university of groningen / faculty of science and engineering

Signed Threshold Graphs and their Logic

A study into the logic of behavior diffusion through a social network of friends and enemies

Master's Thesis

Bonne Jongsma

s3501647

March 31, 2025

First Supervisor: Ass. prof. Zoé Christoff
Second Supervisor: Prof. dr. Rineke Verbrugge

Artificial Intelligence
University of Groningen

Abstract

This thesis aims to bridge the gap in social network analysis between balance theory and diffusion models. Balance theory is based on signed graphs, where agents can have positive (friends) and negative relations (enemies) to other agents. Diffusion models study how a behavior, opinion or technology spreads through a network using a threshold: agents adopt if a threshold of adopted neighbors is reached. This results in a model that is updated in discrete timesteps.

The combination of these two models results in signed threshold graphs, which depict how a behavior diffuses through a network of friends and enemies. The key idea is that enemies, unlike friends who copy each other, want to have opposing behavior. It is proven that the optimal state of behavior diffusion can be reached iff the graph is balanced. Afterwards, a dynamic extension of social network logic is developed for reasoning about these models. An axiomatisation is given that is proven to be sound and complete. The properties of non-overlapping, balance and completely connected are defined by including the difference operator.

Keywords— Balance theory, Modal logic, Threshold models, Social network analysis, Dynamic logic, Completeness

Contents

1 Introduction

Imagine that your partner has a hobby that they are very passionate about. Hearing your partner enthusiastically talk about it, you feel enticed to try it, too.

This is an example of *social influence*: the intentional or unintentional communication that produces change(s) in another person's attitudes, beliefs, intentions, motivations, or behaviors. In this case, it is the behavior of participating in this particular hobby. While any person is more likely to be influenced by someone that they *like* [16], they are influenced by everyone around them, whether it happens consciously or not. This includes the influence of people that they do not like.

Continuing this scenario, imagine that your partner has a little friend group consisting of 3 friends and themselves. Your partner is not the only one that engages in this hobby, their entire friend group does so, too. But after finally meeting these friends, you have found a mutual dislike of each other. The fact that you dislike your partner's friends has both direct and indirect consequences. It creates tension in this social situation: your partner likes all of you and would like you to get along. It also causes you to feel demotivated to try the hobby. While your partner might have convinced you to try the hobby before, knowing that their friends participate too makes you feel less inclined to do so. This situation is a case study in how someone's network of friends and rivals causes them to be influenced about a behavior.

This thesis aims to analyse these situations in a more general sense. This is done by combining the idea of *balance* with *behavior diffusion* in a network. Both ideas have associated models, which will be combined into one model. Both of these models have associated logics, which will be combined into one logic that is semantically defined on the combined model.

The rest of this section will explain the basic ideas behind balance and behavior diffusion, aand the research questions will be posed afterwards. Section 2 gives all necessary information on balance theory and its logic. Section 3 will do the same for threshold models. The research questions will be explored in Section 4 and Section 5. Section 6 concludes the thesis with a discussion and some final remarks.

1.1 Balance

Balance in a network is based on the idea that everyone wants their environment to be harmonious and stable. This social harmony is commonly known by the phrases 'The friend of my friend is my friend' and 'The enemy of my enemy is my friend'. This idea of social harmony in a group of three is known as *local balance*. On the other hand, 'The friend of my friend is my enemy' and 'The enemy of my enemy is my enemy' are not locally balanced.

Figure 1.1 shows these four situations using a *signed graph*. The *nodes* represent three people in the situation. The *edges* between them depict their relationships. These edges are labeled with a either a '+' or a '-' and are henceforth called *signed edges*. They represent respectively a friendship or a rivalry between the people. Formally, local balance is defined on a signed graph as follows.

Definition 1.1 (Local balance, [7, 13])**.** Name a set of three nodes that are connected by three signed edges a *triad*. A triad is *locally balanced* if either all three edges are labeled $+$ or exactly one edge is labeled $+$.

The example described above with you, your partner and their friend, can be depicted using a signed graph, shown in Figure 1.2. Your partner's friends and you dislike one another, but otherwise everyone in the graph has a positive relationship. It is clear some of the triads are not locally balanced. For example, the triad of you, your partner and friend 1 contains exactly 2 edges labeled '$+$'.

Local balance in the form of these four triads is the simplest example of balance in a graph. But balance can be extended to huge, potentially infinite graphs. Balance theory translates the uncomfortable feeling of tension in a social network to a mathematical rule to follow. It allows someone to look only at the structure of a network to determine potential problems in teamwork. As the balance property is only based on the structure of the network [13], it is sometimes referred to as *structural balance*. Additionally, balance in a network can be connected to the global property of *friend groups*: clusters of people in a network who are friends with each other but enemies with everyone else. Individual relations thus translate to bigger structural phenomena. Section 2.1 consists of all necessary preliminaries of signed graphs and balance theory.

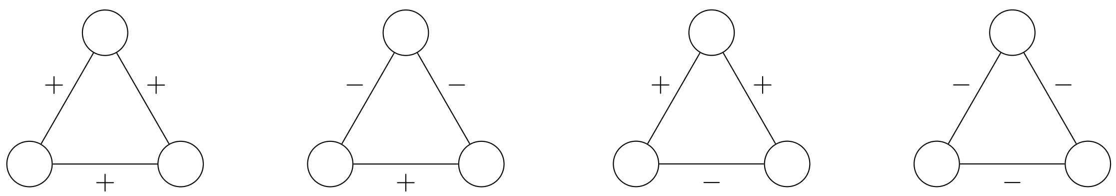

Figure 1.1: The four triads. The two triads on the left are balanced.The two triads on the right are not balanced.

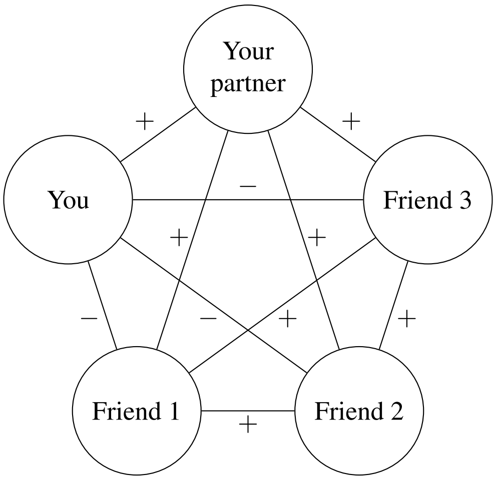

Figure 1.2: A tense situation between you, your partner and their friends.

Balance theory has natural applications in psychology [23, 1], sociology and mathematics [19, 11, 7]. It also has applications in marketing, politics and artificial intelligence.

Marketing campaigns try to link a product to a likeable, good-looking celebrity to create a positive association with the product. If someone likes a celebrity but does not know the product, balance theory explains how a person is more inclined to feel positive towards the product. On the other hand, if someone likes a celebrity but already dislikes the product, this situation is not balanced. The person is motivated to either start disliking the celebrity endorser or start liking the product. This choice is usually made by which relationship is stronger.

Another application is in politics. A country aligning itself with two warring countries would cause a lot of conflict. An example of this is described by Moore in [27]: "For example, the United States' somewhat surprising support of Pakistan in the Bangladesh conflict of 1972, becomes less surprising when one considers that the USSR was China's enemy, China was India's foe and India had traditionally bad relations with Pakistan. Since the US was at that time improving its relations with China, it supported the enemies of China's enemies."

An overlap between these instances of political alignment and business marketing is found in celebrity and brand boycotts. In the recent war between Israel and Palestina, many pro-Palestinian protesters have called to boycott McDonald's and Starbucks, citing their perceived alligiance to Israel [36]. Celebrities that have spoken out in support of Israel have gotten backlash from Palestian supporters while some celebrities that have spoken out in support of Palestina, have had their contracts terminated [6]. Celebrity endorsement of politicians is also believed to help their campaign: Oprah's support of Barack Obama in the 2008 elections helped him get over a million voters [15]. These matters come down to allegiance: a triad between a celebrity, the cause they support and the public or associated companies is formed. The desire for balance causes these relationships to change when necessary.

Balance theory as a field was founded by Harary [19], and was later expanded upon by Cartwright [7] and Davis [11]. A short historical overview of the relevant research done on balance theory is given in Section 2.2.

Several logics for signed balanced models have been proposed. These include a logic of allies and enemies (**LAE**) [37] and positive and negative relations logic ($\mathbf{pnl}_n$) [39, 29, 28], which take a very different approach to signed graphs. **LAE** includes a third type of edge for agents who do not know each other, raising the number of possible triad configurations from 4 to 10. This logic also makes the distinction between balance and stability in a network, because it looks at how the structure of the network changes through time. It is based on Computation Tree Logic (CTL). On the other hand, $\mathbf{pnl}_n$ is based on modal logic. To capture balance, this thesis will focus on the logic $\mathbf{pnl}_n$. It can express a lot with very little language, and can be straight-forwardly combined with a dynamic logic to obtain a dynamic social network logic. In Section 2.3 the syntax and semantics of $\mathbf{pnl}_n$ are discussed, in addition to some possible extensions. Section 2.4 gives a short overview of **LAE**.

1.2 Diffusion model

A diffusion model captures have a behavior, opinion or technology spreads through the agents in a network through social influence. This matter to be influenced on will be referred to as simply 'a behavior' from here on. Agents that have some relationship between them are called *neighbors*. The key idea is that one agent adopting a behavior can inspire a neighboring agent to also adopt it. Specifically, they need a critical number of neighbors to have adopted.

This makes a *threshold model*: it is a graph with agents of two different types (agents who have already adopted and agents who have not) and a critical threshold.

Let us look at the example of you, your partner and their friends, and imagine that all of you get along. However, still only your partner and their friends have adopted the behavior , you have not. This can be depicted using a threshold model, shown in Figure 1.3. The red nodes show people who have adopted, while the black nodes show people who have not (yet) adopted. However, the figure shows an incomplete threshold model: there is no threshold.

Let us put the threshold at $\frac{9}{10}$. Then, in order for you to be influenced into adopting, you would need to have at least $\frac{9}{10}$ of your friends to have adopted already. This is not the case, and you will not adopt. On the other hand, if we let the threshold be $\frac{3}{5}$, the critical threshold will be reached and you will adopt. Both of these scenario's suggest that there is a next *timestep* where you adopt or not. This is defined using an *updated threshold model*, where the graph and threshold remain the same but the set of adopted agents is updated. There can be multiple timesteps, and with each timestep the model is updated again. This sequence of updated models shows a domino effect: a few agents adopt, and step-by-step this causes the other agents in the network to adopt. Section 3.1 consists of all the necessary preliminaries to threshold models.

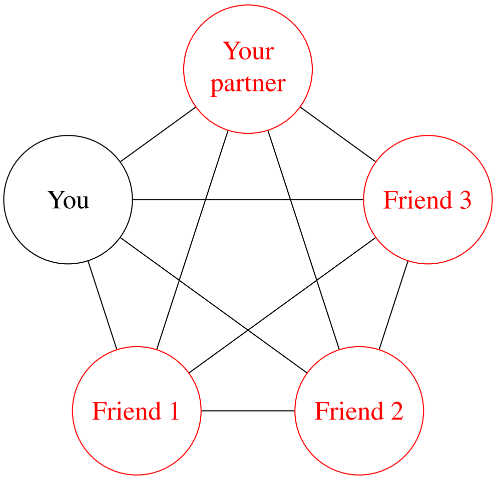

Figure 1.3: A network of you, your partner and their friends. Only you have not adopted the behavior of playing the computer game.

There are two threshold models that have gained wide-spread notoriety: Schelling's segregation model [33, 34] and Granovetter's riot model [18]. Though they differ a lot, they both depend on threshold-limited influence. Section 3.2 goes into the details on both models and their differences.

A threshold is inherent to a threshold model. How does an agent decide on a threshold? Essentially, both options of adopting and not adopting have an associated *payoff*. If the choice to adopt has a much higher payoff than not adopting, an agent would have a lower threshold. This decision between two choices and their respective payoffs comes from *game theory*, defined in *Networks, Crowds & Markets* as: "Game theory is designed to address situations in which the outcome of a person's decision depends not just on how they choose among several options, but also on the choices made by the people they are interacting with" [13]The threshold is thus a strategic element of the threshold model. Section 3.3 takes a game-theoretical approach to the threshold and the variables it depends on.

Multiple logics have been proposed to describe threshold-limited influence [4, 38, 31, 41, 3, 10, 25, 35], often including an *epistemic* angle. Epistemic logic includes operators about knowledge and belief. This allows for *higher-order reasoning*: if an agent knows what their neighbors know, they can predict their neighbors actions and act on those predictions. One of those logics is the Logic of Threshold-Limited Influence (L_θ) [4]. It distinguishes itself by allowing for any uniform threshold, and for being minimal: it is a propositional logic with one dynamic operator. This is in contrast to other logics on social influence, which use hybrid logic [10, 35] or assume a set threshold [18, 25]. This allows for a minimal logic while maintaining expressivity. Section 3.4 details the syntax and semantics of L_θ.

1.3 Research question

In a threshold model, the links between agents can be seen as 'friendships'. This corresponds to the idea that agents want to coordinate with the behavior of their neighbors. What would happen if some of the links are not friendships but rivalries? It is natural to assume that agents do not want to adopt the behavior that their enemies are displaying. Instead of playing a so-called 'coordination game' with their neighbors, agents will play an 'anti-coordination' with their enemies. Thus a link between a signed graph and a threshold model graph is made. These two models have only been considered independently so far and a formal connection has not been made yet.

To this end, the research question is posed:

Is it possible to combine a signed graph and a threshold model, to obtain a model of friends and enemies that influence each others decision to adopt a behavior?

This question is explored in Section 4. The model is defined in Section 4.1. Afterwards, the following question is explored in Section 4.2:

Is there a connection between the structure of such a model and the ability of agents to achieve a high payoff?

Lastly, Section 4.3 introduces a simplification of the games an agent engages in when deciding on a threshold. This simplification allows for a concrete update rule.

Furthermore, both a signed graph and a threshold model have been captured by their respective logics. This raises the follow-up question:

Can the logics $\mathbf{pnl}_n$ and L_θ be combined to form a meaningful logic that describes this combined model?

