|} \geq \tau, & \text{if } N[i] \neq \emptyset \\ i \in V_T(r), & \text{otherwise.} \end{cases}
$$

Note that the condition that $N[i] \neq \emptyset$ was included to account for non-reflexive relations. In this thesis the reflexive positive relation in the model (see definition 3.6) automatically results in a non-empty neighbourhood for each agent and therefore, the condition is not strictly necessary.
Through the extension of the language $\mathcal{L}$ with dynamic modality $\langle \dagger_\tau \rangle$, we obtain the language $\mathcal{L}^\dagger := \mathcal{L} + \langle \dagger_\tau \rangle$, which allows us to describe the effects of the social influence operation. Following Smets and Velázquez-Quesada (2020), the modality $\langle \dagger_\tau \rangle$ (for $\tau \in [0,1]$) is semantically interpreted as

$$M \Vdash \langle \dagger_\tau \rangle \varphi \text{ iff}_{\text{def}} M^{\dagger(\tau)} \Vdash \varphi$$

where the formula $\langle \dagger_\tau \rangle \varphi$ denotes that φ is true after an act of social influence.

Axiom System. Following Smets and Velázquez-Quesada (2020), the DEL technique of *recursion axioms* is used to build the axiom system for the language with the new dynamic modality. Recursion axioms define a truth-preserving translation from the language with the dynamic modality to the language without it, through valid formulas and validity-preserving rules. The following abbreviations, based on Smets and Velázquez-Quesada, are used in the axiom $\dagger_{i_r}$, which describes the conditions under which an agent changes his position (see table 1). The axioms $\dagger_{i \rhd^P j}$ and $\dagger_{i \rhd^N j}$, provided in table 1, are signed adaptations to the axiom provided by Smets and Velázquez-Quesada which denotes that the influence operation does not affect existing relations. The other axioms are equivalent to those provided by Smets and Velázquez-Quesada.
For $i, j \in \mathcal{A}; \mathcal{B}, \mathcal{B}' \subseteq \mathcal{A}; T \in \mathcal{T}$, define

- $i \rhd^P \mathcal{B} := \bigwedge_{j \in \mathcal{B}} i \rhd^P j \wedge \bigwedge_{j \in \mathcal{A} \setminus \mathcal{B}} \neg(i \rhd^P j)$ (so $M \Vdash i \rhd^P \mathcal{B}$ iff $N_P[i] = \mathcal{B}$)
- $i \rhd^N \mathcal{B} := \bigwedge_{j \in \mathcal{B}} i \rhd^N j \wedge \bigwedge_{j \in \mathcal{A} \setminus \mathcal{B}} \neg(i \rhd^N j)$ (so $M \Vdash i \rhd^N \mathcal{B}$ iff $N_N[i] = \mathcal{B}$)
- $i \rhd^{P \cup N} j := i \rhd^P j \vee i \rhd^N j$ (so $M \Vdash i \rhd^{P \cup N} j$ iff $j \in N_P[i] \cup N_N[i]$)
- $i \rhd^{P \cup N} \mathcal{B} := \bigwedge_{j \in \mathcal{B}} i \rhd^{P \cup N} j \wedge \bigwedge_{j \in \mathcal{A} \setminus \mathcal{B}} \neg(i \rhd^{P \cup N} j)$ (so $M \Vdash i \rhd^{P \cup N} \mathcal{B}$ iff $N[i] = \mathcal{B}$)
- $sst_r^T(\mathcal{B}', \mathcal{B}) := \bigwedge_{j \in \mathcal{B}'} j_r \wedge \bigwedge_{j \in \mathcal{B} \setminus \mathcal{B}'} \neg j_r$ (so $M \Vdash sst_r^T(\mathcal{B}', \mathcal{B})$ iff $\begin{cases} \mathcal{B}' \subseteq V_T(r), \text{ and} \\ (\mathcal{B} \setminus \mathcal{B}') \subseteq \overline{V_T(r)} \end{cases}$)

Theorem 3.4. *The axiom system characterising the validities of the language $\mathcal{L}^\dagger$ in SNMs, provided by the propositional axiom system schema, the recursion axioms, and the rules in table 1, is sound and strongly complete.*

Table 1: Recursion axioms for $\langle\dagger_\tau\rangle$ w.r.t. SNMs.

Name of axiom	Axiom in the language $\mathcal{L}^\dagger$
$(\dagger_{i_r})$	$\vdash \langle\dagger_\tau\rangle i_r \leftrightarrow ((i \rhd^{P\cup N} \emptyset \wedge i_r) \vee (\neg(i \rhd^{P\cup N} \emptyset) \wedge \bigvee_{\mathcal{B}\subseteq\mathcal{A}} (i \rhd^{P\cup N} \mathcal{B} \wedge \bigvee_{\mathcal{B}'\subseteq\mathcal{B}:\frac{\lvert\mathcal{B}'\rvert}{\lvert\mathcal{B}\rvert}\geq\tau} sst_r^T(\mathcal{B}',\mathcal{B}))))$
$(\dagger_{i\rhd^P j})$	$\vdash \langle\dagger_\tau\rangle i \rhd^P j \leftrightarrow i \rhd^P j$
$(\dagger_{i\rhd^N j})$	$\vdash \langle\dagger_\tau\rangle i \rhd^N j \leftrightarrow i \rhd^N j$
$(\dagger_\neg)$	$\vdash \langle\dagger_\tau\rangle\neg\varphi \leftrightarrow \neg\langle\dagger_\tau\rangle\varphi$
$(\dagger_\wedge)$	$\vdash \langle\dagger_\tau\rangle(\varphi\wedge\psi) \leftrightarrow (\langle\dagger_\tau\rangle\varphi \wedge \langle\dagger_\tau\rangle\psi)$
$(\dagger_{SPE})$	From $\vdash \psi_1 \leftrightarrow \psi_2$ infer $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more occurrences of ψ_1 in φ with ψ_2.

Proof. Soundness follows from the validity and validity-preserving properties of the axioms and rule, as shown by Smets and Velázquez-Quesada (2020). The $\dagger_{i_r}$ axiom, in particular, is equivalent to the axiom provided by Smets and Velázquez-Quesada (2020), as it describes the new valuation on the updated model (see definition 3.3): an agent $i \in \mathcal{A}$ is influenced to take a position $r \in \mathcal{R}$ depending on the number of agents in the set $N[i]$ of agents related to him, supporting position r. Furthermore, axioms $\dagger_{i\rhd^P j}$ and $\dagger_{i\rhd^N j}$ denote that the social influence operation does not affect existing social relations.
Completeness follows from the validity-preserving translation from $\mathcal{L}^\dagger$ to $\mathcal{L}$, defined by the *rewrite rules* consisting of the axioms and the rule $\dagger_{SPE}$: by sequentially eliminating occurrences of $\langle\dagger_\tau\rangle$ - starting with the deepest occurrence - in a formula in the language $\mathcal{L}^\dagger$, an equivalent formula in the language $\mathcal{L}$ is obtained. As the propositional system of classical logic is complete, $\mathcal{L}^\dagger$ is complete as well (Smets and Velázquez-Quesada, 2020). Example 3.5 illustrates how the rewrite rules provide a translation from a $\mathcal{L}^\dagger$-formula to the language $\mathcal{L}$. $\square$

Example 3.5. Consider the formula $\langle\dagger_\tau\rangle\neg(a \rhd^P b \wedge \langle\dagger_\tau\rangle a_p)$ in the language $\mathcal{L}^\dagger$. We translate this formula to $\mathcal{L}$ as follows.

$$\begin{aligned}
\langle\dagger_\tau\rangle\neg(a_p \wedge \langle\dagger_\tau\rangle a \rhd^P b) &\leftrightarrow \langle\dagger_\tau\rangle\neg(a_p \wedge a \rhd^P b) && (\dagger_{i\rhd^P j})\\
&\leftrightarrow \neg\langle\dagger_\tau\rangle(a_p \wedge a \rhd^P b) && (\dagger_\neg)\\
&\leftrightarrow \neg(\langle\dagger_\tau\rangle a_p \wedge \langle\dagger_\tau\rangle a \rhd^P b) && (\dagger_\wedge)\\
&\leftrightarrow \neg(\langle\dagger_\tau\rangle a_p \wedge a \rhd^P b) && (\dagger_{a\rhd^P j})\\
&\leftrightarrow \neg((a \rhd^{P\cup N} \emptyset \wedge a_r) \vee (\neg(a \rhd^{P\cup N} \emptyset) \wedge \bigvee_{\mathcal{B}\subseteq\mathcal{A}} (a \rhd^{P\cup N} \mathcal{B}\\
&\quad \wedge \bigvee_{\mathcal{B}'\subseteq\mathcal{B}:\frac{\lvert\mathcal{B}'\rvert}{\lvert\mathcal{B}\rvert}\geq\tau} sst_r^T(\mathcal{B}',\mathcal{B})))) \wedge a \rhd^P b) && (\dagger_{i_r})
\end{aligned}$$

3.2.2 Friendship Selection

Similarly to Smets and Velázquez-Quesada (2020), agents in the network acquire and lose relationships with one another according to a similarity-based policy, where the *distance* between two agents denotes the difference in their opinions. As our model includes negative relations, the distance measure is used to determine if a positive (friendship), negative (acquaintanceship), or no relation at all is formed. The positive and negative relation are non-overlapping.
Different from Smets and Velázquez-Quesada, the total set of positions $\Lambda(i)$ is regarded instead of $\Lambda_T(i)$ per topic $T \in \mathcal{T}$ and thus, similarity and relations are not defined on separate topics, but on the entire set of topics.