In this thesis, a logic is considered meaningful when it it *sound* and *complete*. This question is explored in Section 5. First, the syntax, semantics and the model are defined in Section 5.1. Section 5.2 is a short discussion of possible update rules. Section 5.3 the axiomatisation that defines the logic. Section 5.4 and Section 5.5 prove respectively the soundness and strong completeness of the logic with respect to the class of balanced signed threshold graphs.

Finally, Section 6 concludes the thesis. Section 6.1 summarises the research done in Section 4 and Section 5. Section 6.2 gives a discussion on the research and its limitations. Lastly, Section 6.3 explores several avenues for further research in this field are explored.

1.4 Acknowledgements

I want to thank my supervisor Zoé Christoff, for her academic guidance, open ear for discussions and her support of me in many different aspects. Her endless patience with me and our monstrous e-mail chain have gotten this thesis to completion.

I want to thank Mina Young Pedersen for her writing style [29, 28], as its clarity and readability inspired me a lot while writing. She also kindly helped me with some of the graphs shown in this thesis. I want to thank Anton Chernev for kindly helping me with completeness proofs and canonical models on multiple occasions. Most importantly, the proof of Lemma 5.23 would not have happened without him.

I want to thank my father for his support throughout my entire studies, and for encouraging me throughout my entire life to challenge myself, both academically and personally. Thank you for always being there for me. I want to thank my study buddy Bas, I would never have actually started my thesis if not for your recurring presence at the Bernoulliborg. I want to thank Elke, Liisa and my fellow students that listened to my ups and downs every Monday at the thesis support group. Lastly, I want to thank my partner Dirk. You were there when it was (very) hard, and you were there to celebrate every small and big achievement with me. You make it enjoyable to show up every day.

2 Signed balanced graphs and their logic

This section will detail all relevant information about signed graphs, balance theory and its associated logic. In Section 2.1, an overview of all the preliminary information regarding signed graphs and balance theory is given. This is followed by a chronological overview of the relevant literature in Section 2.2. In subsection Section 2.3, an introduction to $\mathbf{pnl}_n$ is given. Section 2.4 discusses further research into logic of signed graphs.

This section assumes a basic knowledge of graph theory and modal logic. If unfamiliar with graph theory, the reader is referred to Chapter 2 of [13] for the basic definitions. The first few chapters of [30] clearly explain all necessary information on modal logic.

2.1 Preliminaries of Balance Theory

Though Fritz Heider [23] laid down the psychological foundation on how tension in social networks is formed through network structure, the mathematical formalisation was done by Harary. In order to discuss the mathematics of balance, it is necessary to first define a *signed graph*.

Definition 2.1 (Signed graph, [19, 39]). A signed graph is a tuple $G = (A, R^+, R^-)$, where A is a set of nodes, $R^+ \subseteq A \times A$ is a set of positive edges and $R^- \subseteq A \times A$ is a set of negative edges. A signed graph has the following properties:

1. $R^+ \cap R^- = \emptyset$ (non-overlapping)
2. $(x,y) \in R^* \iff (y,x) \in R^*$ for all $x,y \in A$ and $i \in \{+,-\}$ (symmetry)
3. $(x,x) \in R^+$ for all $x \in A$ (positive reflexivity)

The assumption of mutual friendship or rivalry might not be intuitive, but assuming anything other than symmetry turns out to be not that interesting regarding balance theory. The possibility of balance in signed *directed graphs* (digraphs) was discussed in [7]. A directed graph is a graph where the existence of edges (x,y) and (y,x) are independent of each other. Symmetry in a directed graph is therefore not necessary.

In an undirected graph, asymmetrical relationships can exist: agent x likes (or dislikes) agent y but agent y doesn't reciprocate these feelings. As it turns out, even one asymmetrical relationship in an undirected graph will result in the graph not being balanced [7]. Thus balance can only properly be studied in graphs that assume symmetry. Additionally, most of the relevant literature also considers symmetry to be a basic assumption for signed graphs. This thesis will therefore assume a signed graph to be symmetric, and thus undirected.

Notation. The shorthand infix notation xR^+y and xR^-y will be used, where xR^+y means that $(x,y) \in R^+$ and xR^-y means that $(x,y) \in R^-$. Thus, xR^+y means: 'Agent x is friends with agent y'. Furthermore, the notation xR^*y will be used to say $(x,y) \in R^+ \cup R^-$; there is an edge between x and y. Two nodes x and y are *adjacent* if xR^*y. These agents are then commonly referred to as *neighbors*.

Originally, a balanced graph was defined using cycles, which is a type of path. A *path* $x_1 \dots x_k$ with $x_1, \dots, x_k \in A$ is a sequence of distinct nodes connected by edges, $x_i R^* x_{i+1}$ for all $i \leq k-1$ [13, 39]. We can think of a path as walking along the nodes and edges of a graph. If you make a path where you end at the node that you started with, you have a *cycle*.

Definition 2.2 (Cycle). A *cycle* in a graph is a path of nodes $x_0 \dots x_k = x_0$ where the end node x_k is equal to the start node x_0. The length k of a cycle is defined as the number of edges it covers. A cycle of length k is called a k-cycle.

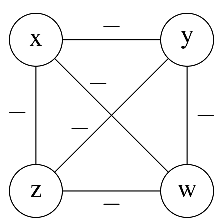

Figure 2.1: A signed graph.

Example 2.3. In Figure 2.1, a signed graph is given. The set of nodes consists of $A = \{x,y,z,w\}$. The set of positive relations is given by $R^+ = \{(x,y),(y,x),(x,z),(z,x),(z,w),(w,z),(y,w),(w,y)\}$ while $R^- = \{(x,w),(w,x)\}$. An example of a path in this graph would be $ywxz$, while an example of a cycle would be $xwzx$.

In a signed graph, a cycle covers both positive and negative edges. Based on how many positive and negative edges exist in a cycle, we can divide the class of cycles into two subclasses: positive and negative cycles. A cycle is *positive* if there is an even number of negative edges in the cycle, and *negative* if there is an odd number of negative edges in the cycle. Here is an intuitive way to look at this definition: give positive edges a positive weight and negative edges a negative weight. Then, as you walk along the path, multiply all the weights of a path until the first node is reached again. Thus, the *sign of a cycle* can be viewed as 'the sign of the product of all the edges in a cycle'. Note that a signed graph is different from a weighted graph, as in weighted graphs the weights of a path are added instead of multiplied.

With this, it is possible to define a *balanced graph*. Though local balance was defined in Section 1, this was done using triangle relationships. Triangles are 3-cycles, therefore they are a subclass of all cycles. Defining balance using the local property of triangles is beneficial when you want to introduce the topic, examine the graph on a very local level or when the graph is *completely connected*: all nodes in the graph are neighbors. For example, the graph from Figure 2.1 is not completely connected, because z and y are not neighbors. For real-life applications, graphs are completely connected only when you look at a small community: an office, a friend-group or a classroom. If a signed graph is completely connected, you can check whether the whole graph is balanced by looking at 3-cycles only. In a graph that is not completely connected, like bigger networks often are, this cannot be done. To this end, a definition of balance that does not rely on triangle relationships, is given.

Definition 2.4 (Balanced graph [19])**.** A signed graph $G = (A, R^+, R^-)$ is balanced iff it has no negative cycles.

A signed graph is *unbalanced* if it is not balanced. The graph from Figure 2.1 is unbalanced, as it contains the negative cycles *xwzx* and *xywx*. Additionally, note that any signed graph without cycles, called a *tree*, is by definition balanced.

There are multiple characterizations to balance in a signed graph. One characterization for a completely connected signed graph was already mentioned: a completely connected signed graph is balanced iff each of its triangles is balanced [13]. In Theorem 2.5, these equivalencies for signed graphs are collected.

Theorem 2.5 (Balance theorem [19, 39])**.** *Let $G = (A, R^+, R^-)$ be a signed graph. The following are equivalent:*

1. *Graph G is balanced.*
2. *There are no negative cycles in G.*
3. *There exists $A_1 \subseteq A$ such that*
   - *for all $(x,y) \in R^+$, $x \in A_1$ iff $y \in A_1$, and*
   - *for all $(x,y) \in R^-$, $x \in A_1$ iff $y \in A \setminus A_1$.*
4. *There exists a completely connected, balanced signed graph $G' = (A, R^{+'}, R^{-'})$ such that $R^+ \subseteq R^{+'}$ and $R^- \subseteq R^{-'}$.*
5. *For all distinct agents $x, y \in A$, all paths between x and y have the same sign.*

The second statement is the definition of a balanced graph. The third statement is a formal generalization of the famous *structure theorem* by Harary. It states that a balanced graph can be partitioned into two disjoint subsets A_1 and A_2, where one of the sets is possibly empty, where all edges between nodes in the same subset are positive and all edges between nodes of different subsets are negative: the two groups like everyone from their own group, and hate everyone in the other group. The sets A_1 and A_2 that make up this partition are called *friend groups* from here on. An example of this is depicted in Figure 2.2. All the agents on the left form the set A_1 and all agents on the right form the set A_2.

Notation. Whenever a graph $G = (A, R^+, R^-)$ is balanced, the set A can be partitioned into two sets A_1 and A_2 as described in Theorem 2.5, part 3. This is called a *balance partition*, and the notation $A = A_1 \cup_\triangle A_2$ will be used to describe it.

This is an important result in balance theory. It is used and referenced often, as it gives insight on how local friendships and rivalries impact the structure of a network.

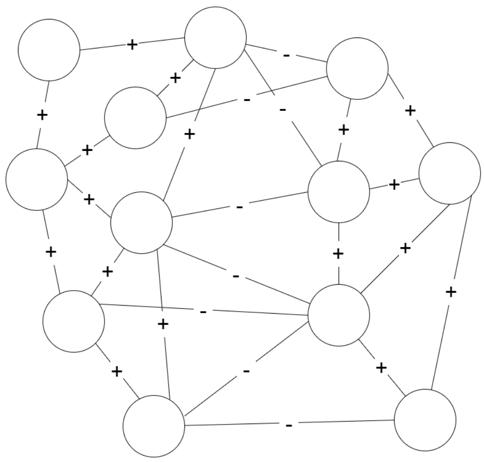

Figure 2.2: A balanced signed graph.

The fourth statement tells us two things. Firstly, any (non-completely connected) subgraph of a completely connected, balanced graph is also balanced. Secondly, we can fill in the 'missing edges' in a balanced graph to end up with a completely connected, balanced graph. For example, take a graph where some agent is friends with two other agents in the graph. Those two agents can then become friends to complete the triangle and maintain balance. If an agent has both an enemy and a friend, the friend is

encouraged to become an enemy of this enemy. Like this, the whole network can be filled in, keeping structural balance in place.
The fifth statement is another result from Harary on balanced graphs. Here is a sketch of the proof: Take two different paths between agent x and agent y. When a path is created from combining both paths, like $x \dots y \dots x$, a cycled has been created. If the two paths have different signs, then the cycle must be negative. Complete proofs for these equivalences can be found in [19].

2.2 An overview of Balance Theory in the literature

Over the years, the literature on signed graphs has become extensive. For a clear and concise documentation, see [40]. It is the most complete collection of literature on signed graphs, and this section will detail the research relevant to this thesis.

Signed graphs have applications in different fields, and have thus been studied from different perspectives. Frank Harary was not just a mathematician, he was one of the fathers of modern graph theory, therefore his work in balance theory is done from a mathematical perspective. His paper 'On the notion of balance of a signed graph' [19] from 1953 was the paper that founded the field of balance theory. It defined a signed graph, defined balance and proved the results as detailed in Theorem 2.5.

In 1956, mathematician Harary collaborated with social psychologist Dorwin Cartwright. Together, they wrote 'Structural balance: a generalization of Heider's theory' [7]. It connects the psychological foundational theory of psychologist Fritz Heider [23], where balance is viewed specifically as a concept between groups of two or three people or objects, to formal graph theory. They define balance on graphs of 3 agents only (Definition 1.1) and then generalizing this concept to bigger graphs (Definition 2.4). The paper first details Heider's approach, and then discusses the five ambiguities they find in it; symmetry (as discussed in Section 2.1), balance in a network bigger than 3 agents, how a negative relation is defined, different types of relations and the possibility to describe a social system.

Harary also wrote another paper balance theory: 'On local balance and n-balance in signed graphs' [20], where two additional weaker versions of balance in a graph are defined. The first is *local balance at a node p*, where all cycles going though p are positive. The second definition he gave is on *n-balance*.

Definition 2.6 (n-balance [20])**.** A signed graph G is n-balanced if each cycle in G of length n or shorter is balanced.

Notation. In order to make the distinction with weaker forms of balance like n-balance, the term 'balance' as defined in Definition 2.4, is sometimes referred to as *full balance*.

Though n-balance and balance at a node were proposed by Harary, a more popular and well-known weaker form of balance was defined by Davis in [11]. The key idea is that big groups of people often break up into several subgroups that dislike every other group. This is opposed to Harary's definition of structural balance, as it only allows for the structure of a network to be partitioned into two friend groups, instead of multiple. Davis proved that clustering into multiple friend groups occurs when there are no cycles in a graph with exactly one negative edge. Though Davis defined this new form of balance as *clustering*, it is commonly referred to as simply 'weak balance' [13].

Definition 2.7 (Weak Balance [11])**.** A signed graph G is weakly balanced if G contains no cycle having exactly one negative edge.

Weak balance is very comparable to balance. It is so comparable, that almost every definition and theorem regarding balance can be adjusted slightly to make it hold for weak balance. Local balance in terms of the triangles from Definition 1.1 and Figure 1.1, the situation where three people dislike each other would be classified as weakly balanced. It is stated that this situation is less likely to change than the triangle with one negative edge. Secondly, the first four statements of Theorem 2.5 can be adjusted slightly such that they hold for weak balance. Most notably, part 3, which states that any balanced graph can be divided into two disjoint subsets, can be rephrased as: any weakly balanced graph can be divided into a finite number of disjoint subsets.

A good introduction to balance theory is found in the textbook *Networks, Crowds and Markets: Reasoning about a Highly Connected World* by Easley and Kleinberg [13, Ch. 5]. In order to prove balance in a non-complete graph using triangles, an algorithm is described to find whether a graph is balanced or not. They first find *supernodes* of connected positive components, and after checking that there are no negative edges within these supernodes, apply breadth-first search is to the supernodes. If any two supernodes in the same layer are connected, the graph is unbalanced. The structure theorem and an approximation of balance are also mentioned and proven. Informally, this approximation states that if a high percentage of the cycles are balanced, then the network can be divided into two groups where most agents like each other and dislike the agents from the other group.