Definition 3.6 (Friendship Selection)**.** Let θ_1 and θ_2 be real numbers in [0, 1] and let $M = \langle N_P, N_N, V\rangle$ be a SNM. For $i, j \in \mathcal{A}$, define

$$\begin{aligned} \mathrm{diff}(i,j) &:= \{r \in \mathcal{R} \mid r \in \Lambda(i) \veebar r \in \Lambda(j)\}, \\ \mathrm{dist}(i,j) &:= |\mathrm{diff}(i,j)| \end{aligned}$$

where $\veebar$ represents the exclusive disjunction. Thus, $\mathrm{dist}(i,j)$ is given by the cardinality of the set of positions on which agents i and j disagree. The set $\mathrm{diff}(i,j)$ is a subset of the union of the positions of all topics, $\mathcal{R}$.
The *friendship selection* operation is a model transformation which acts on a given SNM $M = \langle N_P, N_N, V\rangle$, and given $\mathcal{A}, \mathcal{T}, \mathcal{R}_T$ and θ_1 and θ_2, and results in an updated SNM $M^{\#(\theta_1,\theta_2)} = \langle N'_P, N'_N, V\rangle$, where for any $i, j \in \mathcal{A}$ in M, the new positive relations are given by

$$N'_P ij \text{ iff}_{\text{def}} \frac{\mathrm{dist}(i,j)}{|\mathcal{R}|} \leq \theta_1$$

and the new negative relations are given by

$$N'_N ij \text{ iff}_{\text{def}}\ \theta_1 < \frac{\mathrm{dist}(i,j)}{|\mathcal{R}|} \leq \theta_2$$

with $\theta_1 < \theta_2$. Thus, in order to form a positive relation between two agents, a smaller distance between the agents is required than for negative relations.

In particular, in this thesis we work with the case studies in which we define three configurations of θ_1 and θ_2: we define (1) $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{2}{3}$, (2) $\theta_1 = 0$ and $\theta_2 = \frac{2}{3}$, and (3) $\theta_1 = 0$ and $\theta_2 = \frac{1}{3}$.
In the first configuration, for example, $\mathcal{R} = \{p, q, r\}$ gives us the following: a negative relation requires a difference in at most two propositions, a positive relation requires a difference in at most one proposition, and a difference in all three

propositions does not induce any relation. As the difference of an agent to himself is always 0, the friendship relation is reflexive. Consequentially, an agent cannot have an acquaintance relation with himself.
Through the extension of the language $\mathcal{L}$ with dynamic modality $\langle\#_{\theta_1,\theta_2}\rangle$, we obtain the language $\mathcal{L}^{\#} := \mathcal{L} + \langle\#_{\theta_1,\theta_2}\rangle$, which allows us to describe the effects of the friendship selection operation. Following Smets and Velázquez-Quesada (2020), the dynamic modality $\langle\#_{\theta_1,\theta_2}\rangle$ (for $\theta_1, \theta_2 \in [0,1]$) is semantically interpreted as

$$M \Vdash \langle\#_{\theta_1,\theta_2}\rangle\varphi \text{ iff}_{\text{def}} M^{\#(\theta_1,\theta_2)} \Vdash \varphi$$

where the formula $\langle\#_{\theta_1,\theta_2}\rangle\varphi$ denotes that φ is true after an act of friendship selection.

Axiom System. The axiom system for the dynamic modality $\langle\#_{\theta_1,\theta_2}\rangle\varphi$ is built using the same technique of recursion axioms. The following abbreviation is used in the axioms $\#_{i\rhd^P j}$ and $\#_{i\rhd^N j}$ (see table 2) which describe the conditions under which social connections are created and dissolved. These axioms are signed variations of the axiom provided by Smets and Velázquez-Quesada. The other axioms in table 2 are unchanged from Smets and Velázquez-Quesada.
For $i, j \in \mathcal{A}; S \subseteq \mathcal{R}$, define

$$\text{diff}(i, \mathcal{S}, j) := \bigwedge_{r\in\mathcal{S}} (i_r \veebar j_r) \wedge \bigwedge_{r\in\mathcal{R}\setminus\mathcal{S}} (i_r \leftrightarrow j_r)$$

so $M \Vdash \text{diff}(i, \mathcal{S}, j)$ iff $\mathcal{S} = \text{diff}(i, j)$.

Table 2: Recursion axioms for $\langle\#_{\theta_1,\theta_2}\rangle$ w.r.t. SNMs.

Name of axiom	Axiom in the language $\mathcal{L}^{\#}$
$(\#_{i_r})$	$\vdash \langle\#_{\theta_1,\theta_2}\rangle i_r \leftrightarrow i_r$
$(\#_{i\rhd^P j})$	$\vdash \langle\#_{\theta_1,\theta_2}\rangle i \rhd^P j \leftrightarrow \bigvee_{\{\mathcal{S}\subseteq\mathcal{R}:\frac{\lvert\mathcal{S}\rvert}{\lvert\mathcal{R}\rvert}\leq\theta_1\}} \text{diff}(i,\mathcal{S},j)$
$(\#_{i\rhd^N j})$	$\vdash \langle\#_{\theta_1,\theta_2}\rangle i \rhd^N j \leftrightarrow \bigvee_{\{\mathcal{S}\subseteq\mathcal{R}:\theta_1<\frac{\lvert\mathcal{S}\rvert}{\lvert\mathcal{R}\rvert}\leq\theta_2\}} \text{diff}(i,\mathcal{S},j)$
$(\#_\neg)$	$\vdash \langle\#_{\theta_1,\theta_2}\rangle\neg\varphi \leftrightarrow \neg\langle\#_{\theta_1,\theta_2}\rangle\varphi$
$(\#_\wedge)$	$\vdash \langle\#_{\theta_1,\theta_2}\rangle(\varphi\wedge\psi) \leftrightarrow (\langle\#_{\theta_1,\theta_2}\rangle\varphi \wedge \langle\#_{\theta_1,\theta_2}\rangle\psi)$
$(\#_{SPE})$	From $\vdash \psi_1 \leftrightarrow \psi_2$ infer $\vdash \varphi \leftrightarrow \varphi[\psi_2/\psi_1]$, with $\varphi[\psi_2/\psi_1]$ any formula obtained by replacing one or more occurrences of ψ_1 in φ with ψ_2.

Theorem 3.7. *The axiom system characterising the validities of the language $\mathcal{L}^{\#}$ in SNMs, provided by the propositional axiom system schema, the recursion axioms, and the rules in table 2, is sound and strongly complete.*

Proof. Soundness follows from the validity and validity-preserving properties of the axioms and rule (Smets and Velázquez-Quesada, 2020). In particular, the axiom $\#_{i_r}$ expresses that the opinions of the agents are not affected by the friendship selection operation, and the axioms $\#_{i \rhd^P j}$ and $\#_{i \rhd^N j}$ rewrite the definitions of the new positive and negative relations in the model, respectively. $\langle \#_{\theta_1,\theta_2} \rangle i \rhd^P j$, for instance, is true if and only if the distance between agents i and j, divided by the total number of propositions, is less than or equal to a threshold θ_1.
Similarly to the translation defined from $\mathcal{L}^\dagger$ to $\mathcal{L}$ (see theorem 3.4), completeness follows from the validity-preserving translation from $\mathcal{L}^\#$ to $\mathcal{L}$, defined by the axioms and the rule $\#_{SPE}$. □

3.3 Influential Nodes

In this thesis we define influential nodes in a social network model through a characterization of natural influencers. Recall the distinction between natural and designated influencers in social networks: whereas designated influencers obtain their influential power through assignment, natural influencers obtain (and may lose) influential power through their social position in a network. As opposed to the existing quantitative definitions, we provide a qualitative definition.
We define natural influencers in the language $\mathcal{L}^{\#\dagger} := \mathcal{L} + \langle \#_{\theta_1,\theta_2} \rangle + \langle \dagger_\tau \rangle$. This language requires no additional axioms besides those provided in tables 1 and 2 (Smets and Velázquez-Quesada, 2020).
In this section we define two notions of influencers, namely *rigid* and *flexible* influencers. Both are have the ability to increase the proportion of friends among their neighbours within a given time frame. However, whereas *rigid influencers* do so without letting the influence operation affect any of their opinions, *flexible influencers* are allowed to change some of their opinions provided that they do not change their position regarding at least one topic.
Furthermore, we investigate the duration of influencership: we define a first type of influencers who increase their proportion of friends in at least one update of the model (consisting of an act of social influence and an act of friendship selection), as well as a second type of influencers who keep increasing their proportion of friends during at least two updates. The definition of the first type of influencer considers only one instance of the social influence operation, during which agents are unable to reach agents beyond their nearest neighbours. Consequentially, this definition resembles the quantitative degree centrality measure, which considers nearest neighbours only, and tends to overvalue the influential power of agents with a large number of nearest neighbours, but only few next nearest neighbours (Chen et al., 2012). The definition of the second type of influencer is thus expected to highlight agents with higher influential power.
The above definitions can be formalized by making use of the following abbrevia-

tions. For $i \in \mathcal{A}, r \in \mathcal{R}$,

- $\text{inf}(i) := \bigvee_{\mathcal{B},\mathcal{B}',\mathcal{C},\mathcal{C}' \subseteq \mathcal{A}} (i \rhd^P \mathcal{B} \wedge i \rhd^N \mathcal{B}' \wedge \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle (i \rhd^P \mathcal{C} \wedge i \rhd^N \mathcal{C}') \wedge \frac{|\mathcal{C}|}{|\mathcal{C} \cup \mathcal{C}'|} > \frac{|\mathcal{B}|}{|\mathcal{B} \cup \mathcal{B}'|})$
- $\text{rigid}(i, r) := i_r \leftrightarrow \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle i_r$

Given a SNM $M = \langle N_P, N_N, V \rangle$ with $\mathcal{A}, \mathcal{T}, \mathcal{R}_T$, we have for $i \in \mathcal{A}$ in M that $M \Vdash \text{inf}(i)$ if and only if the proportion of friends of i among his neighbours increases within the following update, consisting of one act of social influence and one subsequent act of friendship selection. Furthermore, for $i \in \mathcal{A}, r \in \mathcal{R}$ in M we have $M \Vdash \text{rigid}(i, r)$ if and only if i does not change his position concerning r within the following update, consisting of one act of social influence and one subsequent act of friendship selection.
A rigid influencer increases their proportion of friends without changing his position on any proposition, within one update of the model.

Definition 3.8 (Rigid Influencer)**.** For $i \in \mathcal{A}$, define

$$RI(i) := \text{inf}(i) \wedge \bigwedge_{r \in \mathcal{R}} \text{rigid}(i, r)$$

so $i \in \mathcal{A}$ is a rigid influencer if $M \Vdash RI(i)$. Consequentially, i is a two-step rigid influencer if

$$M \Vdash RI(i) \wedge \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle RI(i).$$

A flexible influencer, on the other hand, increases their proportion of friends without changing his position on at least one proposition, whithin one update of the model.

Definition 3.9 (Flexible Influencer)**.** For $i \in \mathcal{A}$, define

$$FI(i) := \text{inf}(i) \wedge \bigvee_{r \in \mathcal{R}} \text{rigid}(i, r)$$

so $i \in \mathcal{A}$ is a flexible influencer if $M \Vdash FI(i)$. Consequentially, i is a two-step flexible influencer if

$$M \Vdash FI(i) \wedge \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle FI(i).$$

Generalized to n steps, we define the following:

Definition 3.10 (n-step Influencer)**.** For $i \in \mathcal{A}$, $n \geq 2$, i is a n-step rigid influencer if

$$M \Vdash RI(i) \wedge (\langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle)^{n-1} RI(i)$$

where

$$(\langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle)^1 RI(i) := \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle RI(i)$$

and

$$(\langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle)^{n+1} RI(i) := \langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle (\langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle)^n RI(i).$$

Similarly, i is a n-step flexible influencer if

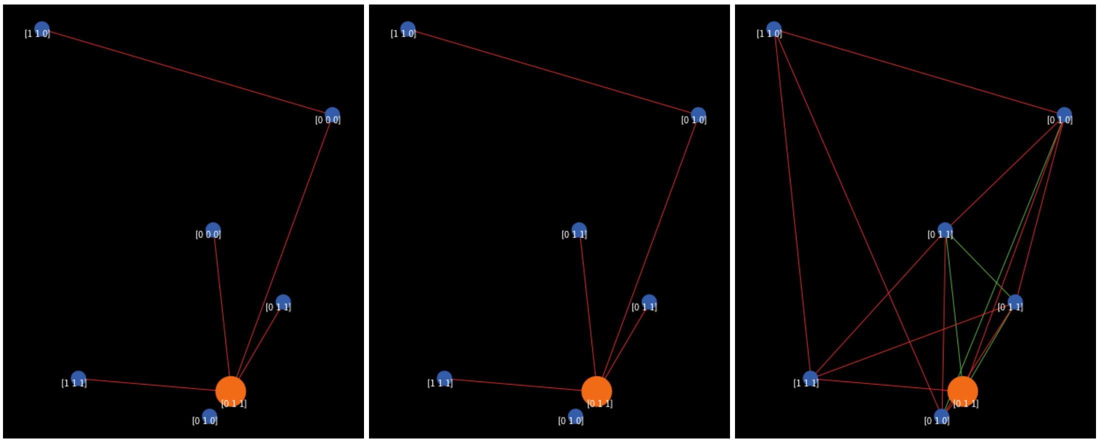
$$M \Vdash FI(i) \wedge (\langle \dagger_\tau \rangle \langle \#_{\theta_1,\theta_2} \rangle)^{n-1} FI(i).$$

3.4 Intertwining Dynamics with Influential Nodes

In the following example let $\theta_1 = 0$ and $\theta_2 = \frac{1}{3}$, as in configuration 3, and let $\tau = \frac{1}{2}$. Consider the SNM $M = \langle N_P, N_N, V \rangle$ with the ER random network structure shown in figure 3. $\mathcal{A}$ has length seven, the number of initial links is five and $\mathcal{T} = \{P, Q, R\}$ with $\mathcal{R}_P = \{p\}$, $\mathcal{R}_Q = \{q\}$, and $\mathcal{R}_R = \{r\}$. The following figures show the results of applying four rounds of social influence and friendship selection to the model. The figures are snapshots of a run of the NetLogo implementation of the model, of which a description is provided in section 3.6. The label of each agent represents his opinion, with indexes 0, 1, and 2 denoting p, q and r respectively. 1 at index i denotes support of the position represented by index i.
Before each update, the influencers at the current state of the model are computed and visualised. Influencers maintain their status until the update has completed and thus, the nodes visualized as influencers during any update are those that were considered influencers in the state of the model previous to that update. Though not visualized, each agent has a friendship relation with himself.

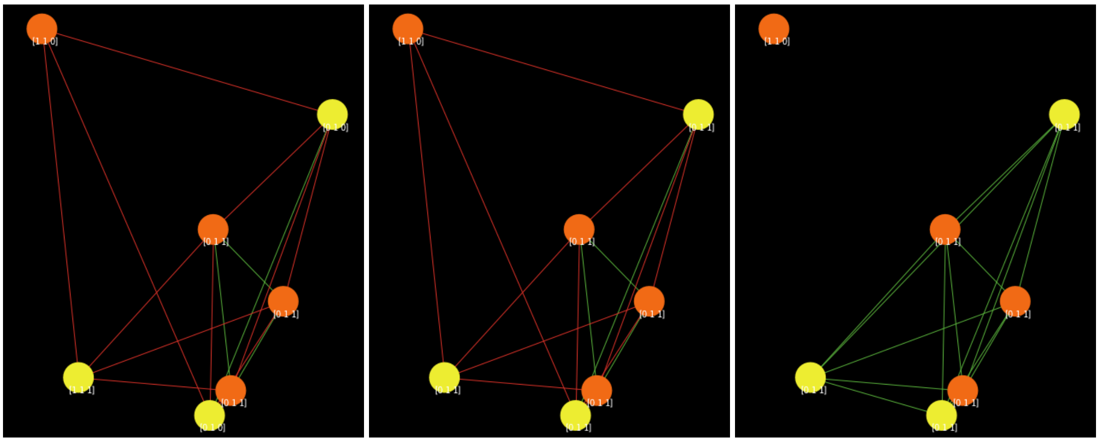
In the initial model (see figure 3a), the agent visualized in orange is a rigid influencer. Indeed, during update 1 he does not change his opinion (see figure 3b), but increases his proportion of friends (see figure 3c). The other agents change their opinions in accordance with their own neighbourhood. As $\theta_1 = 0$, only agents with identical opinions form friendships.

After the first update, each agent has become either a rigid (orange) or flexible (yellow) influencer. Two blocks are formed within the second update (see figure 4): the upper left agent keeps the opinions $[p, q, \neg r]$, while the other rigid influencers keep the opinions $[\neg p, q, r]$, which are adopted by the flexible influencers. First,



(a) M (b) $M^{\dagger(\frac{1}{2})}$ (c) $(M^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})}$

Figure 3: Initial model (a), the intermediate result of update 1 (b), and the final result (c).

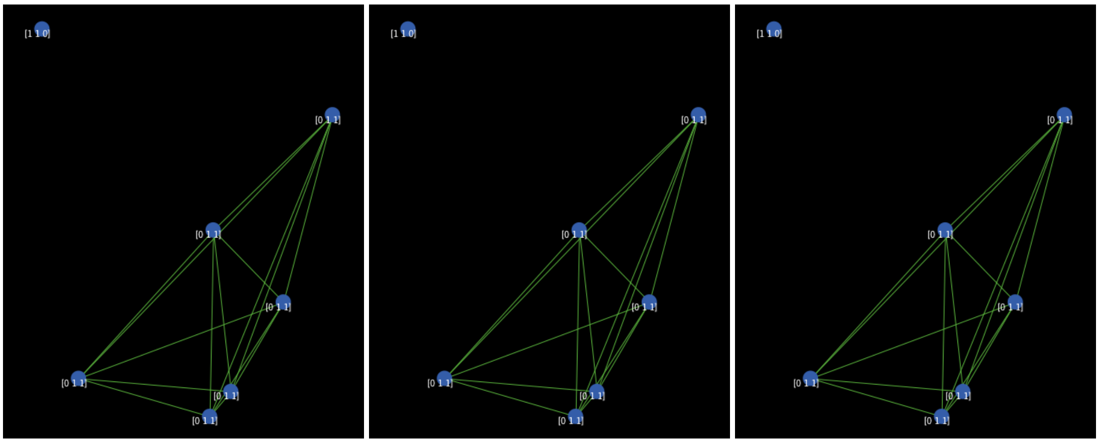


(a) $(M^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})}$ (b) $((M^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})})^{\dagger(\frac{1}{2})}$ (c) M'

Figure 4: The state of the model after update 1 (a), the intermediate result of update 2 (b), and the final result, $M' := (((M^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})})^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})}$ (c).

note that the upper left agent is an influencer, because friendship relations are reflexive by definition 3.6, and thus, he increases his proportion of friends from $\frac{1}{4}$ to 1 within update 2. Second, note that after the update, the distance between the two existing opinion sets is too large for acquaintanceships to be formed between the two blocks (see figure 4c). As a consequence, the upper left agent becomes isolated from the component.