Looking at the class of all unbalanced signed graphs, some are closer to being balanced than others. In 1959, Harary wrote [21], where three different indices to quantify 'the amount of balance' are described.

The first index is the *degree of balance*, which is the ratio of number of positive cycles to the total number of cycles. While relatively easy to compute, it doesn't take into account properties like the length of a cycle and overlapping cycles. The importance of these properties will be further explained in Section 2.3.

Secondly is the *line index of balance*, which encompasses both the *line-negation index* and the*line-deletion index*. The line-negation index is the minimum number of edges that need to be sign-changed in order for the graph to be balanced. The line-deletion index is the minimum number of edges that needs to be deleted to the graph balanced.

Example 2.8. The signed graph in Figure 2.1 is unbalanced, but if the edge *xw* is changed to a positive edge, the graph would be balanced. Thus, the line-negation index is one. However, the edge could also be entirely deleted to make the graph balanced. The line-deletion index is thus also one. In this case, the line-negation index and the line-deletion index are equal.

In fact, the paper proved that the two indices are always equal. Thus, these two indices are grouped together to form the line index of balance: the minimum number of edges that needs to be sign-changed or deleted in an unbalanced graph in order to make the graph balanced.

Lastly, the *point index of balance* is defined similarly: what is the minimum amount of nodes that would need to be deleted from the graph in order to make the graph balanced? Removing an agent from a network is not always a practical solution in real life.

Of these three proposed indices, the line index of balance has been most commonly used afterwards. Another term used for the line index is *frustration*. While Harary was writing this paper, he corresponded with social psychologists Robert Abelson and Milton Rosenberg, who wrote the paper 'Symbolic Psycho-Logic: A model of attitudinal Cognition'[1] around the same time. It is split into two parts: the psychological model and the mathematical model. These parts work in harmony together, where the latter half details the psychological ideas using mathematical models and matrices. It constructs a graph and its relations using symmetrical matrices and devices operations on these matrices that can determine balance in a network. This paper also measures the minimum number of changes necessary to achieve balance, and calls it *complexity*, a term that is now obsolete. This paper approaches social networks and balance from an angle of linear algebra instead of graph theory, which is not often used in the following literature. But it mirrors a lot of the work done in the field, and it can serve as a useful bridge between the psychological foundations of the field and the mathematical analysis done on those foundations.

2.3 A logic for signed graphs

2.3.1 The invention of $\mathbf{pnl}_n$

A logical formalisation of balance can be found in the paper 'On the Logic of Balance in Social Networks' by Xiong and Ågotnes [39]. I will summarise the relevant parts of $\mathbf{pnl}_n$ in this section.

The paper defines a modal logic for positive and negative relations, where the models consist of signed graphs. The axiomatisation of the logic is proven to be sound and weakly complete with respect to the class of n-balanced graphs.

The weaker form n-balance (Definition 2.6) is chosen because the authors argue that full balance is too strong of a requirement for practical purposes. In real life, networks are almost never fully balanced. After all, one unbalanced cycle can ruin the balance of an entire network, possibly millions of nodes big. They also raise the point that longer cycles are more likely to be unbalanced than short cycles, increasing the likelihood a network is unbalanced the bigger it gets. Instead, the authors analyse signed graph with a logic that defines balance up to a degree. The language $\mathcal{L}_{\text{PNL}}$ of the resulting logic is defined as follows.

Definition 2.9 (Syntax of $\mathbf{pnl}_n$, [39])**.** Let $\textsc{At}$ be a countable set of propositional letters. We define the well-formed formulas of the language $\mathcal{L}_{\text{PNL}}$ to be generated by the following grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \diamond\!\!\!+\varphi \mid \diamond\!\!\!-\varphi$$

where $p \in \textsc{At}$. Additionally, the following shorthand notation for the propositional connectives and duals will be used:

$$\begin{aligned}
\varphi\vee\psi &\iff \neg(\neg\varphi\wedge\neg\psi)\\
\varphi\to\psi &\iff \neg\varphi\vee\psi\\
\top &\iff \varphi\vee\neg\varphi\\
\bot &\iff \varphi\wedge\neg\varphi\\
\boxplus\varphi &\iff \neg\diamond\!\!\!+\neg\varphi\\
\boxminus\varphi &\iff \neg\diamond\!\!\!-\neg\varphi
\end{aligned}$$

The propositions in $\textsc{At}$ are understood to be properties of agents, like 'has blonde hair' or 'studies at the University of Groningen'. The symbols $\neg$, $\wedge$, $\vee$, $\to$, $\leftrightarrow$, $\top$ and $\bot$ are defined as usual.

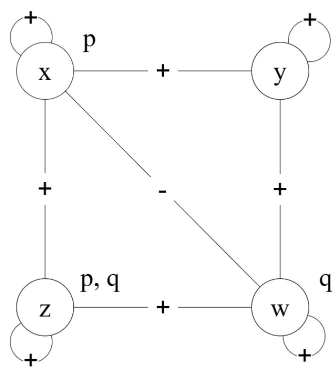

Figure 2.3: An example of a model.

Furthermore, $\Diamond^{+}\varphi$ means an agent has a friend with property φ. Similarly, $\Diamond^{-}\varphi$ means an agent has an enemy with property φ. The dual modalities $\boxplus\varphi$ and $\boxminus\varphi$ mean that all of the agents friends or enemies respectively have property φ.

In order to determine whether a statement is true or not, a language of a logic is interpreted on a *model*. Formally, this is defined as follows:

Definition 2.10 ((n-balanced) model, [39]). A model $M = (F,V)$ is a pair of a frame $F = (A,R^+,R^-)$ and a valuation function $V : \mathrm{AT} \to \mathcal{P}(A)$, where $\mathcal{P}(A)$ refers to the power set of A. A frame is a signed graph, where R^+ is reflexive. A valuation function assigns the propositions in AT to sets of agents. A model is n-balanced if the graph produced by the frame is n-balanced.

Example 2.11. An example is given in Figure 2.3. The frame $F = (A,R^+,R^-)$ is given by $A = \{x,y,z,w\}$, $R^+ = \{(x,y),(y,x),(x,z),(z,x),(z,w),(w,z),(y,w),(w,y),(x,x),(y,y),(z,z),(w,w)\}$ and $R^- = \{(x,w),(w,x)\}$. The valuation function is given by $V(p) = \{x,z\}$ and $V(q) = \{z,w\}$.

A model is thus defined to consist of a set of agents, a set of positive relations, a set of negative relations and a valuation function. The notations $M = (F,V)$ and $M = (A,R^+,R^-,V)$ will be used interchangeably.
Now that the model is defined, we can define the semantics of **pnl**: this determines inductively whether a formula is true or not, by connecting the formulas to the model.

Definition 2.12 (Semantics of **pnl** [39]). For a model $M = (A,R^+,R^-,V) = (F,V)$, and formulas $\varphi,\psi \in \mathcal{L}_{\mathrm{PNL}}$, the formula φ is satisfied by an agent $x \in A$, written $M,x \vDash \varphi$, as follows:

$$
\begin{array}{ll}
M,x \vDash p & \Longleftrightarrow x \in V(p) \\
M,x \vDash \neg\varphi & \Longleftrightarrow M,x \nvDash \varphi \\
M,x \vDash \varphi \wedge \psi & \Longleftrightarrow M,x \vDash \varphi \text{ and } M,x \vDash \psi \\
M,x \vDash \Diamond^{+}\varphi & \Longleftrightarrow \exists b \in A \text{ such that } xR^+y \text{ and } M,y \vDash \varphi \\
M,x \vDash \Diamond^{-}\varphi & \Longleftrightarrow \exists y \in A \text{ such that } xR^-y \text{ and } M,y \vDash \varphi \\
M,x \vDash \boxplus\varphi & \Longleftrightarrow \forall y \in A : xR^+y \text{ implies } M,y \vDash \varphi \\
M,x \vDash \boxminus\varphi & \Longleftrightarrow \forall y \in A : xR^-y \text{ implies } M,y \vDash \varphi
\end{array}
$$

A formula is *satisfiable on a frame F* when there exists some model that makes it true. Namely, there exists some valuation function $V : \mathrm{AT} \to \mathcal{P}(A)$ and agent $x \in A$ such that $(F,V),x \vDash \varphi$.

Given this language, certain frame properties can be expressed using formulas. For example, a frame has positive reflexivity iff the formula $p \to \Diamond^{+}p$ is satisfied for every agent and every valuation on the frame. The property of 'positive reflexivity' is then *modally definable* [8].
However, the properties of non-overlapping and n-balanced models are proven to be *modally undefinable*. In order to still make an axiomatisation of $\mathbf{pnl}_n$, the authors define the new inference rule $\mathbf{Nb}_n$, see [39] for details. An axiomatisation $\mathbf{pnl}_n$ is provided using this rule. The article then proves that for all $n \in \mathbb{N}^+$, $\mathbf{pnl}_n$ is sound and weakly complete with respect to the class of all n-balanced models.

2.3.2 Extensions of $\mathbf{pnl}_n$

Pedersen, Smets, and Ågotnes extended the logic of $\mathbf{pnl}_n$ in their paper "Modal logics and group polarization" [29]. The underlying focus of this paper is to use the concept of balanced graphs for polarization in society. Positive and negative edges would then represent agreement and disagreement between agents.

The paper proposes multiple extensions, but the focus is on extending $\mathbf{pnl}_n$ with *hybrid logic*. The purpose of these extensions is to find a solution for the main two downsides of **pnl**: the property of full balance is not modally definable,and there is no axiomatisation of the class of fully balanced models. The n-balanced model and $\mathbf{pnl}_n$ are first extended to include weak balance. There is also some research done into the frame properties of completely connected frames and non-overlapping relations. These relate to varying measures of balance.

One of the proposed extensions is to incorporate the difference operator $\langle D \rangle$. Using this operator, it is proven that completely connectedness of a graph and non-overlapping are now modally definable.

This modality is both globally useful and captures both desired properties. Semantically, it is defined by:

Definition 2.13 (Semantics of the difference operator, [29])**.** For a model $M = (A, R^+, R^-, V)$ and formula φ, the formula $\langle D \rangle \varphi$ is satisfied when

$$M, x \vDash \langle D \rangle \varphi \text{ iff } \exists y \in A \text{ such that } x \neq y \text{ and } M, y \vDash \varphi$$

A hybrid logic extension of **pnl**, called $\mathbf{pnl}_i$, is introduced, which includes nominals and the hybrid operators $\downarrow$ and $@$. For further details on this hybrid logic extension, see [29]. Finally, multiple measures of balance, are discussed: how far is an unbalanced network from being polarized?

2.4 Further research into signed network logic

A different approach to find a logic of social balance theory can be found in [37]. The proposed logic, called Logic of Allies and Enemies (LAE), adapts a form of temporal logic. As it is written from a computer science perspective, there is a focus on computational complexity.

It is mentioned here for two reasons: to give a complete overview of the research done into these kind of logics, and because this logic already has a temporal aspect to it, foreshadowing the subject of Section 4 and Section 5. However, there are multiple reasons why this thesis choses to analyse signed graphs using $\mathbf{pnl}_n$ instead of LAE.

Firstly, the model on which LAE is defined, defines the absence of a relation between two agents as a neutral bond. It is represented in drawings by the absence of a line. Thus, in every network, all agents are connected in some way. This impacts the logic in several ways. It adds 6 more possible triads to the previous 4 triads from Figure 1.1. It also adds two possible categories apart from balanced and unbalanced triads: the class of partially balanced triads, where one out of three edges is neutral, and the class of pressure-free triads, where either two or three edges are neutral. The commonly-used definitions are also altered to accommodate neutral bonds. The definition of a completely connected network, a connected component, and even the definition of balance are changed. To solve this, a separate definition of *semi-balance* is introduced. Due to the inclusion of neutral bonds, mosts definition and theorems from this paper seem more complicated than the ones previously described in Sections 2.1 and 2.2.

Instability in a network causes the network to change, giving rise to a *time evolution* of *successors*. It is therefore a natural choice for LAE to be based on Computation Tree Logic (CTL), a logic that expresses changes through time. The time-evolution of CTL is dependent on stability, and stability depends on the amount of friends and enemies agents have. To this purpose, LAE includes an extra quantifier and some abbreviations next to the basis of CTL. The quantifier, the inclusion of neutral bonds in the network and the general language of CTL together complicate formulas. When comparing, $\mathbf{pnl}_n$ is more simplistic.

However, $\mathbf{pnl}_n$ does not characterize full structural balance, while LAE does. Furthermore, it makes a difference between balance and stability. Those ideas are fully dependent on each other in the context of balance, so a distinction is rarely made. However, the paper argues that a network can be stable without tension while being unbalanced. Finally, LAE already includes the idea of unbalanced networks changing through time, which this thesis also does in Section 4 and Section 5.

3 Threshold models and their logic

This section is an introduction to threshold models. In Section 3.1, all necessary elements of a threshold model are given and the most important theorem regarding threshold models is explained: the cluster theorem. Section 3.2 describes a few notable variations on threshold models from the relevant literature. Afterwards, Section 3.3 gives motivation on how to decide on a threshold, after which Section 3.4 details a logical formalisation of threshold models, called L_θ.

3.1 Preliminaries of Threshold Models

A threshold model describes how people are influenced by their neighbours to adopt a new behavior. The most general definition consists of a network, with nodes of two different types, a threshold and an update rule determined by the threshold. The structure of the network can differ from a small graph to a theoretically infinite, completely connected grid. The number of nodes of both types can change or stay the same. The threshold can be uniform throughout the network, differ per node, or be defined according to some network property. In short, threshold models can be found in varying forms. However, a core assumption in threshold models is that of *threshold-limited influence*: "agents adopt a behavior or opinion whenever a critical fraction of their neighbors in the network have adopted it already" [4].

the structure of a threshold model thus relies on agents and the relationship between them. This is modeled as a graph. However, edges are defined not as a set of neighboring agents, but as a function from agents to all their neighbors. This structure is called a *network*.

Definition 3.1 (Network [4, 13]). A network is a pair (A,N) of a non-empty finite set of agents A and a neighborhood function $N : A \to \mathcal{P}(A)$ such that N has the following properties:

1. $x \notin N(x)$ (irreflexive)
2. $y \in N(x) \iff x \in N(y)$(symmetric)
3. $N(x) \neq \emptyset$ (serial)

The properties of this function tell us an agent is not their own neighbor, as this would imply an agent also influences themselves. Further, the neighbor connection goes both ways, and every agent has at least one neighbor. Given the context that threshold models describe behavior of agents depending on their neighbors, it would not be useful to discuss an agent without any neighbors.

In order for agents to be influenced by their neighbors behavior, some behavior needs to be defined. This is given by a set of agents $B \subseteq A$ that have adopted a new behavior. The notation B is used both to mean the behavior itself and the set of agents that have adopted it. This set B defines the two different types of agents: the ones who are part of B and the ones who are not.

Now take an arbitrary agent x in such a network. Agent x assesses the behavior of all of their neighbors, and when a sizable enough portion of their neighbors have adopted, she will follow suit. But a 'sizable enough portion' is vague, and could mean anything.

For this purpose, a threshold $\theta \in [0,1]$ is added to the model. An agent adopts behavior B iff fraction θ of their neighbors have already adopted B. This variable is thus the defining factor for when a non-adopting agent will adopt. Usually, it is assumed that the threshold is *uniform*: every agent in the network has the same threshold.

Together, the network, set B and threshold θ form the *threshold model*.

Definition 3.2 (Threshold Model, [4]). A threshold model is a tuple $M = (A,N,B,\theta)$ where (A,N) is a network, $B \subseteq A$ is a behavior and $\theta \in [0,1]$ is a uniform adoption threshold.

Example 3.3. An example of a threshold model is given in Figure 3.1. Here network is comprised of a set of agents $A = \{1,2,\dots,9\}$. The neighbors of agent 1 are $N(1) = \{3,4,6\}$ and the neighborhood functions of the other agents are defined similarly. The set $B = \{3,7\}$, these are colored red in the figure. These agents, who have adopted the new behavior on their own, are called the *initial adopters*. The threshold θ is unknown.

Thus, a threshold model is established. However, a static threshold model does not make an interesting threshold model. A threshold model studies the change of the set B as more and more agents adopt the new behavior, thus a formal way to describe the transition of B at different timesteps is needed. During a time step, a new model M' is created from M using an *update rule*. The new model M' has the same network and the same uniform threshold as model M, but the set B has changed to include agents who have now adopted the new behavior.

Definition 3.4 (Threshold Model Update, [4]). The update of threshold model $M = (A,N,B,\theta)$ is the threshold model $M' = (A,N,B',\theta)$, where B' is defined by

$$B' = B \cup \left\{ x \in A : \frac{|N(x) \cap B|}{|N(x)|} \geq \theta \right\}$$

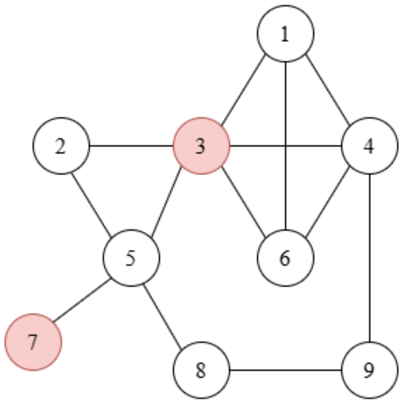

Figure 3.1: An example of a threshold model. The set B only has two agents.

Note that due to seriality of the network, $|N(x)|$ is never zero.

Example 3.5. Take the threshold model from Example 3.3 with a threshold of $\theta = \frac{1}{2}$. Only two agents satisfy the requirement that half of their neighbors have adopted: agents 2 and 5. The next timestep, these agents are included in the set B. This is also depicted in Figure 3.2, where red agents are in set B and white agents are not in set B.

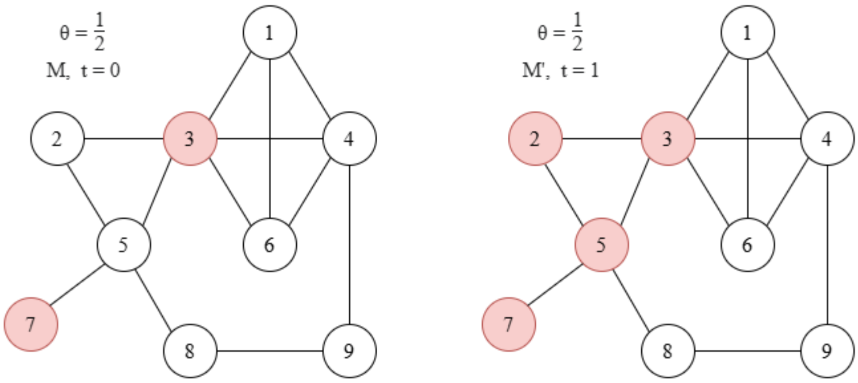

Figure 3.2: A threshold model M after one timestep.

By applying the update rule to the updated model M' again, a new model M'' is created. Doing this repeatedly to every new model, results in a *diffusion sequence*.

Definition 3.6 (Diffusion Sequence, [4])**.** Take threshold model $M = (A, N, B, \theta)$. A diffusion sequence $M^{(0)}, M^{(1)}, M^{(2)}, \ldots$ is a sequence of threshold models such that, for any $n \in \mathbb{N}$, $M^{(n)} = (A, N, B^{(n)}, \theta)$ where $B^{(n)}$ is given by $B^{(n+1)} = B^{(n)\prime}$ and $B^{(0)} = B$.

Eventually, this sequence stabilizes to a stable model, in the sense that no changes occur during an update anymore. This is called a *fixed point model* [4]. In this state, either every agent has adopted the behavior, called a *complete cascade*, or there is a group of agents that have not and will not ever adopt the behavior.

Example 3.7. These two two cases are shown in models M_1 and M_2 in Figure 3.3. Models M_1 and M_2 are almost the same, except there is an edge between agents 4 and 9 in model M_1, but not in M_2.

On top, model M_1 starts with the initial adopters of agents 3 and 7. With a threshold of $\theta = \frac{1}{2}$, agents 2 and 5 adopts this behavior in the next time step. Agent 8, agent 9, agent 4 all follow suit step by step and finally in the fifth timestep, agents 1 and 6 adopt the new behavior. After this timestep, the model does not change anymore, so a fixed point model is reached, and it is complete cascade.

On the other hand, model M_2 does not reach a complete cascade. After agent 9 adopts the new behavior, agent 4 is not influenced since there is no connection between agent 4 and agent 9 in this model. The fixed point model is reached after only 3 timesteps, and the agents 1, 4 and 6 do not adopt.

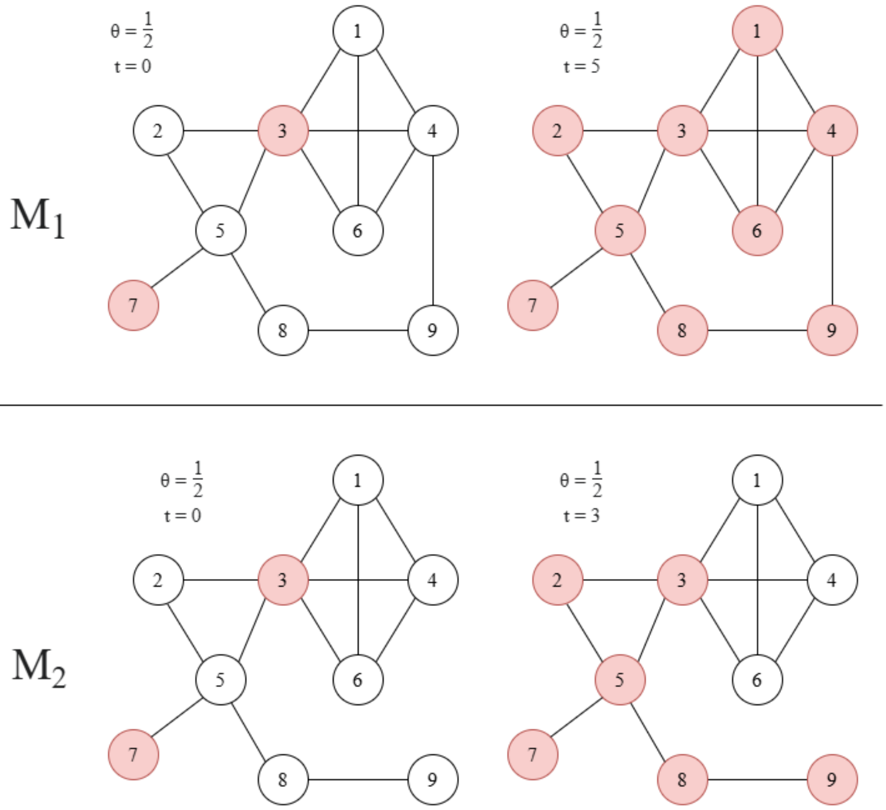

Figure 3.3: Two threshold models and their fixed point models. On top, model M_1 reaches a complete cascade after 5 timesteps. On the bottom, model M_2 reaches a point state where agents 1, 4 and 6 have not adopted the new behavior. Both models have a uniform threshold of $\theta = \frac{1}{2}$.

Why one of these models reaches a complete cascade while the other does not, depends entirely on the existence of a *cluster* in the remaining network.

Definition 3.8 (Cluster, [13]). In a network (A,N), a *cluster of density* $p \in (0,1]$ is a set of nodes $X \subseteq A$ such that each node $x \in X$ has at least p fraction of its neighbors in X.

In the threshold model M_1 from Figure 3.1, the agents 1, 4 and 6 form a cluster of density $\frac{1}{2}$. The density of a cluster can be found by looking for an agent who has the most connections outside of the cluster. In this case, agent 4 has 2 neighbors outside of the 146 cluster and 2 neighbors inside the cluster, thus it is the 'weak link'. Clusters can form an obstacle to complete cascades.

If the density of cluster of agents 1,4,6 from Example 3.7 is compared between models M_1 and M_2, we see that it is different. For model M_1, the cluster has a density of $\theta = \frac{1}{2}$, while it has a density of $\theta = \frac{2}{3}$ in model M_2. In both cases, the cluster $\{1,4,6\}$ lies outside of $A \setminus B$, but only in model M_2 does the density cluster exceed $1-\theta = \frac{1}{2}$, causing M_1 to have a complete cascade while M_2 does not.

In fact, they are the only obstacle to a behavior achieving a cascade.

Theorem 3.9 (Cluster Theorem, [13]). *Given a threshold model $M = (A,N,B,\theta)$, a complete cascade is reached eventually iff the network $A \setminus B$ does not contain a cluster of density greater than $1-\theta$.*

This is an important theorem for threshold models, as it allows someone to infer from just the structure of the graph and its initial adopters whether a new behavior will eventually spread throughout the entire network or not. In short, this theorem tells us that a dense enough cluster is the only thing stopping a behavior from effectively spreading throughout an entire network.

3.2 Variations of Threshold Models in the literature

In this section, two historically important models will be discussed that can both be qualified as threshold models, but differ a lot from each other.

The first idea of a threshold model is from the '70s. Though James M. Sakoda first developed a checkerboard dynamic model of social interaction, a version of it became the canonical model widely known as 'the Schelling Model', named after Thomas Schelling [32, 33, 34, 22]. In this model, agents of two types are put on a checkerboard with a threshold for how many neighbors they want of the same type. If they are dissatisfied, they will switch places with another dissatisfied agent that is in a spot where the threshold of the first agent is reached. the agents change position in the next timestep. Agents that previously had their threshold reached, might now have new neighbors, causing them to be dissatisfied. This causes a model that runs in discrete timesteps until everybody is happy. This model could potentially run forever.

The network structure of this model is a grid, where agents are only connected to their direct neighbors. For most agents, this will mean they have 8 neighbors, but some agents on the edge of the grid will have 4 or 6 neighbors. Though the number of agents of both types stays consistent, they move around, giving rise to the dynamic nature of this model. The threshold is uniform throughout the model, and can be absolute or relative. Sometimes, an upper threshold is enforced. For instance, an agent might be satisfied if it the number of neighbors of the same type lies between $\frac{1}{6}$ and $\frac{1}{2}$.

This type of threshold model was used specifically to model segregation. It shows how even mild preference to live close to people like you can lead to a society that is strongly segregated. Though this model was highly influential, its applications are limited due to the grid structure of the network and the stationary number of agents per group [13].

In 1973, Granovetter published the paper 'Threshold Models of collective behavior' [18]. At the time, most literature concerning a spread of preferences through a population were concerned with how people came about these preferences. Instead, Granovetter studied how individual preferences can explain the collective behavior. It is explained how individuals with differing thresholds can lead to cascading behavior in a crowd, where the paper mainly uses the example of a riot being instigated. The example sketched is of a crowd of 100 agents, each one having a threshold of 'I will join the riot if x other people have already joined'. One agent has threshold 0, one agent has threshold 1, one agent has threshold 2, up to one agent with threshold 99. The first agent, with threshold 0, will adopt in timestep 0. The next timestep, the agent with threshold 1 follows. This continues until everyone in the crowd has joined in. However, if the agent with threshold 0 also takes on a threshold of 1, no one will ever join the riot. Though only agent is different, the two threshold models have drastically different outcomes. This shows how local changes can have global effects.

This model more closely resembles a threshold model as introduced in the previous subsection. Granovetter names several applications of threshold models apart from rioting behavior: "innovation adoption, rumor or disease spreading, strikes, voting, going to college, leaving social occasions, migration, or conformity". He also highlights the usefulness of this new model: "Bandwagon effects can be imagined for each, but the sensitivity of such effects to exact distributions of preferences is rarely appreciated. Threshold models may be of particular value in understanding situations where the average level of preferences clearly runs strongly in favor of some action, but the action is not taken" [18].

There are many differences in these two models. Firstly, the network structure. Assuming that every agent only can see the behavior of their direct neighbors. In a Schelling Model this results in a grid, while Granovetter's threshold model more closely resembles a completely connected graph, everyone can see the behavior of every other agent.

Secondly, the threshold itself differs a lot. In a Schelling Model, the threshold is uniform fraction. In Granovetter's threshold model, it is an absolute number that varies per agent.