The third update (see figure 5) does not affect the model: as consensus was reached within the component of agents with opinions $[\neg p, q, r]$, none of these



(a) M' (b) $(M')^{\dagger(\frac{1}{2})}$ (c) $((M')^{\dagger(\frac{1}{2})})^{\#(0,\frac{1}{3})}$

Figure 5: The state of the model after update 2 (a), the intermediate result of update 3 (b), and the final result (c).

agents can be caused to change their opinion. Since the upper left agent is isolated, he will not change his opinion by definition. Thus, the social relations in the model will remain unchanged as well: the model has stabilized. Since none of the agents are able to increase their proportion of friends, no influencers are detected before the second update.

Note that the only two-step influencer in the model (the orange node in figure 3) is one of only two agents who have the opinions $[\neg p, q, r]$ in the initial state of the model, and then isolates the only agent who has not been influenced to adopt these opinions within two updates. This observation may support the hypothesis that two-step influencers have more influential power over a network than one-step influencers.

3.5 Balance Theorem

Polarization in our social network model is measured by means of the Balance Theorem (Easley and Kleinberg, 2019): the network is balanced if and only if a balanced division exists, that is, if the network can be divided into two groups X and Y of agents, such that there exist no negative edges within X or within Y, and there exist no positive edges between X and Y. The variation of the theorem defined for non-complete networks is used, as the defined social network is generally non-complete.

Definition 3.11 (Balanced Division (Non-Complete Networks))**.** A balanced division of a non-complete SNM $M = \langle N_P, N_N, V \rangle$ with a set $\mathcal{A}$ of agents is a division of $\mathcal{A}$ into $X \subseteq \mathcal{A}$ and $Y = \mathcal{A} \setminus X$, such that

1. for any $i, j \in X$ we have $\neg(i \rhd^N j)$, and
2. for any $i, j \in Y$ we have $\neg(i \rhd^N j)$, and
3. for any $i \in X, j \in Y$ we have $\neg(i \rhd^P j)$.

The following propositions concern particular SNMs and configurations, and justify parts of the Python and NetLogo implementations of the balance measure. In this thesis, we restrict our attention to the set $\mathcal{M}$ of SNMs such that for each SNM $M = \langle N_P, N_N, V \rangle$ with $M \in \mathcal{M}$, we have the associated set of topics $\mathcal{T} = \{P, Q, R\}$ with positions $\mathcal{R}_P = \{p\}$, $\mathcal{R}_Q = \{q\}$, and $\mathcal{R}_R = \{r\}$. Consequently, each topic $T \in \mathcal{T}$ yields two possible opinions: an agent may or may not adopt the position associated with a topic. The following propositions hold only for SNMs in $\mathcal{M}$.
Furthermore, with respect to friendship selection, we consider only three configurations of θ_1 and θ_2: we define (1) $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{2}{3}$, (2) $\theta_1 = 0$ and $\theta_2 = \frac{2}{3}$, and (3) $\theta_1 = 0$ and $\theta_2 = \frac{1}{3}$.
Proposition 3.12 also holds for any other configuration of θ_1 and θ_2 besides 1, 2 and 3. The others specify the configurations concerned.

Proposition 3.12. Let $M = \langle N_P, N_N, V \rangle$ be a connected[5] SNM such that $M \in \mathcal{M}$, with a set $\mathcal{A}$ of agents. If M is balanced, then there exists exactly one balanced division of $\mathcal{A}$.

Proof. Let $M = \langle N_P, N_N, V \rangle$, $M \in \mathcal{M}$, be a SNM and suppose M is balanced. We construct the unique division of $\mathcal{A}$ into X and Y as follows. Let $a \in \mathcal{A}$ be arbitrary and let $a \in X$. We show that for any agent $b \in \mathcal{A}$, b is then forced to be placed in either group X or Y. To show this let $b \in \mathcal{A}$, $a \neq b$, be arbitrary. As M is connected, there exists a path $a(\rhd^{P \cup N})^+ b$ (where $(\rhd^{P \cup N})^+$ denotes the transitive closure of $\rhd^{P \cup N}$). We prove by induction that for any length n of the path, b is forced into one group.
For $n = 1$, we have $a \rhd^{P \cup N} b$. By definition, either $a \rhd^P b$ or $a \rhd^N b$. In the former case, by $a \in X$ we have $b \in X$, as X and Y cannot be connected by a positive edge. In the latter case, we have $b \in Y$, as two agents within a group cannot be connected by a negative edge.
For the induction step, suppose that for any path $a(\rhd^{P \cup N})^+ i$ of length n, i is forced into a group. Consider $a(\rhd^{P \cup N})^+ b$ with length $n + 1$. Then there is some $c \in \mathcal{A}$ such that we have $a(\rhd^{P \cup N})^+ c \rhd^{P \cup N} b$ with $a(\rhd^{P \cup N})^+ c$ having length n. Then by the induction hypothesis, c is forced into a group. Suppose without loss

[5] Given a *signed connected* graph M containing positive and negative binary relations $\rhd^P$ and $\rhd^N$ respectively, and a set of agents $\mathcal{A}$, there exists a path $a(\rhd^{P \cup N})^+ b$ between any two nodes $a, b \in \mathcal{A}$, where $(\rhd^{P \cup N})^+$ denotes the transitive closure of $\rhd^{P \cup N}$.

of generality that $c \in X$. Now we have either $c \rhd^P b$ or $c \rhd^N b$. Following the argument from the base case, in the former case we have $b \in X$, whereas in the latter we have $b \in Y$. The cases for $c \in Y$ are analogous. Thus, in any case, with a path $a(\rhd^{P\cup N})^+b$ of length $n+1$, b is forced into one group.
Therefore, we have for any length of the path $a(\rhd^{P\cup N})^+b$, that b is forced into one group. As b was arbitrary and as M is connected, we have for all $b \in \mathcal{A}$, $a \neq b$, that b is forced into a group by the placement of an arbitrary agent a. Thus, we have a unique balanced division of $\mathcal{A}$ into two groups X and Y. □

Both proposition 3.13 and proposition 3.14 concern only SNMs $M \in \mathcal{M}$ with θ_1 and θ_2 configured as in 1 or 2, that is, configurations such that two agents a and b do not form a relation if and only if $\text{diff}(a, b) = 3$.

Proposition 3.13. Let $M = \langle N_P, N_N, V\rangle$ be a disconnected SNM generated by at least one round of social influence and friendship selection, consisting of exactly two components $X \subseteq \mathcal{A}$ and $Y = \mathcal{A} \setminus X$. Let $M \in \mathcal{M}$ and let θ_1 and θ_2 be configured as in 1 or 2. Then both X and Y are fully connected components[6] and M is balanced, with X and Y providing the balanced division of $\mathcal{A}$.

Proof. Let $M = \langle N_P, N_N, V\rangle$ be a SNM generated by at least one round of social influence and friendship selection, with $M \in \mathcal{M}$, θ_1 and θ_2 configured as in 1 or 2, with a set $\mathcal{A}$ of agents. Suppose M consists of two components $X \subseteq \mathcal{A}$ and $Y = \mathcal{A} \setminus X$. We show that X and Y provide a balanced division of $\mathcal{A}$.
We distinguish two cases.

- Case 1: M is the result of an act of friendship selection.
  We first show that for all $i \in X, j \in Y$ we have a distance $\text{dist}(i, j)$=3, for all $i, j \in X$ we have $\text{dist}(i, j)$=0, and for all $i, j \in Y$ we have $\text{dist}(i, j)$=0.
  - Claim: for all $i \in X, j \in Y$ we have $\text{dist}(i, j)$=3.
    Proof of claim: by $M \in \mathcal{M}$, we have for both configurations 1 and 2 of θ_1 and θ_2 that a relation is formed between two agents a and b if and only if $\text{dist}(a, b)$<3. This follows from having $\theta_2 = \frac{2}{3}$ in both configurations: as $M \in \mathcal{M}$, we have three propositions $\mathcal{R} = \{p, q, r\}$. Then by definition of the friendship selection update, the maximum distance allowed for an acquaintance relation $a \rhd^N b$ is comprised of two of the three propositions, i.e. $\text{dist}(a, b)$=2. As the friendship relation requires a smaller distance by definition, $\text{dist}(a, b)$=2 is the maximum distance for a and b to form a relation $a \rhd^{P\cup N} b$. As there exists no edge connecting the components X and Y, we have for all $i \in X, j \in Y$

[6] A *fully connected*, or *complete*, component is a component of a graph which is complete when considered as a graph in isolation (Easley and Kleinberg, 2019).

that $\neg(i \rhd^{P\cup N} j)$ and thus, for all $i \in X, j \in Y$ we have dist$(i,j)> 2$, which, having three propositions, means dist(i,j)=3.

  - Claim: for all $i,j \in X$ we have dist(i,j)=0, and for all $i,j \in Y$ we have dist(i,j)=0.
    Proof of claim: let $a \in X$ be arbitrary and suppose without loss of generality that a supports positions p,q and r, i.e., suppose $M \Vdash a_p \wedge a_q \wedge a_r$. Now let $b \in Y$ be arbitrary. As we have shown, it follows that dist(a,b)=3. Thereby, we have $M \Vdash \neg b_p \wedge \neg b_q \wedge \neg b_r$. As b was chosen arbitrarily, we have $M \Vdash \bigwedge\limits_{b\in Y}(\neg b_p \wedge \neg b_q \wedge \neg b_r)$. But then, as for all $i \in X, j \in Y$ we have dist(i,j)=3, we must have $M \Vdash \bigwedge\limits_{a\in X}(a_p \wedge a_q \wedge a_r)$.
    Thus, we have both dist(i,j)=0 for all $i,j \in X$ and for all $i,j \in Y$. The cases for other stances with respect to positions $\mathcal{R}$ are analogous.