Thirdly, while both models use two different types of agents, they take a very different approach to this concept. In Granovetter's model, the two types of agents reflect whether they have made a choice or not. The network structure stays the same, but an agent can 'switch sides'. This is not possible in a Schelling model. Given that the different types of agents reflect different racial groups in the model, this property fits. The change in the model is the agents moving around, instead.

Thus the threshold model from Definition 3.2 is close to the model from Granovetter. The main differences are the assumption of a uniform threshold and that the network is not completely connected.

3.3 Defining the threshold using Game Theory

A threshold is inherent to a threshold model. How do agents set this threshold? Or is it set for them? Setting the threshold to be any value might seem arbitrary. Game theory provides us with an answer to how such a threshold would come to be.

Take any agent x, from any network, who has at least one neighbor. Agent x and her neighbor agent y, can both choose between options A and B. A *coordination game* between agent x and agent y ensues: if the two neighbors choose the same option, they both receive a high payoff, whereas them choosing differently results in a low payoff for both agents [13]. A commonly-used example of such a coordination game is two colleagues choosing between two different technologies that don't interact well. Choosing the same technology means they can communicate better and work on projects together, while choosing opposing technologies makes this a lot more difficult.

In game theory, often the goal is that every agent wants to achieve the highest possible payoff for themselves. In this case, that naturally means that both agents want to pick an option that the other agent will also pick. The difficulty in this coordination game lies in the fact that there are two ideal outcomes, making it hard for the agents to predict which option will be chosen. If the payoff b for behavior B is higher than the payoff a for behavior A, then we can assume both agents would choose behavior B. This would maximize payoff for both of them.

Definition 3.10 (Coordination game, [13])**.** A coordination game is defined by the payoff matrix given in Table 3.1, along with the following constraints: $p_{AA} > p_{BA}, p_{BB} > p_{AB}$ where $p_{AA}, p_{AB}, p_{BA}, p_{BB} \in \mathbb{R}$.

		Agent y: A	Agent y: B
Agent x	A	p_{AA}, p_{AA}	p_{AB}, p_{BA}
	B	p_{BA}, p_{AB}	p_{BB}, p_{BB}

Table 3.1: A payoff matrix between two agents. Payoff is listed as Agent x, Agent y.

The point is that it is beneficial for both agents to coordinate and choose the same. If an agent knows that their neighbor has chosen option A, then the agent would maximize their payoff if they also choose A, and ditto for option B.

We can fit a coordination game to a network of agents, where each pair of neighbors plays this coordination game with each other. We introduce the following simplification: The default behavior A has payoff a, while the new behavior B has payoff b. When the agents pick different options, they both receive 0 payoff. When $a \neq b$, it is called an *unbalanced coordination game*. If an agents wants to maximize the total payoff they get, they need to account for the amount of neighbors with a behavior, and the payoff associated with this.

		Agent y: A	Agent y: B
Agent x	A	a, a	0,0
	B	0,0	b, b

Table 3.2: A payoff matrix between agents x and y.

To describe it formally, take any agent x in a threshold model $M = (A, N, B, \theta)$, who has $N = |N(x)|$ neighbors. Agent x has $N_\beta = |N(x) \cap B|$ neighbors with behavior B and $N - N_\beta$ neighbors with behavior A. If x plays a coordination game with a neighbor, they have a payoff matrix as described in Table 3.2. If the agent plays this coordination game with all of their neighbors simultaneously, they would get a total payoff of either $N_\beta b$ when choosing option B or $(N - N_\beta)a$ when choosing default behavior A. So an agent chooses behavior B only if

$$\begin{aligned} N_\beta b &\geq (N - N_\beta)a \\ \frac{N_\beta}{N} &\geq \frac{a}{a+b} \end{aligned} \tag{1}$$

Define $\theta := \frac{b}{a+b}$. Then this corresponds to an agent choosing behavior B only if the fraction of neighbors with behavior B exceeds the threshold θ. Eq. (1) directly corresponds to the update rule:

$$\frac{N_\beta}{N} = \frac{|N(x) \cap B|}{|N(x)|} \geq \theta = \frac{a}{a+b}$$

If the payoffs a and b are the same for all agents, this would mean there is a uniform threshold of $\theta = \frac{1}{2}$ in the network. In general, a threshold lower than $\frac{1}{2}$ means that behavior B gives a higher payoff than behavior A does. For example, a threshold of $\theta = \frac{2}{5}$ would correspond to payoffs $a = 2$ and $b = 3$.

This take on a coordination game within a network gives every agent equal weight. In social settings, this might not be the case. The opinion of a close friend or family member would hold more weight than that of a colleague.

3.4 A logic for threshold models

A logical formalisation of threshold models and their updates is given in the paper ‘Dynamic Epistemic Logics of Diffusion and Prediction in Social Networks’ [4]. It approaches these threshold models from a logical and epistemic point of view. This subsection will explain the relevant parts of this paper.

A minimal logic is introduced to express this threshold model as described in Section 3.1. In order to describe the dynamic nature of a threshold model, the dynamic modality $[adopt]$ is included in the language of $\mathcal{L}_{[]}$.

Definition 3.11 (Syntax of $\mathcal{L}$ and $\mathcal{L}_{[]}$ [4])**.** Let A be a finite set and let atoms be given by $\Phi = \{N_{xy} : x,y \in A\} \cup \{\beta_x : x \in A\}$. We define the well-formed formulas of the language $\mathcal{L}_{[]}$ to be generated by:

$$\varphi := N_{xy} \mid \beta_x \mid \neg\varphi \mid \varphi \wedge \varphi \mid [adopt]\varphi$$

Any formula not involving the $[adopt]$ modality is a formula in $\mathcal{L}$.

Notably, there are no atomic propositions representing possible properties of agents. The only properties that are given using the atoms are whether two agents are neighbors and whether an agent has adopted or not. There is no need to define a seperate model, as the semantics are defined on a threshold model.

Definition 3.12 (Semantics of $\mathcal{L}_{[]}$ [4])**.** Given a model $\mathcal{M} = (\mathcal{A}, N, B, \theta)$, $N_{xy}, \beta_x \in \Phi$ and $\varphi, \psi \in \mathcal{L}_{[]}$:

$$\begin{aligned}
\mathcal{M} &\models \beta_x && \iff x \in B\\
\mathcal{M} &\models N_{xy} && \iff y \in N(x)\\
\mathcal{M} &\models \neg\varphi && \iff \mathcal{M} \not\models \varphi\\
\mathcal{M} &\models \varphi \wedge \psi && \iff \mathcal{M} \models \varphi \text{ and } \mathcal{M} \models \psi\\
\mathcal{M} &\models [adopt]\,\varphi && \iff \mathcal{M}' \models \varphi \text{ where } \mathcal{M}' \text{ is the updated threshold model (Definition 3.4)}
\end{aligned}$$

Note that the formulas are not satisfied at a specific agent, the atomic propositions are all satisfied in the entire model.

The set B bears some resemblance to a valuation function in modal logic: while the network (frame) stays the same, the set B (valuation V) gives meaning (truth value) to the agents: who partakes in the new behavior (which agents have formula p to be true)?

The following abbreviations are also introduced:

$$[adopt]^0\,\varphi = \varphi$$
$$[adopt]^n + 1 = [adopt]\,[adopt]^n\,\varphi$$

$$\beta_{N(x)\geq\theta} = \bigvee_{\{\mathcal{G}\subseteq\mathcal{N}\subseteq\mathcal{A}:\, \frac{|\mathcal{G}|}{|\mathcal{N}|}\geq\theta\}} \left(\bigwedge_{y\in\mathcal{N}} N_{xy} \wedge \bigwedge_{y\notin\mathcal{N}} \neg N_{xy} \wedge \bigwedge_{y\in\mathcal{G}} \beta_y \right)$$

The formula $\beta_{N(x)\geq\theta}$ is satisfied if the proportion of neighbors of agent a who have adopted is equal to or greater than θ.

Using a provided axiomatisation, the logic of threshold-limited influence L_θ is defined. These axioms include Network axioms detailing the properties of a network as defined in Definition 3.1, Reduction axioms regarding model updates and Inference rules. These are combined with any axiomatization of propositional logic to provide a complete and sound axiomatisation of L_θ. The class of threshold model $\mathcal{C}_\theta$ is defined as the class of models containing all and only models with the same threshold θ, for some $\theta \in [0,1]$. The paper also goes into a logical formalisation of the Cluster Theorem (Theorem 3.9).

Finally, the epistemic angle is discussed: what does an agent know about the network? The logic L_θ is extended to include a knowledge operator. If an agent has additional knowledge over the network, agents are able to predict which agents will adopt in the future. They can then adopt the behavior at an earlier timestep than they would have in L_θ.

4 Signed Threshold Graphs

Section 4.1 will define a signed threshold graph. In Section 4.2, it is described how agents can achieve their highest possible payoff in such a network, and under which conditions this is possible for every agent in a graph. In Section 4.3, a simplification of the coordination and anti-coordination games is described, together with an associated update rule and a discussion on the implications of this update rule.

4.1 Defining a Signed Threshold Graph

Within a threshold model, agents aim to be similar to their neighbors. The underlying assumption here is that agents have a favorable relationship to their neighbors. People usually try to be like their friends, but distance themselves from people they do not like.

The idea of a threshold model with positive and negative relations arises. The central question in such a model is: if an agent has multiple friends and enemies, some of which have adopted and some have not, will the agent adopt themselves?

To this end, the notion of a signed graph (Definition 2.1) is combined with a threshold model (Definition 3.2).

Definition 4.1 (Signed Threshold Graph)**.** A signed threshold graph is a tuple $M = (A, R^+, R^-, B, \theta)$ where A is a set of agents, $R^+ \subseteq A \times A$ is a set of positive edges, $R^- \subseteq A \times A$ is a set of negative edges, $B \subseteq A$ is a behavior and $\theta \in [0,1]$ is a uniform adoption threshold. It has the following properties:

1. $R^+ \cap R^- = \emptyset$ (non-overlapping)
2. $(x,y) \in R^* \iff (y,x) \in R^*$ for $i \in \{+,-\}$ (symmetric)
3. $(x,x) \notin R^+ \cup R^-$ for all $x \in A$ (irreflexive)

Notably, the relations R^+ and R^- of a signed graph define the network structure. The properties non-overlapping and symmetry are inherited from a signed graph. The model is irreflexive because an agents choice depends solely on their neighbors, which does not include their own adoption status. A signed graph is positively reflexive, but this property can be safely removed as it does not alter the sign of a cycle, and thus does not impact whether a graph is balanced or not.

Though the network structure of signed graphs has been applied, the neighborhood sets of the threshold model networks are still very useful, so we introduce them here as notation.

Notation. For a model $M = (A, R^+, R^-, B, \theta)$, the neighbor sets of agent x are defined as follows:

$$N^+(x) = \{y \in A : (x,y) \in R^+\} \qquad \text{(set of friends)}$$
$$N^-(x) = \{y \in A : (x,y) \in R^-\} \qquad \text{(set of enemies)}$$
$$N(x) = N^+(x) \cup N^-(x) \qquad \text{(set of neighbors)}$$

It is simple to make the transition from a signed threshold graph to a threshold model: set $R^- = \emptyset$ and define the neighborhood function such that $y \in N(x)$ iff xR^+y. Similarly, transitioning from a signed threshold graph to a signed graph can be done by setting $B = \emptyset$ and removing the threshold, leaving only the graph structure.

In a signed threshold graph, a behavior B spreads through both friends and enemies. Agents thus do not only play a coordination game with their friends, they also play an *anti-coordination game* with their enemies. This is defined similar to how a coordination game was defined in Definition 3.10.

Definition 4.2 (Anti-coordination game, [13])**.** An anti-coordination game is defined by the payoff matrix given in Table 4.1, along with the following constraints: $p_{BA} > p_{AA}$ and $p_{AB} > p_{BB}$ for $p_{AA}, p_{AB}, p_{BA}, p_{BB} \in \mathbb{R}$

		Agent y: A	Agent y: B
Agent x	A	p_{AA}, p_{AA}	p_{AB}, p_{BA}
	B	p_{BA}, p_{AB}	p_{BB}, p_{BB}

Table 4.1: The payoff matrix for an anti-coordination game, where payoff is given as $p_{\text{choice x, choice y}}$

In fact, the payoff matrix is the exact same as that of a coordination game, just the constraints have changed: both agent x and their enemy y get the highest payoff if they do not cooperate.

4.2 Maximal payoff

Since an agent plays the coordination game with friends at the same time as they play the anti-coordination game with their enemies, it is helpful to have both payoff matrices side-by-side. These numbers can be different, so a new notation of $f_{\text{choice x, choice y}}$ will be used for payoff from a friend and $e_{\text{choice x, choice y}}$ for payoff from an enemy.

In accordance with the definitions given in Definition 3.10 and Definition 4.2 respectively, the constraints for a coordination game is defined by Table 4.2 and the constraints $f_{AA} > f_{BA}$, $f_{BB} > f_{AB}$ for $f_{AA}, f_{AB}, f_{BA}, f_{BB} \in \mathbb{R}$, while an anti-coordination game is defined by Table 4.3 and the constraints $e_{BA} > e_{AA}$ and $e_{AB} > e_{BB}$ for $e_{AA}, e_{AB}, e_{BA}, e_{BB} \in \mathbb{R}$.