  We now show that X and Y meet the three requirements of a balanced division.
  For the first requirement, let $a,b \in X$ be arbitrary. As we have dist(a,b)=0, and given that M is the result of a friendship selection update, we have $a \rhd^P b$, so $\neg(a \rhd^N b)$. Thus, we have for all $a,b \in X$ that $\neg(a \rhd^N b)$, i.e., there exist no negative edges between any two agents within X. The case for the second requirement, which is for no negative edges to exist between any two agents within Y, is analogous.
  For the third requirement, let $a \in X, b \in Y$ be arbitrary. With X and Y being components, we have $\neg(a \rhd^{P\cup N} b)$ and thus, $\neg(a \rhd^P b)$. As $a \in X, b \in Y$ were arbitrary, we have no positive edges between X and Y.
  Therefore, if M is the result of an act of friendship selection, then by definition of a balanced division, X and Y provide a balanced division of $\mathcal{A}$.
  Furthermore, X and Y are both fully connected components, as for arbitrary $a,b \in X$ we have $a \rhd^P b$ and for arbitrary $a,b \in Y$ we have $a \rhd^P b$.

- Case 2: M is the result of an act of social influence.
  Consider the SNM M' such that $M = (M')^{\dagger(\tau)}$ for some $\tau \in [0,1]$, i.e., consider the state of the model prior to the social influence update (this state exists, as the model is by assumption generated by at least one round of social influence and friendship selection). As the social influence update does not affect existing relations between agents, we know that M' is disconnected as well as M, with the same components X and Y. As M' is the result of a friendship selection update, case 1 is in effect. This means that X and Y provide a balanced division of $\mathcal{A}$. As X and Y remain unchanged during the transformation of M' into M, X and Y also provide a balanced division of $\mathcal{A}$ for M.

X and Y being fully connected components also follows from the social influence update not affecting existing relations between agents: as they were fully connected components in M', they are in M as well.

$\square$

Proposition 3.14. Let $M = \langle N_P, N_N, V\rangle$ be a SNM generated by at least one round of social influence and friendship selection, with $M \in \mathcal{M}$ and θ_1 and θ_2 configured as in 1 or 2. Then either M is connected, or M consists of exactly two components, which are both fully connected.

Proof. Let $M = \langle N_P, N_N, V\rangle$ be a SNM generated by at least one round of social influence friendship selection, with $M \in \mathcal{M}$, θ_1 and θ_2 configured as in 1 or 2, and a set $\mathcal{A}$ of agents. Again, we distinguish two cases.

- Case 1: M is the result of an act of friendship selection.
  If M is connected we are done, and if M consists of less than two components, M is connected as well and we are done; so suppose that M is disconnected and suppose for a contradiction that M contains more than two components. Let $X, Y \subseteq \mathcal{A}$, $X \cap Y = \emptyset$, be two components of M. As we proved for proposition 3.13, we know that with $M \in \mathcal{M}$ and $\theta_2 = \frac{2}{3}$, we have by definition of the friendship selection update that $\text{dist}(i,j)$=3 for all $i \in X, j \in Y$ and $\text{dist}(i,j)$=0 for all $i, j \in X$ as well as all $i, j \in Y$. Let $a \in X, b \in Y$ be arbitrary. Then we have $\text{dist}(a,b)$=3. Now consider a third component $Z \subseteq \mathcal{A}$ and let $c \in Z$. Then, again by proposition 3.13, we have $\text{dist}(a,c)$=3 as well as $\text{dist}(b,c)$=3. However, by $M \in \mathcal{M}$, one of only two opinions is possible for each topic T. Thereby, $\text{dist}(a,b)$=3 and $\text{dist}(b,c)$=3 give us $\text{dist}(a,c)$=0: we have reached a contradiction. Thus, Z cannot exist. Therefore, M consists of exactly two components X and Y. By proposition 3.13, X and Y are fully connected.

- Case 2: M is the result of an act of social influence.
  Consider the SNM M' such that $M = (M')^{\dagger(\tau)}$ for some $\tau \in [0,1]$, i.e., consider the state of the model prior to the social influence update (this state exists, as the model is by assumption generated by at least one round of social influence and friendship selection). As the social influence update does not affect existing relations between agents, we know that M' is disconnected as well as M, and composed of the same components. As M' is the result of a friendship selection update, case 1 is in effect: M has exactly two components X and Y, which are, by proposition 3.13, both fully connected. As X and Y remain unchanged during the transformation of M' into M, M consists of the same two fully connected components.

□

Corollary 3.15. Let $M = \langle N_P, N_N, V \rangle$ be a SNM generated by at least one round of social influence and friendship selection, with $M \in \mathcal{M}$ and θ_1 and θ_2 configured as in 1 or 2. If M is disconnected, then M is balanced.

Proof. Let $M = \langle N_P, N_N, V \rangle$ be a SNM generated by at least one round of social influence and friendship selection, with $M \in \mathcal{M}$ and θ_1 and θ_2 configured as in 1 or 2. Suppose M is disconnected. Then by proposition 3.14, M consists of exactly two fully connected components X and Y. Thus, by proposition 3.13, M is balanced with X and Y providing a balanced division. □

3.6 Implementation and Visualization

Two implementations of the social network model were constructed; the initial implementation was written in Python and contains only the main algorithms. These were rewritten and extended in the final implementation, which was written in the multi-agent modelling environment NetLogo and provides a visualization. Both implementations were defined specifically for social network models in the set $\mathcal{M}$ and thus, the set of topics $\mathcal{T} = \{P, Q, R\}$ and their respective positions $\mathcal{R}_P = \{p\}$, $\mathcal{R}_Q = \{q\}$, and $\mathcal{R}_R = \{r\}$ are fixed in both implementations.
In this section we provide a description of the NetLogo implementation, focusing on the main algorithms: the social influence update, the friendship selection update and the polarization measure. A description of the Python implementation is provided in Appendix A.
The NetLogo implementation runs two main functions: `setup` and `go`. The former resets the visualization and manages the setup for a new run. The latter is the main function of the program, which orders the algorithms and is designed to iteratively execute the program; each run of the `go` function is called a `step`. A built-in `tick` variable keeps track of the number of completed steps in one run.

3.6.1 Setup and Go

The `setup` function of our model resets all variables, generates the specified network and determines the shape and opinions of the `turtles` (agents). The network type, thresholds, and numbers of turtles and links are managed on the user interface (see figure 6). The network types include the ring, star, double star, fully connected and ER random network; the influence threshold (τ) takes a value in the range [0,1] and the friendship selection thresholds θ_1 and θ_2 take values in accordance with configurations 1, 2 and 3.
The ER random network can be run with a number of nodes in the range [1,500] and a number of links in the range [0,800]. As the other network types have a

fixed structure which does not change with the number of nodes and links, and where the number of nodes determines the number of links (consider, for example, the fully connected network), these parameters are not expected to highly affect the behaviour of these networks. Therefore, the ring, star, double star, and fully connected network each run with a fixed number of ten nodes and the according number of links depending on the network type. Each network is initiated with the existing links representing acquaintanceships.
Each turtle is randomly assigned an opinion on each topic by means of a "coin flip": for each turtle, a randomly chosen number from the set $\{0, 1\}$ (of which 1 denotes support of the associated position) is stored for each of the positions p, q and r.
The `go` function organizes the algorithms of which the model is composed: at every `step`, first the influencers at the current state of the model are identified and visualized. Next, the social influence update is performed, followed by the friendship selection update. Afterwards, an attempt is made to create a balanced division; if the attempt succeeds, the network has polarized and the created groups X and Y are visualized. Furthermore, in case of polarization, the components composing the network are stored (where a fully connected network consists of only one component), and for each component, in case of consensus, the mutually shared opinion is stored.

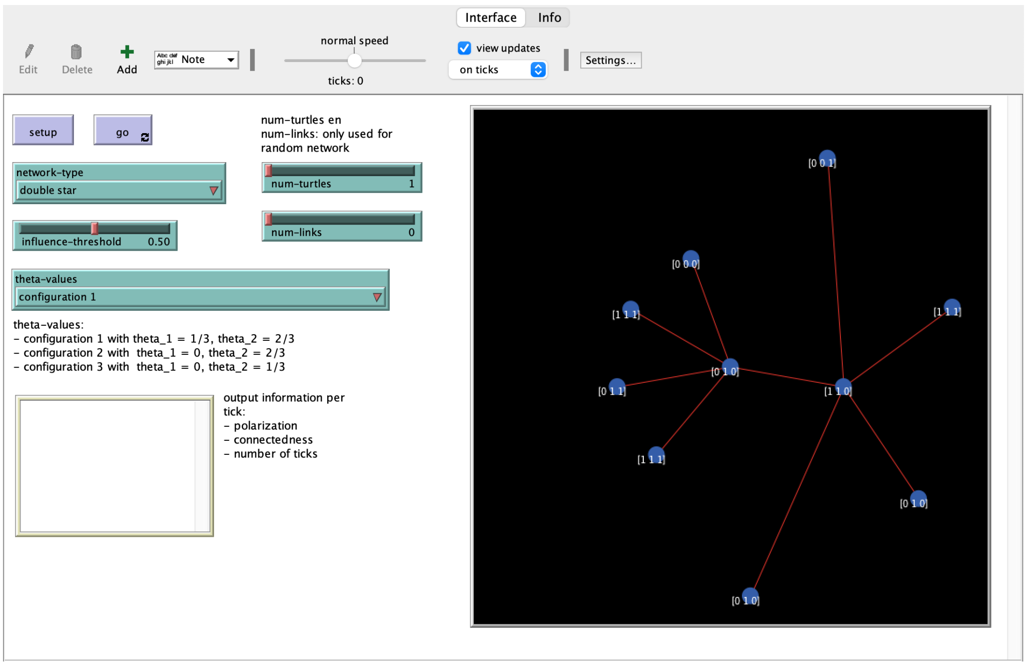


Figure 6: NetLogo user interface after running the `setup` button with the currently selected parameter values.