		Friend y	
		A	B
Agent x	A	f_{AA}, f_{AA}	f_{AB}, f_{BA}
	B	f_{BA}, f_{AB}	f_{BB}, f_{BB}

Table 4.2: A coordination game between agent x and their friend y, where payoff is given as $f_{\text{choice x, choice y}}$

		Enemy y	
		A	B
Agent x	A	e_{AA}, e_{AA}	e_{AB}, e_{BA}
	B	e_{BA}, e_{AB}	e_{BB}, e_{BB}

Table 4.3: An anti-coordination game between agent x and their friend y, where payoff is given as $e_{\text{choice x, choice y}}$

What would the payoff of an agent in a signed threshold graph then be? It depends on what how many friends and enemies an agent has, as well as the portion of friends and enemies that have adopted. To make the corresponding calculations, the following notation is used:

$$\begin{aligned} N_f &= |N^+(x)| && \text{(number of friends)} \\ N_{f,\beta} &= |N^+(x) \cap B| && \text{(number of adopted friends)} \\ N_e &= |N^-(x)| && \text{(number of enemies)} \\ N_{e,\beta} &= |N^-(x) \cap B| && \text{(number of adopted enemies)} \end{aligned} \tag{2}$$

Given a signed threshold graph M with set of adopted agents B, take any agent $x \in A$. If $x \in B$, the payoff they get from friends is $f_{BA}(N_f - N_{f,\beta}) + f_B N_{f,\beta}$ Similarly, x has an enemy payoff of $e_{BA}(N_e - N_{e,\beta}) + e_B N_{e,\beta}$.
Similar calculations can be made for when $x \notin B$, resulting in the payoff formula given in Equation (3).

$$\text{PAYOFF}(x) = \begin{cases} \text{PAYOFF}_{friend}(x \in B) + \text{PAYOFF}_{enemy}(x \in B) & \text{if } x \in B \\ \text{PAYOFF}_{friend}(x \notin B) + \text{PAYOFF}_{enemy}(x \notin B) & \text{otherwise} \end{cases} \tag{3}$$

where

$$\begin{aligned} \text{PAYOFF}_{friend}(x \in B) &= f_{BA}(N_f - N_{f,\beta}) + f_{BB} N_{f,\beta} \\ \text{PAYOFF}_{enemy}(x \in B) &= e_{BA}(N_e - N_{e,\beta}) + e_{BB} N_{e,\beta} \\ \text{PAYOFF}_{friend}(x \notin B) &= f_{AA}(N_f - N_{f,\beta}) + f_{AB} N_{f,\beta} \\ \text{PAYOFF}_{enemy}(x \notin B) &= e_{AA}(N_e - N_{e,\beta}) + e_{AB} N_{e,\beta} \end{aligned} \tag{4}$$

If we have a signed threshold graph M with a set network structure and set threshold, but the set B is variable, how can we choose the set B such that agent x has the highest possible payoff? Lemma 4.3 proves that this happens if all their friends have made the same choice as them and all their enemies have made the opposite choice.

Lemma 4.3. *Given any signed threshold graph $M = (A, R^+, R^-, B_0, \theta)$. For any agent $x \in A$,*

$$\arg\max_{B \subseteq A} \text{PAYOFF}(x, B) = \begin{cases} \{B \subseteq A : N^+(x) \subseteq B, N^-(x) \cap B = \emptyset\} & \text{if } x \in B \\ \{B \subseteq A : N^-(x) \subseteq B, N^+(x) \cap B = \emptyset\} & \text{otherwise} \end{cases}$$

Proof. Equation (3) shows the most general payoff for any agent in any signed threshold graph for any set B. In order to determine for which B this payoff is highest, PAYOFF needs to be a function of the set B. Thus, $\text{PAYOFF}(x) = \text{PAYOFF}(x, B)$, where the variables are $N_{f,\beta}$ and $N_{e,\beta}$, since these are dependent on B.

Assume $x \in B$. The payoff of x is highest if both

$$\text{PAYOFF}_{friend}(x \in B, B) = f_{BA}(N_f - N_{f,\beta}) + f_{BB} N_{f,\beta} = f_{BA} N_f + (f_{BB} - f_{BA}) N_{f,\beta}$$

and

$$\text{PAYOFF}_{enemy}(x \in B, B) = e_{BA}(N_e - N_{e,\beta}) + e_{BB}N_{e,\beta} = e_{BA}N_e + (e_{BB} - e_{BA})N_{e,\beta}$$

are maximized.

Since $f_{BB} > f_{BA}$, $\text{PAYOFF}_{friend}(x \in B, B)$ is a strictly increasing function of $N_{f,\beta} \in \{0, \ldots, N_f\}$ with the highest payoff when

$$N_{f,\beta} = N_f \iff |N^+(x) \cap B| = |N^+(x)| \iff N^+(x) \subseteq B$$

On the other hand, $e_{BA} > e_{BB}$, so $\text{PAYOFF}_{enemy}(x \in B, B)$ is strictly decreasing function of $N_{e,\beta} \in \{0, \ldots, N_e\}$. Thus, the highest possible payoff is when

$$N_{e,\beta} = 0 \iff |N^-(x) \cap B| = 0 \iff N^-(x) \cap B = \emptyset$$

Combining the payoff from both friends and enemies, $x \in B$ has maximum payoff if both $N^+(x) \subseteq B$ and $N^-(x) \cap B = \emptyset$.

Assume $x \notin B$. The payoff of x is highest if both

$$\text{PAYOFF}_{friend}(x \notin B, B) = f_{AA}(N_f - N_{f,\beta}) + f_{AB}N_{f,\beta} = f_{AA}N_f + (f_{AB} - f_{AA})N_{f,\beta}$$

and

$$\text{PAYOFF}_{enemy}(x \notin B, B) = e_{AA}(N_e - N_{e,\beta}) + e_{AB}N_{e,\beta} = e_{AA}N_e + (e_{AB} - e_{AA})N_{e,\beta}$$

are maximized.

Since $f_{AA} > f_{AB}$ by definition, $\text{PAYOFF}_{friend}(x \notin B, B)$ is a strictly decreasing function of $N_{f,\beta} \in \{0, \ldots, N_f\}$, thus the function reaches it maximum at

$$N_{f,\beta} = 0 \iff |N^+(x) \cap B| = 0 \iff N^+(x) \cap B = \emptyset$$

On the other hand, $e_{AB} > e_{AA}$ by definition, making $\text{PAYOFF}_{enemy}(x \notin B, B)$ a strictly increasing function of $N_{e,\beta} \in \{0, \ldots, N_e\}$. Thus, the function reaches its maximum if

$$N_{e,\beta} = N_e \iff |N^-(x) \cap B| = |N^-(x)| \iff N^-(x) \subseteq B$$

Combining the payoff from both friends and enemies, $x \notin B$ has maximum payoff if $N^-(x) \subseteq B$ and $N^+(x) \cap B = \emptyset$ □

While Lemma 4.3 gives a result for maximal payoff of an individual agent, Theorem 4.6 gives a result for maximal payoff for *all* agents in a balanced signed graph. We define a balanced signed threshold graph.

Definition 4.4 (Balanced Signed Threshold Graph). A signed threshold graph $M = (A, R^+, R^-, B, \theta)$ is balanced if the signed graph (A, R^+, R^-) is balanced.

We also define the state of a signed graph such that every agent receives maximal payoff.

Definition 4.5 (Social Optimum). Define $\text{PAYOFF}(x, B) \geq \text{PAYOFF}(x, C)$ to be true iff agent x in model (A, R^+, R^-, B, θ) receives equal or higher payoff than the same agent x in model (A, R^+, R^-, C, θ) receives. A signed threshold graph $M = (A, R^+, R^-, B, \theta)$ is a *social optimum* iff $\forall x \in A, \forall C \subseteq A : \text{PAYOFF}(x, B) \geq \text{PAYOFF}(x, C)$.

The set $C \subseteq A$ is here ‘an alternative set of adopted agents’. A social optimum is reached when every agent has the maximum payoff possible. In other words, no agent would get higher payoff when the set B is changed.

Theorem 4.6. *Given a balanced signed threshold graph $M = (A, R^+, R^-, B, \theta)$ with a balance partition $A = A_1 \cup_\triangle A_2$,*

$$M \text{ is a social optimum} \iff B = A_1 \text{ or } B = A_2$$

Proof. First, we investigate the case where $B = \emptyset$, implying that $B = A_1$ or $B = A_2$ iff A_1 or A_2 is empty. By definition of $A_1 \cup_\triangle A_2$, this is true iff there are no negative edges in the graph. $R^- = \emptyset$ iff $N^-(x) \subseteq B$ and $N^+(x) \cap B = \emptyset$ for all $x \in A$. Additionally, $x \notin B$ for all $x \in A$. Lemma 4.3 tells us that this is true iff $\text{PAYOFF}(x, B) \geq \text{PAYOFF}(x, C)$ for all $x \in A$, for all $C \subseteq A$. Thus, M is a social optimum.

Assume $B \neq \emptyset$ and without loss of generality, assume $B = A_1$.

Assume $x \in B$.
$x \in B$ iff $x \in A_1$. By Definition of balance partition, $x \in A_1$ if and only if for all $y \in N^+(x) : y \in A_1$ and for all $y \in N^-(x) : y \notin A_1$. This holds iff $N^+(x) \subseteq A_1$ and $N^-(x) \cap A_1 = \emptyset$. By Lemma 4.3, this is true if and only if $A_1 \in \arg\max_{B \subseteq A} \text{PAYOFF}(x,B)$. By Definition of the arg max, $\text{PAYOFF}(x,A_1) \geq \text{PAYOFF}(x,C)$ for all $x \in B$, for all $C \subseteq A$.

Assume $x \notin B$.
$x \in B$ if and only $x \notin A_1$. By Definition of balance partition, $x \notin A_1$ if and only if for all $y \in N^+(x) : y \notin A_1$ and for all $y \in N^-(x) : y \in A_1$. This holds iff $N^+(x) \cap A_1 = \emptyset$ and $N^-(x) \subseteq A_1$. By Lemma 4.3, this is true iff $A_1 \in \arg\max_{B \subseteq A} \text{PAYOFF}(x,B)$. By Definition of the arg max, $\text{PAYOFF}(x,A_1) \geq \text{PAYOFF}(x,C)$ for all $x \notin B$, for all $C \subseteq A$.

Since $\text{PAYOFF}(x,A_1) \geq \text{PAYOFF}(x,C)$ for all $C \subseteq A$ holds for all $x \in A$, we can conclude that $B = A_1$ or $B = A_2$ iff M is a social optimum. □

In a balanced signed threshold graph, it is thus possible for all agents to achieve maximal payoff. Theorem 4.7 says this is possible only when a signed threshold graph is balanced.

Theorem 4.7. *Let $M = (A, R^+, R^-, B, \theta)$ be a signed threshold graph.*

$$\exists B_0 \subseteq A \text{ such that } M_0 = (A, R^+, R^-, B_0, \theta) \text{ is a social optimum} \iff M \text{ balanced.}$$

Proof. ($\Rightarrow$) Take any signed threshold graph $M = (A, R^+, R^-, B, \theta)$ such that M is a social optimum: $\forall x \in A$ and $\forall C \subseteq A$, $\text{PAYOFF}(x,B) \geq \text{PAYOFF}(x,C)$. This holds if and only if $B \in \arg\max_{B \subseteq A} \text{PAYOFF}(x,B)$ for all $x \in A$.

First, we investigate the case where $B = \emptyset$. B is empty iff $x \notin B$ for all $x \in A$. Thus, by Lemma 4.3, $\emptyset \in \arg\max_{B \subseteq A} \text{PAYOFF}(x,B)$ iff $N^-(x) \subseteq \emptyset$ and $N^+(x) \cap \emptyset = \emptyset$ for all $x \in A$. Since $N^+(x) \cap \emptyset = \emptyset$ is trivially true, B is empty if and only if $N^-(x) \subseteq \emptyset$ for all $x \in A$, meaning no agent has an enemy. By Definition 4.4 and Balance Theorem part 2, this implies that M is balanced.

Now take B to be non-empty and let $x \in B$. Lemma 4.3 implies that $N^+(x) \subseteq B$ and $N^-(x) \cap B = \emptyset$. If $(x,y) \in R^+$, then $y \in B$. If $(x,y) \in R^-$, then $y \notin B$. Since this holds for all $(x,y) \in R^+$ and all $(x,y) \in R^-$, M is balanced according to the Balance Theorem, part 3.

($\Leftarrow$) Assume M is balanced, then there exists a balance partition $A = A_1 \cup_{\triangle} A_2$ by Balance Theorem, part 3. Without loss of generality, take $B = A_1$. By Theorem 4.6, M is a social optimum. □

This proves that maximal payoff for all agents is only possible if the graph is balanced. Any unbalanced graph has at least one agent in that does not receive maximal payoff.

4.3 A simplified adoption game

In the previous sections, only the most general definition of a coordination game and anti-coordination game were used. This gave us some results that hold for any signed threshold graph. But if we want to investigate a practical example, we need to make a simplification.

The simplification is based on the idea that there are three levels of payoff: positive, neutral or negative. An agent would get positive payoff if they cooperate with a friend. Coordinating with an enemy or anti-coordinating with a friend is unfortunate, but neutral: it doesn't do any harm and no payoff is lost or received. Finally, coordinating with an enemy is bad, and an agent would get negative payoff.

For example; an agent has two friends, one of which has adopted. Whether the agent chooses to adopt or not, they coordinate with only one friend, and the agent and their coordinating friend both get some positive payoff. However, the agent does not lose any payoff from not coordinating with the second friend: while they gain no *advantage*, there is also no *disadvantage*. It is a neutral choice.

Friend y

Agent x		A	B
	A	a,a	0,0
	B	0,0	b,b

Enemy y

Agent x		A	B
	A	$-a,-a$	0,0
	B	0,0	$-b,-b$

Table 4.4: The payoff of agent x in a signed threshold graph.

This leads to the following simplifications: $f_{AA} = a = -e_{AA}, f_{BB} = b = -e_{BB}$ and $f_{AB} = f_{BA} = e_{AB} = e_{BA} = 0$, as are shown in Table 4.4. We assume that $a, b > 0$ to maintain the original constraints of the games. Note that the payoff matrix for the coordination game coincides with the payoff matrix of a neighbor (**??**).