3.6.2 Social Influence

The social influence update requires a simultaneous implementation. To ensure this, the future positions of each turtle are computed based on current opinions and temporarily stored. Afterwards, the turtles simultaneously adopt their updated positions.
Furthermore, as the NetLogo environment prohibits reflexive relations by design, the reflexive nature of friendship relations in the social network model is explicitly taken into account in the implementation of the update function: during computation of the support of a position in the neighbourhood of a turtle, the number of relations of that turtle is increased by 1, and if the turtle initially supports the given position, the support is increased by 1 as well.

3.6.3 Friendship Selection

The friendship selection update asks all turtles to update their relations with the other turtles in the network, computing for each pair of turtles if the requirements (see definition 3.6) for either a friendship or an acquaintanceship are met. The function uses the built in `ask` function of NetLogo to iterate over the agents during the update: this function asks a specified set of agents to run a specified set of commands. The main update function is as follows:

```
to perform-friendship-selection
  ask turtles [
    ask other turtles [
        update-relation myself self
    ]
  ]
end
```

The built-in variable `myself` indicates the turtle giving the command (one of `turtles`) and `self` indicates the turtle executing the command (one of `other turtles`). Thus, all turtles are sequentially asked to update their relation with all other turtles in the network. This algorithm updates a number of relations multiple times. Although this construction would be computationally complex in a Python implementation, NetLogo specifically designed the `ask` function to efficiently have sets of agents perform series of commands, while the NetLogo equivalent of a for loop (the `foreach` function) is only suited to lists (Wilensky, 1999).
Note that the updates are not affected by this sequential construction: an updated relation between two agents cannot be affected by other previously updated relations in the model, as a relation update is solely based on the positions of the two agents in question, and these positions are not affected by the relationship update.

For the same reason, a relation that is updated multiple times within one update will not change after its first update. In other words, the update is idempotent (Smets and Velázquez-Quesada, 2020).

3.6.4 Identifying Influencers

The algorithm computes for all turtles $i \in \mathcal{A}$ if $\inf(i)$ is true (see definitions 3.8 and 3.9) and if so, computes if i is rigid or flexible. Furthermore, for each turtle, the numbers of consecutive steps during which they have been a rigid and flexible influencer, respectively, is stored.

Influencers are visualized by larger dots and different colours: an orange dot denotes a rigid influencer, while a yellow dot denotes a flexible influencer.

3.6.5 Polarization Measure

The implementation of the balance measure makes use of breadth-first search[7] in order to determine if the network is connected, and to efficiently traverse the network while attempting to find a balanced division. First, starting from an arbitrary turtle in the network, breadth-first search is used to order the component it is connected to in layers. If the set of turtles occurring in the constructed layers equals the total set of turtles in the network, the network is connected.

In case of a connected network, we have by proposition 3.12 that if the network is balanced, there exists only one balanced division into two groups X and Y. This division is created using the layers constructed through breadth-first search: after placing the root node (the turtle forming the first layer) in group X, the algorithm iterates over the turtles in the constructed layers while placing them in either group X or Y following definition 3.11. If no contradiction occurs, the network is balanced.

In case of a disconnected network, the taken approach differs for each of the defined configurations of θ_1 and θ_2.

Consider a disconnected SNM $M \in \mathcal{M}$ in configuration 1 or 2, that is, a situation such that either (1) $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{2}{3}$, or (2) $\theta_1 = 0$ and $\theta_2 = \frac{2}{3}$. By corollary 3.15, if M is generated by at least one round of friendship selection and social influence, then M is balanced. By propositions 3.13 and 3.14, M consists of exactly two fully connected components X and Y, which provide the balanced division of M. As the layered component forms one of these two components, there must be one complementary component which includes the turtles that are not present in

[7] *Breadth-first search* is a method to organize a graph by arranging the nodes based on their distance from a given node. With this node forming the first layer, successive layers are composed of the nodes that (1) do not occur in previous layers and (2) have an edge to a node in the previous layer (Easley and Kleinberg, 2019).

the layered component. Since the two components provide a balanced division, the turtles in the layered component are assigned to group X and the others are assigned to group Y.
In case of configuration 3, where $\theta_1 = 0$ and $\theta_2 = \frac{1}{3}$, the network may consist of more than two components. Furthermore, contrary to configurations 1 and 2, turtles within one component are not necessarily friends. A balanced division for this configuration is created as follows: first, the turtles in the layered component are placed into groups X and Y as though the component were a connected network. If no contradiction occurs, the algorithm then iterates over those not included in the layered component and places them in either group X or Y: if a turtle is not related to any previously placed turtles, it is automatically placed in X, unless Y is still empty. If it is related to any previously placed turtles, the turtle is placed in accordance with definition 3.11. If no contradiction occurs, the network is balanced.
In any case, if the network is balanced, it has polarized and the two groups are visualized by colouring the agents in group X and Y green and red respectively.

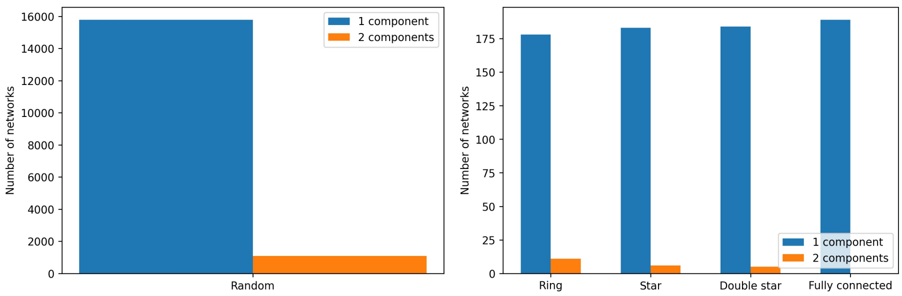
3.7 Evaluation

In order to explore the relation between the several parameters of the model, the influencers and polarization of the model, a number of simulations are executed using the BehaviorSpace software tool integrated with NetLogo. BehaviorSpace systematically varies the parameters of the model, storing the results in a dataset. Five experiments are run, each concerning a different model structure. Each experiment is run for the duration of eleven steps (0 through 10) with the following varying parameter values: the influence threshold values vary in the range [0,1], incrementing by 0.05, and theta-configurations 1, 2, and 3 are considered. Additionally, the ER random network is run with varying numbers of turtles and links. The former varies in the range [1,500], incrementing by 50; the latter varies in the range [0,800], incrementing by 100. Each combination of parameter values is run three times, as the random assignment of opinions to the turtles generally results in different outcomes per run. Considering all possible combinations and repeating those three times results in 17,010 runs of the ER random network experiment and 189 runs of each of the other experiments. Simple statistical analysis is used to interpret the resulting data.

4 Results

Figures 7-9 concern the development of polarization and connectedness of SNMs throughout one run. As shown in figure 7, the majority of networks that polarize

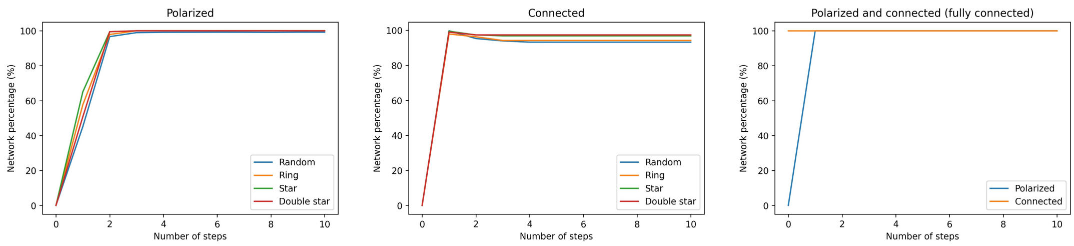
within eleven steps finish as a connected network. Moreover, none of the initially fully connected networks result in a disconnected network.
None of the network simulations in this experiment result in a disconnected network composed of more than two components, even though that result is theoretically possible for networks with θ-configuration 3.



(a) ER Random network (b) Other

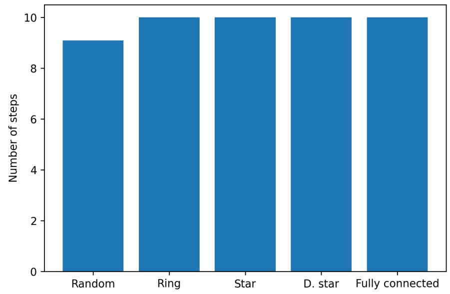
Figure 7: Connectedness among networks that polarize within one run, per structure.

Figure 8 shows that the majority of networks of each structure polarize within the first three steps (steps 0-2). As shown in figure 9b, for networks ending the run with consensus in each component, it takes roughly the same number of steps for the eventually consensual opinions to emerge. However, figure 9a shows that the agents take up to the final run to settle in their final component.

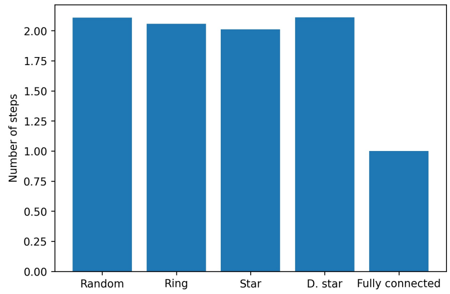


(a) Polarized networks (b) Connected networks (c) Fully connected

Figure 8: The percentages of connected (a) as well as polarized (b) networks, given a step in the run. The fully connected networks (c) stay connected throughout the entire run.