Given these payoff matrices, what would the choice between A and B look like for agent x in a model $M = (A, R^+, R^-, B, \theta)$? Using Equation (4) and Table 4.4, we can see that behavior A will give a total of $(N_f - N_{f,\beta})a$ payoff from friends and $-(N_e - N_{e,\beta})a$ payoff from enemies. Choosing behavior B will give $N_{f,\beta}b$ payoff from friends and $-N_{e,\beta}b$ payoff from enemies. B is then a better choice if:

$$N_{f,\beta}b - N_{e,\beta}b \geq (N_f - N_{f,\beta})a - (N_e - N_{e,\beta})a$$
$$N_{f,\beta}b + N_{f,\beta}a - N_{e,\beta}b - N_{e,\beta}a \geq N_f a - N_e a$$
$$(N_{f,\beta} - N_{e,\beta})(a+b) \geq (N_f - N_e)a$$
$$N_{f,\beta} - N_{e,\beta} \geq \frac{a}{a+b}(N_f - N_e) \tag{5}$$

Since all agents make the choice between A and B simultaneously, the model update include all agents for which the above equation is true. Define $\theta = \frac{a}{a+b}$. Then this corresponds to the following update rule:

Definition 4.8 (Update Rule for a Signed Threshold Graph)**.** The update of a signed threshold graph $M = (A, R^+, R^-, B, \theta)$ is the signed threshold graph $M' = (A, R^+, R^-, B', \theta)$ where B' is given by

$$B' = B \cup \{x \in A : |N^+(x) \cap B| - |N^-(x) \cap B| \geq \theta(|N^+(x)| - |N^-(x)|)\}$$

4.3.1 Examples

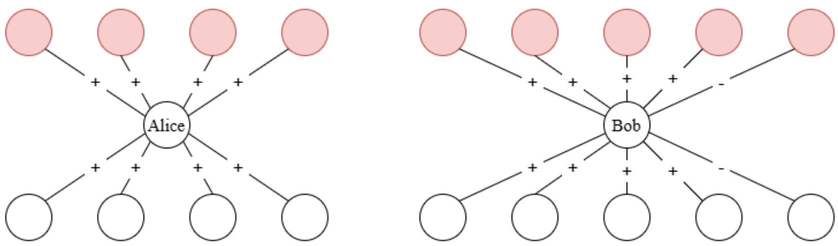

Figure 4.1: Two signed threshold models. Agents who have adopted are colored red, agents who have not are colored white.

Example 4.9. The agent Alice has 8 friends, half of which have the new behavior. She has no enemies. Then $|N^+(x) \cap B| - |N^-(x) \cap B| = 4$ and $|N^+(x)| - |N^-(x)|$ is 8. Alice only adopts if her threshold is lower than or equal to $\frac{1}{2}$.

Example 4.10. Agent Bob has 10 neighbors, 2 enemies and 8 friends. For both his friends and enemies, half of them have adopted the new behavior. Thus, $|N^+(x) \cap B| - |N^-(x) \cap B| = 3$ and $|N^+(x)| - |N^-(x)| = 6$, so Bob will only adopt if his threshold is lower than or equal to $\frac{1}{2}$.

These two examples are also depicted in Figure 4.1. Bob and Alice have the same amount of friends, and the same fraction of these friends have adopted the new behavior. For Alice, the situation intuitively makes sense: if she has a threshold of $\theta = \frac{1}{2}$, and half of her friends have adopted, then she adopts. Bob is in a similar situation, with half of his friends that have adopted, but he also has one enemy that has adopted. Bob's situation with regards to these two enemies is that whether Bob adopts or not, he gets 0 payoff from one agent and negative payoff from the other. Thus, his situation does not change from that of Alice: it still depends on which option has the highest payoff, and thus depends on the same threshold.

Example 4.11. We have a network, where everyone dislikes each other. This network is not balanced. For example, take the signed threshold graph on the left in Fig. 4.2. Every agent in this network has no friends, and has three enemies, none of which have adopted. An agent in this situation will adopt whenever $0 \geq -3\theta$. Given that $\theta \geq 0$, this is always true, so every agent adopts. The updated graph is depicted in the right graph of the same figure. Since every agent has adopted at the same time (to rebel against their enemies), every agent now is stuck in the situation where they are accidentally coordinating with their enemies. Even if these agents would be able to unadopt, they would all unadopt at the same time and the graph would return to the graph depicted on the left. The diffusion sequence would get stuck in a perpetual loop between these two graphs.

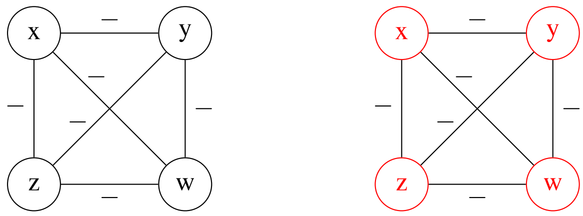

Figure 4.2: Two signed threshold graph, nodes in red are nodes that are in the set B. The right graph is the model update of the left graph.

4.3.2 Discussion

Note that if an agent has no enemies, this update rule corresponds to the original update rule on threshold models for Definition 3.4. This is a direct result of the payoff matrix with a friend from Table 4.4 is equal to the payoff matrix with a neighbor given in Table 3.2. On the other hand, if an agent has no friends an only enemies, the agents adopt iff $\frac{|N^-(x) \cap B|}{|N^-(x)|} \leq \theta$. This mirrors the friends-only rule, and corresponds with the idea that if too much of your enemies have adopted, you do not want to follow along. Additionally, the threshold θ is defined as the fraction $\frac{a}{a+b}$, just as θ was previously defined is Section 3.3. In these ways, the model mirrors the threshold model as described in Section 3.

However, there is a big difference with a threshold model: an agent can adopt 'out of the blue'. A side effect of adopting if many of your enemies have not adopted, is that they need not have any neighbors with behavior B to adopt the behavior themselves.

Firstly, agents adopting out of the blue directly implies that there is some element of common knowledge in the network: every agent at least knows about the new behavior B. In a threshold model, an agent only knows about their immediate neighbors and their behavior. This epistemic angle of threshold models is explored in [4]. For further usage, if common knowledge is not desired, this could easily be remedied by including some notion of 'at least one neighbor has adopted' to the update rule for an agent.

Thus, the notion of a cascade is not applicable here. A behavior does not 'spread' through a network, but it pops up in some places. Specifically, if an agent only has enemies and none of them have adopted, the agent adopts by default.

Lastly, the threshold would act as a tie-breaker in an ideal situation; it is a constant that shows whether option B or option A has higher payoff. So if an agent has the same amount as friends and enemies that have adopted the behavior, it makes sense for an agent to choose the option that has the highest payoff. For this update rule, if either the right-hand side or the left side of Equation (5) becomes 0, the threshold θ, and thus the payoff from either options, plays no role in decision anymore: an agent is solely decision based on the number of (adopted) neighbors. This is a result of the symmetry of the friend and enemy payoff matrices.

In the case where an agent has equal number of friends and enemies, they adopt only if the number of adopted friends is higher than the number of adopted enemies. In the case where an agent has equal number of adopted friends and enemies, they adopt only if the total number of friends is smaller than the number of enemies: the fraction of adopted friends is higher than the fraction of adopted enemies.

If both sides of the equation are zero, the tie is simply decided by the $\geq$-sign. This favors the new option B.

4.3.3 Achieving maximal payoff

Previously, it was proven that maximal payoff for every agent is only possible if the graph is balanced, and thus the set of agents A can be divided into two friendgroups A_1 and A_2. Maximal payoff can then be achieved iff the divide between the adopted and unadopted agents is equal to the divide between the two friendgroups. But how attainable is it to reach such an optimal situation?

When an agent adopts, they cannot unadopt. It is clear that two agents from the two different friendgroups both adopting the behavior then poses an obstacle to attaining maximal payoff. Naturally, this raises the question: if only agents from one friendgroup have adopted, what would cause an agent from the other friendgroup to adopt? We want to find networks for which this cannot happen, as these are the graphs where a social optimum is possible.

In Section 4.3.2 it was already stated that an agent with only enemies can adopt a behavior out of the blue. Assuming that the set B starts as a subset of A_1, the goal is to keep behavior B within friendgroup A_1. But any agent of friendgroup A_2 with only unadopted enemies would cause the behavior to spread outside of A_1. This suggests that lacking friends makes an agent more susceptible to adopting the behavior. To this end, we define a *cluster of friends*.

Definition 4.12 (Cluster of Friends)**.** A cluster of friends is a set $C \subseteq A$ such that for all $x \in C : |N^+(x)| > |N^-(x)|$ and $N^+(x) \subseteq C$.

The following lemma proves that a cluster of friends form an obstacle to adopting, similarly to how clusters in threshold models block the cascade.

Lemma 4.13. *Let $M = (A, R^+, R^-, B, \theta)$ be a balanced signed threshold graph where $A = A_1 \cup_{\triangle} A_2$, $\theta > 0$ and $B \subseteq A_1$. Let $C \subseteq A_2$ be a cluster of friends. Let $B^{(k)}$ be the set of agents that have adopted after k model updates. Then,*

$$C \cap B^{(k)} = \emptyset \text{ for all } k \in \mathbb{N}$$

Proof. Proof by contradiction. Given any balanced graph $M = (A, R^+, R^-, B, \theta)$ such that $A = A_1 \cup_{\triangle} A_2$, $\theta > 0$ and $B \subseteq A_1$. Assume that there exists a set of agents $C \subseteq A_2$ such that C is a cluster of friends, then $C \cap B = C \cap B^{(0)} = \emptyset$.

Assume that $C \cap B^{(k)} \neq \emptyset$ for some $k \geq 1$. Take i such that $M^{(i)} = (A, R^+, R^-, B^{(i)}, \theta)$ is the first model update where $C \cap B^{(i)}$ is non-empty, and take agent x to be in this intersection, which implies $x \in B^{(i)}$. This is true if and only if $x \in B^{(i-1)}$ or $|N^+(x) \cap B^{(i-1)}| - |N^-(x) \cap B^{(i-1)}| \geq \theta(|N^+(x)| - |N^-(x)|)$. Agent x has not adopted at timestep $i-1$ since i is the earliest timestep where x has adopted. Thus, $|N^+(x) \cap B^{(i-1)}| - |N^-(x) \cap B^{(i-1)}| \geq \theta(|N^+(x)| - |N^-(x)|)$ must be true. Since $\theta > 0$ and $|N^+(x)| - |N^-(x)| > 0$, this implies that $|N^+(x) \cap B^{(i-1)}| > |N^-(x) \cap B^{(i-1)}|$ is true. The assumptions of $N^+(x) \subseteq C$ for all $x \in C$ and $C \cap B^{(i-1)} = \emptyset$ implies that $|N^+(x) \cap B^{(i-1)}| = 0$. Thus $|N^+(x) \cap B^{(i-1)}| > |N^-(x) \cap B^{(i-1)}|$ is only true if $|N^-(x) \cap B^{(i-1)}| < 0$ which contradicts cardinality of a set. Therefore, there does not exist k for which $C \cap B^{(k)}$ is non-empty.

($\Leftarrow$) Assume that $C \cap B^{(k)} = \emptyset$ for all $k \in \mathbb{N}$, then this also holds for $k = 0$. □

The following theorem is a natural conclusion:

Lemma 4.14. *Let $M = (A, R^+, R^-, B, \theta)$ be balanced signed threshold graph where $A = A_1 \cup_{\triangle} A_2$, $\theta > 0$ and $B \subseteq A_1$. Let $B^{(k)}$ be the set of agents that have adopted after k model updates.*

$$|N^+(x)| > |N^-(x)| \text{ for all } x \in A_2 \implies B^{(k)} \subseteq A_1 \text{ for all } k \in \mathbb{N}$$

Proof. Given any balanced graph $M = (A, R^+, R^-, B, \theta)$ where $A = A_1 \cup_{\triangle} A_2$, $\theta > 0$ and $B \subseteq A_1$, assume that $|N^+(x)| > |N^-(x)|$ for all $x \in A_2$. Then A_2 is a cluster of friends, since balance partition says that $N^+(x) \subseteq A_2$ holds for any $x \in A_2$. Also, $B \subseteq A_1$ implies that $B \cap A_2$ is empty. Thus, the conditions for Lemma 4.13 are fulfilled, which implies that $A_2 \cap B^{(k)} = \emptyset$ for all $k \in \mathbb{N}$. Thus, $B^{(k)} \subseteq A_1$ for all $k \in \mathbb{N}$. □

Since behavior does not cascade through a model the same way it does in a threshold model, there is no trivial translation of clusters and the cluster theorem to signed threshold graphs. Nevertheless, the above result proves that achieving maximal payoff for every agent is possible under certain circumstances.

5 A Logic for Signed Threshold Models

This section defines a dynamic modal logic. Section 5.1 first defines the language, the model, the model update and the semantics of the language. Section 5.2 discusses how to construct an update rule for the model. Afterwards, Section 5.3 gives an axiomatisation of the class of balanced signed threshold models. Section 5.4 proves the soundness of this axiomatisation. Section 5.5 proves the completeness of this axiomatisation.

5.1 Syntax and Semantics

A logic is defined by formulas in a language. It is necesary to define this language first.

Take AT to be a set of propositional atoms, denoted by letters $p,q,r,\dots$, which are understood to be properties of agents. Take β to be a propositional atom not in AT, which corresponds to the property of 'has adopted behavior B'. The language $\mathcal{L}_{[+/-]}$ is then given as follows.

Definition 5.1 (Syntax of $\mathcal{L}_{[+/-]}$ and $\mathcal{L}_{+/-}$)**.** Let AT be a countable set of propositional atoms, and let $\Phi = \text{AT} \cup \{\beta\}$, where $\beta \notin \text{AT}$ is also a propositional atom. We define the well-formed formulas of the language $\mathcal{L}_{[+/-]}$ to be generated by the following grammar:

$$\varphi := p \mid \beta \mid \neg\varphi \mid \varphi \wedge \varphi \mid \diamond\!\!\!+\,\varphi \mid \diamond\!\!\!-\,\varphi \mid [adopt]\varphi \mid \langle D\rangle\varphi$$

where $p \in \text{AT}$. Additionally, the following shorthand notation for the propositional connectives and duals will be used:

$$\begin{aligned}
\varphi \vee \psi &\iff \neg(\neg\varphi \wedge \neg\psi)\\
\varphi \to \psi &\iff \neg\varphi \vee \psi\\
\varphi \leftrightarrow \psi &\iff (\varphi \to \psi) \wedge (\psi \to \varphi)\\
\boxplus\varphi &\iff \neg \diamond\!\!\!+\,\neg\varphi\\
\boxminus\varphi &\iff \neg \diamond\!\!\!-\,\neg\varphi\\
[D]\varphi &\iff \neg\langle D\rangle\neg\varphi\\
[adopt]^n\varphi &\iff [adopt]\,[adopt]^{n-1}\,\varphi, \quad [adopt]^0\,\varphi = \varphi
\end{aligned}$$

The language $\mathcal{L}_{+/-}$ is defined as all formulas that do not contain the $[adopt]$ modality, and is called the *static part of the language* $\mathcal{L}_{[+/-]}$.

This language contains all elements from $\mathbf{pnl}_n$, and adds the propositional atom β and dynamic operator $[adopt]$ from L_θ. The difference operator $\langle D\rangle$ and its dual $[D]$ are not from either logic, but its inclusion in the language is inspired by the proposal of multiple extensions on $\mathbf{pnl}_n$ (see Section 2.3.2). The key usage for $\langle D\rangle$ is on expressivity.