(a) Final components



(b) Final consensual opinions

Figure 9: The mean number of steps required for a network ending the run with consensus in each component, to arrange its nodes into their final components (a) and to develop eventually consensual opinions for existing components (b).

The configurations of the θ-values did not cause significant differences in the polarization rate of the networks. They did, however, affect the percentage of networks becoming connected within a run (see figure 10): runs with configuration 3 report a significantly lower percentage of networks resulting in one component, than those with configurations 1 and 2 for all network structures. No significant differences were found between configurations 1 and 2.

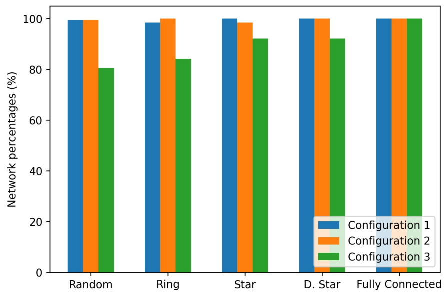


Figure 10: The percentage of connected networks after one run, given the θ-configuration, for each network structure.

Figures 11 and 12 present the effect of different τ-values. Figure 11 shows a decrease of almost 20% in connectedness in ER random networks for values between 0.2 and 0.8, with its low points at values 0.4 and 0.6. A decrease of 5% in polarization was found for the value 0.5. The ring structure shows a significant decrease in connectedness for the same values, though polarization is unaffected. The other structures appear to be unaffected by the threshold values.

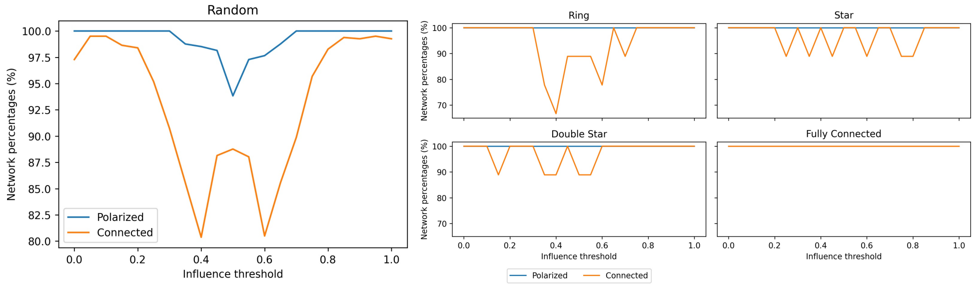


Figure 11: The percentage of networks that polarize within one run, as well as the percentage that is connected after one run, for each structure, given the influence threshold.

Furthermore, figure 12 shows that ER random networks that do polarize with value 0.5 take, on average, roughly 0.4 more steps to polarize than with other values. The ring structure shows a significant increase for the center values as well.

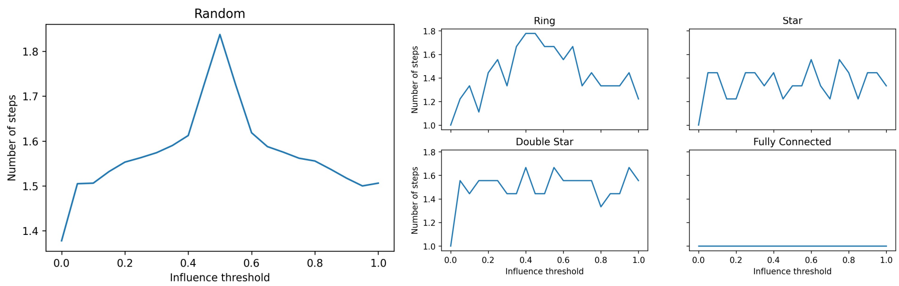
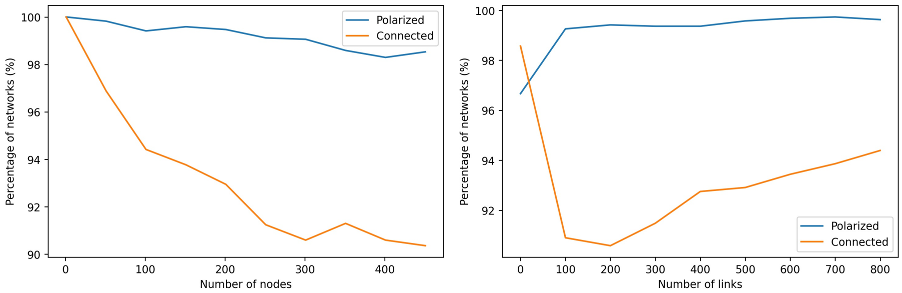


Figure 12: The mean number of steps required for each network structure to polarize given the influence threshold.

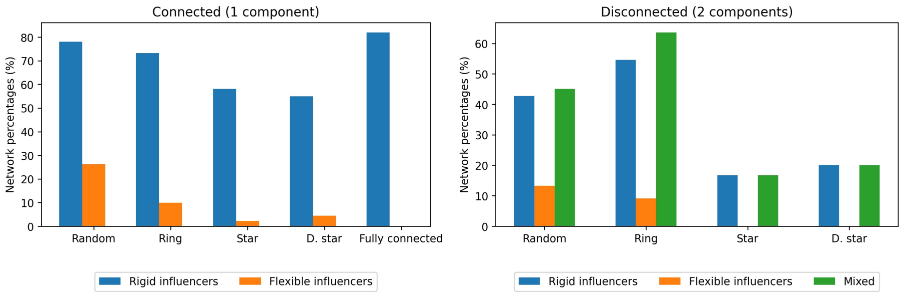
The numbers of nodes and links were varied solely for the ER random network. Figure 13a shows that polarization and connectedness decrease with a larger number of nodes. An initially larger number of links has the opposite effect on polarization (see figure 13b). Connectedness appears to decrease with relatively small numbers of links and increase with amounts larger than 200 links.



(a) Number of nodes (b) Initial number of links

Figure 13: The percentage of random networks that polarize, as well as the percentage of random networks that become connected, within one run, given the number of nodes (a) and the initial number of links (b).

Figures 14 and 15 present the relations between influencers and polarization. Of the polarized networks we consider only the extreme cases of polarization: these are the runs resulting in a balanced division of the network into either one or two components, such that there is consensus in each component. We will call these *extremely polarized networks*.

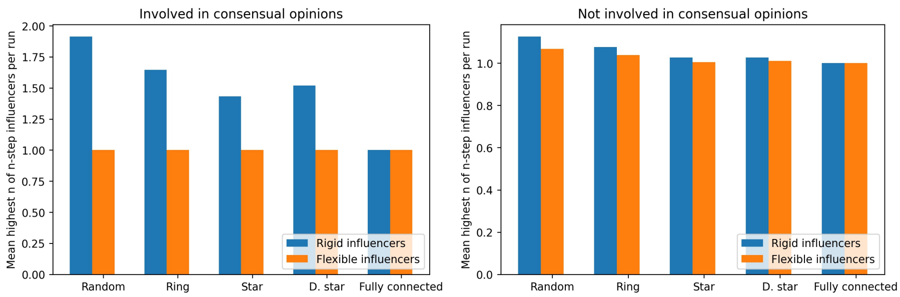


(a) Connected (b) Disconnected

Figure 14: (a) The percentage of extremely polarized connected networks in which the consensual opinion was adopted by one or more influencers within the first two steps; and (b) the percentage of extremely polarized disconnected networks in which for each of the consensual opinions, the respective component contains at least one influencer who adopted the opinion within the first two steps.

Figure 14 shows the relation between opinions adopted by influencers and the final consensual opinions in extremely polarized networks. Only opinions adopted within the first two steps of the run are considered, as the first influencers who adopt an opinion which will eventually become consensual are expected to be the most powerful. The fully connected network is left out of figure 14b, as none of the runs with this structure resulted in a disconnected network.

It is apparent that the fully connected structure has the highest percentage of runs in which at least one rigid influencer shares the final consensual opinion within the first two steps. The percentages concerning flexible influencers are significantly lower for all structures.
As shown in figure 14b, overall, the ring structure has the highest percentages of runs with consensual opinions previously shared by influencers. Again, the percentages concerning flexible influencers are significantly lower. The percentages of mixed runs, which disregard the kind of influencer sharing the opinion, are comparable to those concerning rigid influencers. This indicates that there are only few runs in which only one component contains a rigid influencer sharing the final opinion and the other component contains a flexible influencer meeting the same requirements.



(a) Involved in consensus (b) Not involved in consensus

Figure 15: The mean highest number n of n-step influencers occurring in one run in extremely polarized networks, which do (a), or do not (b) share the final consensual opinion in their final component.

Finally, figure 15 shows the mean of the highest number n per run, such that at least one n-step influencer exists in that run which either do (see figure 15a), or do not (see figure 15b) share the final consensual opinion in their final component, for all runs resulting in extremely polarized networks. Interestingly, on average, rigid influencers who do share the final consensual opinion in their component maintain their position significantly longer than others. No significant differences were found in the flexible influencers.

5 Conclusions

In this thesis we examined under which conditions polarization is more likely to occur and what role natural influencers play in the process. To this end we have proposed a threshold model for social networks based on the model provided by Smets and Velázquez-Quesada (2020), as well as a qualitative definition of natural

influencers. We implemented the social network model in Python before writing the final implementation in the NetLogo modelling environment, in which we ran several simulations of the model. Using a qualitative measure for group polarization, we measured polarization rates of the model under different circumstances. Interestingly, the majority of networks polarized within the first three steps. Among the networks polarizing within one run (consisting of eleven steps), the majority resulted in a connected network. Nevertheless, given θ-configuration 3, a significantly higher percentage of polarized networks resulted in disconnected networks consisting of two opposing components. It can therefore be suggested that requiring further agreement between agents for the formation of relations increases the probability of a network falling apart into two opposing components.
Furthermore, τ-values in proximity of 0.5 appeared to delay and decrease polarization in random networks. The same values decreased connectedness in random networks as well as ring-structured networks. These results, however, are likely to be related to our definition of social influence, which states that a position with respect to a topic is *adopted* if a threshold is exceeded. Given our choice for binary opinions on topics, the threshold limits are skewed towards adoption and rejection, respectively: threshold values approaching 0 will generally cause the agents to eventually adopt all available positions. Similarly, threshold values approaching 1 will generally cause agents to reject all positions. This situation can either be averted by avoiding binary opinions or by changing the definition for social influence; with respect to the former solution, the social network model is designed to account for various sets of topics and according positions. With respect to the latter, a more suitable definition would state that a current opinion is *changed* if a threshold is exceeded. The NetLogo code can straightforwardly be adapted to account for the former solution by following the instructions provided under the "Info" tab of the implementation; the latter requires a different implementation of the social influence update.
Another important finding was that rigid influencers appear to play a more explicit role in polarization than flexible influencers: in more than half of extremely polarized networks that finished connected, the final opinion set had been adopted by at least one rigid influencer within the first two steps of the run. With respect to flexible influencers, this was the case in only a few networks. Furthermore, rigid influencers meeting these requirements tended to maintain their influencer status significantly longer than flexible influencers. This result is surprising, as flexible influencers were expected to keep increasing their proportion of friends for a longer period of time, due to their ability to partly adapt to others.
Contrary to expectations, both kinds of influencers appear to be less influential in extremely polarized networks which end the run consisting of two opposing components. There are several possible explanations for this result. It could be

hypothised that disagreeing influencers are unlikely to coexist until the end of the run. Another possible explanation is that the probability of finding one influencer having adopted a final consensual opinion is simply greater than that of finding two influencers each having adopted different consensual opinions. To develop a full picture of the role of influencers in polarization, additional studies will be needed which test the influential power of influential nodes through interventions on their opinions. Although NetLogo allows commands issued from the command line during a run, it may not be suited to interventions, due to the high speed at which updates are performed.
It might be the case that flexible influencers, as they are currently defined, apply a more subtle form of influence than rigid influencers, by causing the other agents to change only a subset of their opinions. The examination of the development of opinions of flexible influencers is therefore left for future work.
Moreover, as flexible influencers appear to be less influential than expected, further constraints on their behaviour might be necessary. The same holds true for rigid influencers, since the provided definitions allow influencers who see an overall decrease in related agents (see section 3.4), provided that the proportion of their friends increases. A further study with more focus on the qualitative definitions of influencers is therefore suggested. For the NetLogo implementation to run a different definition for flexible influencers, one main function needs to be adapted. Directions are provided under the "Info" tab of the implementation.
Lastly, since DEL techniques were used to define the social network model, epistemic features provide an accessible extension of the model: for example, influencer marketing can be modelled by allowing influencers to lie about their opinions to certain agents. Furthermore, as previously suggested by Smets and Velázquez-Quesada (2020), explicit knowledge of certain features of other agents can be required for a relation to be formed or to be influenced by those agents.

References

U. Wilensky. Netlogo. http://ccl.northwestern.edu/netlogo/, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999.

A. Baltag, Z. Christoff, R.K. Rendsvig, and S. Smets. Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, 107(3):489–531, 2018. doi: 10.1007/s11225-018-9804-x.

S. Smets and F.R. Velázquez-Quesada. A logical analysis of the interplay between social influence and friendship selection. *Lecture Notes in Computer Science Dynamic Logic. New Trends and Applications*, page 71–87, 2020. doi: 10.1007/978-3-030-38808-9_5.

A. Baltag, Z.L. Christoff, J.U. Hansen, and S. Smets. Logical models of informational cascades. *Logic Across the University: Foundations and Applications: proceedings of the Tsinghua Logic Conference, Beijing, 2013*, page 405–432, 2013. URL `https://hdl.handle.net/11245/1.406292`.

S. Tsugawa and K. Kimura. Identifying influencers from sampled social networks. *Physica A: Statistical Mechanics and its Applications*, 507:294–303, 2018. doi: 10.1016/j.physa.2018.05.105.

D. Easley and J. Kleinberg. *Networks, crowds and markets: reasoning about a highly connected world.* Cambridge University Press, 2019.

M.S. Alvim, S. Knight, and F. Valencia. Toward a formal model for group polarization in social networks. *The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy Lecture Notes in Computer Science*, page 419–441, 2019. doi: 10.1007/978-3-030-31175-9_24.

M. Young Pedersen, S. Smets, and T. Ågotnes. Further steps towards a logic of polarization in social networks. *Lecture Notes in Computer Science Logic and Argumentation*, page 324–345, 2020. doi: 10.1007/978-3-030-44638-3_20.

M. Young Pedersen, S. Smets, and T. Ågotnes. Analyzing echo chambers: A logic of strong and weak ties. *Logic, Rationality, and Interaction Lecture Notes in Computer Science*, page 183–198, 2019. doi: 10.1007/978-3-662-60292-8_14.

Z. Christoff and J.U. Hansen. A logic for diffusion in social networks. *Journal of Applied Logic*, 13(1):48–77, 2015. doi: 10.1016/j.jal.2014.11.011.

S. Smets and F.R. Velázquez-Quesada. How to make friends: A logical approach to social group creation. *Logic, Rationality, and Interaction Lecture Notes in Computer Science*, page 377–390, 2017. doi: 10.1007/978-3-662-55665-8_26.

H. van Ditmarsch, W. van der Hoek, and B.P. Kooi. *Dynamic epistemic logic*. Springer, 2008.

D. Chen, L. Lü, M. Shang, Y. Zhang, and T. Zhou. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, 391(4): 1777–1787, Feb 2012. URL `https://doi.org/10.1016/j.physa.2011.09.017`.

Werner Geyser. The state of influencer marketing 2021: Benchmark report, Feb 2021. URL `https://influencermarketinghub.com/influencer-marketing-benchmark-report-2021/`.

Anna Holligan. Dutch celebs face backlash over covid rebellion, Sep 2020. URL `https://www.bbc.com/news/world-europe-54279008`.

K.J.s. Zollman. Network epistemology: Communication in epistemic communities. *Philosophy Compass*, 8(1):15–27, 2013. doi: 10.1111/j.1747-9991.2012.00534.x.

F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191(11):2403–2431, 2014. doi: 10.1007/s11229-014-0432-3.

U. Wilensky and W. Rand. Netlogo random network model, 2008. URL `http://ccl.northwestern.edu/netlogo/models/RandomNetwork`.

Appendix A

Python Implementation

The Python implementation stores its data in two dataframes: the first keeps track of the opinions of every agent, while the second stores the relations between agents. Figure 16 shows an example of the initial dataframes given a SNM $M \in \mathcal{M}$ with agents $\mathcal{A} = \{a, b, c, d\}$. The dataframe of relations (left) is an adjacency matrix where for $i, j \in \mathcal{A}$, the entries at indexes (i, j) and (j, i) both display the relation status between agents i and j: 0 implies that i and j are not related; 1 implies friendship, and -1 implies acquaintanceship.
The dataframe of positions is composed of rows which each display the opinions of an agent, and columns each denoting a topic position. For $i \in \mathcal{A}$ and $s \in \mathcal{R}$ (with $\mathcal{R} = \{p, q, r\}$ being the set of available positions), the entry at index (i, s) shows if agent i supports position s. If so, the entry is 1; otherwise it is 0. The

	p	**q**	**r**
a	0	1	1
b	1	1	1
c	1	1	1
d	0	1	0

	a	**b**	**c**	**d**
a	1	-1	-1	1
b	-1	1	0	0
c	-1	0	1	0
d	1	0	0	1

Figure 16: SNM $M = \langle N_P, N_N, V \rangle$ with agents $\mathcal{A} = \{a, b, c, d\}$, topics $\mathcal{T} = \{P, Q, R\}$ and positions $\mathcal{R}_P = \{p\}$, $\mathcal{R}_Q = \{q\}$, and $\mathcal{R}_R = \{r\}$, such that $\mathcal{R} = \{p, q, r\}$.

friendship selection update computes the updated relations in the model given the current positions of the agents and occurs simultaneously for all agents in the model. To ensure this, the updated relations are computed and stored in a temporary dataframe before performing the actual update. This dataframe is initiated as an identity matrix: as the friendship relation is reflexive, each agent has a friendship relation with himself at any point during a simulation, and thus this relation does not need to be computed. To compute the updated relations, the algorithm iterates over all possible combinations of agents $i, j \in \mathcal{A}$ where $i \neq j$ and, given the dataframe of positions, computes for each combination if the requirements (see definition 3.6) for either a friendship or an acquaintanceship are met. The cells in the temporary dataframe corresponding to agents i and j are filled accordingly. Thereafter, the original dataframe of relations is replaced with the updated dataframe.

The social influence update computes the updated positions of the agents in the model given current relations and occurs simultaneously as well. Similarly to the friendship selection update, updated positions are stored in a temporary dataframe during computation. Following definition 3.3, agents related to none other than themselves keep their current opinions. Otherwise, the algorithm computes for each position $s \in \mathcal{R}$ if the requirements (see definition 3.3) are met to adopt that position. Thereafter, the original dataframe of positions is replaced with the updated dataframe.
The implementation of the polarization measure was written solely for models with θ_1 and θ_2 configurated as in 1, that is, $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{2}{3}$. The algorithm follows the procedure in NetLogo that was designed for this configuration.
Similarly, influencers are computed and updated according to the same algorithm as in the NetLogo implementation. Different from NetLogo, Python supports nested dictionaries and thus, the influencer data is stored in one nested dictionary instead of in various lists.