The formulas are evaluated on a model.

Definition 5.2 ((Signed Threshold) Frame / Model)**.** A (signed threshold) frame is a tuple $F = (A, R^+, R^-)$, where A is a set of agents and $R^+ \subseteq A \times A$ and $R^- \subseteq A \times A$ are sets of respectively positive and negative relations between agents. These relations are irreflexive, non-overlapping and symmetric. A (signed threshold) model $M = (F, V) = (A, R^+, R^-, V)$ is a tuple consisting of a frame and a valuation function $V : \Phi \to \mathcal{P}(A)$, which assigns a subset of agents to every atomic proposition.

Most notably, a signed threshold graph is different from a signed threshold model. The set B in a signed threshold graph represents the set of adopted agents. In this case, the valuation of proposition β determines this set. However, the threshold θ is not included in a signed threshold model, as the model update does not include this. An alternative update rule is defined in Definition 5.3.

Definition 5.3 (Update Rule for a Signed Threshold Model)**.** The update of a model $M = (A, R^+, R^-, V)$ is the model $M' = (A, R^+, R^-, V')$ where V' is given by

$$V'(\varphi) = \begin{cases} V(\beta) \cup \left(\left\{x \in A : |N^+(x) \cap B| \geq 1\right\} \cap \left(\left\{x \in A : |N^-(x)| = 0\right\} \cup \left\{x \in A : |N^-(x) \cap B| < |N^-(x)|\right\}\right)\right) & \text{if } \varphi = \beta \\ V(\varphi) & \text{otherwise} \end{cases}$$

This update rule tells us that an agent adopts (the agent x is in the updated valuation set of β) iff they have already adopted, or they have at least one friend that has adopted, and they have no enemies or at least one enemy has not adopted. For example, if an agent has only friends then one friend adopting is enough to convince the agent to adopt themselves. If the agent has a couple friends and one enemy, then the agent can never adopt if the enemy has adopted. If the agent has only enemies and no friends, the agent can never adopt. If the agent has adopted already, then they will stick to their choice regardless of how many enemies or friends this agent has.

Now we can define the semantics of the language.

Definition 5.4 (Semantics of $\mathcal{L}_{[+/-]}$)**.** For a model $M = (A, R^+, R^-, V)$, and formulas $\varphi, \psi \in \mathcal{L}_{[+/-]}$, the formula φ is satisfied by an agent $x \in A$, denoted $M, x \models \varphi$, as follows:

$$
\begin{aligned}
&M,x \models p && \iff x \in V(p)\\
&M,x \models \beta && \iff x \in V(\beta)\\
&M,x \models \neg\varphi && \iff M,x \not\models \varphi\\
&M,x \models \varphi \wedge \psi && \iff M,x \models \varphi \text{ and } M,x \models \psi\\
&M,x \models \Diamond^{+}\varphi && \iff \exists y \in A \text{ such that } xR^{+}y \text{ and } M,y \models \varphi\\
&M,x \models \Diamond^{-}\varphi && \iff \exists y \in A \text{ such that } xR^{-}y \text{ and } M,y \models \varphi\\
&M,x \models [adopt]\varphi && \iff M',x \models \varphi \text{ where } M' \text{ is the updated signed threshold model (Definition 5.3)}\\
&M,x \models \langle D\rangle\varphi && \iff \exists y \in A \text{ such that } y \neq x \text{ and } M,y \models \varphi
\end{aligned}
$$

Additionally, the abbreviations are satisfied when:

$$
\begin{aligned}
&M,x \models \varphi \vee \psi && \iff M,x \models \varphi \text{ or } M,y \models \psi\\
&M,x \models \varphi \rightarrow \psi && \iff M,x \models \varphi \text{ implies } M,x \models \psi\\
&M,x \models \varphi \leftrightarrow \psi && \iff M,x \models \varphi \text{ iff } M,x \models \psi\\
&M,x \models \boxplus\varphi && \iff \forall y \in A : xR^{+}y \text{ implies } M,y \models \varphi\\
&M,x \models \boxminus\varphi && \iff \forall y \in A : xR^{-}y \text{ implies } M,y \models \varphi\\
&M,x \models [D]\varphi && \iff \forall y \in A : x \neq y \text{ implies } M,y \models \varphi
\end{aligned}
$$

The following formulas are highlighted:

$$
\begin{aligned}
&M,x \models \Diamond^{+}\top && \iff \exists y \in A \text{ such that } xR^{+}y\\
&M,x \models \Diamond^{-}\top && \iff \exists y \in A \text{ such that } xR^{-}y\\
&M,x \models \boxplus\bot && \iff \nexists y \in A \text{ such that } xR^{+}y\\
&M,x \models \boxminus\bot && \iff \nexists y \in A \text{ such that } xR^{-}y
\end{aligned}
$$

It becomes clear why the update rule as defined in Definition 4.8 cannot be expressed in $L_{[\triangle\theta]}$. An agent that adhere to this update rule counts the number of friends, enemies, adopted friends and adopted enemies and then compares these numbers. The semantics of $\mathcal{L}_{[+/-]}$ does not allow for counting. Instead, when using the diamond and box modalities, only general statements about the number of neighbors with property φ can be made. For example, $\Diamond^{-}\beta$ expresses 'at least one enemy has φ' and $\neg\boxminus\beta$ expresses 'not all enemies have φ'.

This is reason to find a different update rule. Though Definition 5.3 expresses one such possible update rule, other update rules can be found. These different rules are then the different thresholds an agent can have. Section 5.2 discusses these different type of thresholds.

5.2 Variants of the Update Rule

$\Diamond^{+}\beta$	At least one friend has adopted
$\boxplus\beta$	All friends have adopted
$\Diamond^{-}\neg\beta$	At least one enemy has not adopted
$\boxminus\neg\beta$	No enemies have adopted

Table 5.1: The four formulas that serve as building blocks to construct an update rule.

The update rule of Definition 5.3 expresses one type of threshold, but this is not the only update rule that can be formulated. This subsection is a guide on how to construct an update rule for a signed threshold model in the language $\mathcal{L}_{[+/-]}$.

A logical update rule is made of 'building blocks' and $\wedge, \vee, \rightarrow$. The building blocks express a minimum or maximum number of friends or enemies, while $\wedge, \vee, \rightarrow$ connect these building blocks to form a requirement. For example, in the sentence 'An agent adopts if all of their friends have adopted or none of their enemies have adopted', the building blocks are 'all of their friends have adopted' and 'none of their enemies have adopted', while 'and' connects these two together to form one requirement. Let us then look at the building blocks: the formulas that can be constructed in $\mathcal{L}_{[+/-]}$ without $\wedge, \vee, \rightarrow$. These formulas are constructed by using $p, \beta, \neg, \Diamond^{+}, \Diamond^{-}$ and $\langle D\rangle$. These formulas should adhere to the idea of a threshold model. An agent needs a *minimum* number of friends and a *maximum* numbers of enemies before the agent adopts. An update rule that includes any of the formulas $\langle - \rangle\beta$, $\boxminus\beta$, $\langle + \rangle\neg\beta$ and $\boxplus\neg\beta$ would not express a useful threshold. Additionally, an update rule that would involve the $\langle D \rangle$ operator would not only depend on the neighbors of an agent, but also on arbitrary different agents in the network. This is not desirable, and we thus exclude $\langle D \rangle$.

This results in the four formulas listed in Table 5.1.

A final note before we can construct an update rule: If an agent has no friends, then $\langle + \rangle\varphi$ is false for every formula φ. Ditto for $\langle - \rangle\varphi$ whenever an agent has no enemies. If an update rule is desired that expresses that 'an agent adopts if not all of their enemies have adopted', the situation where an agent has no enemies is not included. To ensure that this update rule does not rule out this specific case, a formula '$\langle + \rangle\top \to \varphi$' can be included in the reduction axiom, which means 'if an agent has at least one friend, then φ'. Such a formula is then called an *existence condition formula*.

An update rule can then be constructed from $\wedge$, $\vee$,$\to$, the formulas from Table 5.1 and possibly an existence condition formula. Examples of reduction formulas are of the form $[adopt]\beta \leftrightarrow \beta \vee \ldots$

$\langle + \rangle\beta$: At least one friend has adopted.

$\langle + \rangle\top \to \langle + \rangle\beta$: If the agent has any friends, the agent adopts if one friend has adopted. This would imply that if the agent has no friends, they also adopt.

$(\langle + \rangle\top \to \langle + \rangle\beta) \wedge (\neg\langle + \rangle\top \to \boxplus\neg\beta)$: If the agent has any friends, the agent adopts if one friend has adopted, but if the agent has no friends, the agents adopt only if none of their enemies have adopted.

$\boxplus\beta \wedge \boxminus\neg\beta$: All friends have adopted and no enemies have adopted.

$\langle + \rangle\top \to \boxminus\neg\beta$: If the agent has any friends, then they adopt if none of their enemies have adopted.

These reduction axioms can be swapped in the axiomatisation $L_{[\triangle\theta]}$ to account for different update rules.

Finally, while this is not considered in this thesis to be relevant, it is possible to include an atomic proposition p in the update rule. For example, the formula $[adopt]\beta \leftrightarrow \beta \vee (p \wedge \langle + \rangle\beta)$ only allows an agent to adopt when p is true at that agent and they have a friend that has adopted. This could be relevant when β is a new, expensive technology and p is true at an agent iff 'the income of the agent is higher than ...' or a specific skill is needed before adopting a behavior, like being able-bodied or having access to certain information.

5.3 Axiomatisation

In order to make an axiomatisation of the class of balanced signed threshold models, we first need to define an axiom that capture the property of balance. Balance in a signed threshold frame and model is defined as follows.

Definition 5.5 (Balanced Frame / Model)**.** A frame $F = (A, R^+, R^-)$ is balanced if the signed graph (A, R^+, R^-) is balanced. A model $M = (F, V)$ is balanced if the frame F is balanced.

In $\mathbf{pnl}_n$, balance was an undefinable property. The inclusion of the difference operator changes this. Inspired by a hybrid logic axiom from [29], the following notation is introduced first.

Notation ([39, 29]). Let $i, j \in \mathbb{N}$ and $\varphi \in \mathcal{L}_{[+/-]}$. Let

- $(\boxplus;\boxminus)^{i,j}_{\varphi}$ be the set of all formulas obtained by prefixing φ with a sequence of i positive ($\boxplus$) and j negative ($\boxminus$) box modalities in some order.
- $\Box^k\varphi$ be an element of $(\boxplus;\boxminus)^{i,j}_{\varphi}$ for $i + j = k$.
- $(\langle + \rangle;\langle - \rangle)^{i,j}_{\varphi}$ be the set of all formulas obtained by prefixing φ with a sequence of i positive ($\langle + \rangle$) and j negative ($\langle - \rangle$) diamond modalities in some order.
- $\Diamond^k\varphi$ be an element of $(\langle + \rangle;\langle - \rangle)^{i,j}_{\varphi}$ for $i + j = k$.

An example is the formula $\boxplus\boxplus\boxminus p$, which is an element in the set $(\boxplus;\boxminus)^{2,1}_{p}$. This formula is also one possible instance of $\Box^3\varphi$. Note that $M, x \vDash \Box^k\varphi$ iff for every path of length k, the formula φ is true at the end node. Similarly, $M, x \vDash \Diamond^k\varphi$ iff there exists a path of length k such that φ is true at the end node.

We define the balance axiom as follows:

$$(p \wedge \neg\langle D \rangle p) \to \varphi, \quad \varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p} \qquad (B_{\triangle})$$

This formula is made possible by the sub-formula $p \wedge \neg\langle D \rangle p$. It is satisfied at agent x iff

$$M, x \vDash p \wedge \neg\langle D \rangle p \iff x = V(p).$$

Since p is true only at agent x, this formula assigns p as a 'name' to the agent. This is where the promised expressivity of $\langle D \rangle$ comes from. Axiom $B_{\triangle}$ expresses that if p is true only at agent x, then any odd path will end at an agent where p is not true: any odd path starting in x is not a cycle. If this holds for all agents in a graph, the graph is balanced.

We now define the axiomatisations $L_{[\triangle\theta]}$ and $L_{\triangle\theta}$.

	Signed & Difference axioms	
(K_s)	$\Box(p \to q) \to (\Box p \to \Box q)$	Signed K-Distribution
(K_D)	$[D](p \to q) \to ([D]p \to [D]q)$	Difference K-Distribution
$(Dual_s)$	$\Diamond p \leftrightarrow \neg\Box\neg p$	Signed Duality
$(Dual_D)$	$\langle D\rangle p \leftrightarrow \neg[D]\neg p$	Difference Duality
$(S_{\pm})$	$p \to \Box\Diamond p$	Signed Symmetry
(S_D)	$p \to [D]\langle D\rangle p$	D-Symmetry
$(P4)$	$\langle D\rangle\langle D\rangle p \to (p \vee \langle D\rangle p)$	Pseudo-transitivity
(NO)	$(p \wedge \neg\langle D\rangle p) \to (\boxplus\neg\Diamond^{-} p \wedge \boxminus\neg\Diamond^{+} p)$	Non-overlapping
$(T_{\neg})$	$\Diamond p \to \langle D\rangle p$	Irreflexivity
$(B_{\triangle})$	$(p \wedge \neg\langle D\rangle p) \to \varphi, \quad \varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$	Balance
	Reduction axioms	
(R_p)	$[adopt]p \leftrightarrow p$	Red. Ax. p
$(R_{\neg})$	$[adopt]\neg\varphi \leftrightarrow \neg[adopt]\varphi$	Red. Ax. $\neg$
$(R_{\wedge})$	$[adopt](\varphi \wedge \psi) \leftrightarrow ([adopt]\varphi \wedge [adopt]\psi)$	Red. Ax. $\wedge$
$(R_{\Diamond})$	$[adopt]\Diamond\varphi \leftrightarrow \Diamond[adopt]\varphi$	Red. Ax. $\Diamond$
$(R_{\langle D\rangle})$	$[adopt]\langle D\rangle\varphi \leftrightarrow \langle D\rangle[adopt]\varphi$	Red. Ax. $\langle D\rangle$
(R_{β})	$[adopt]\beta \leftrightarrow (\beta \vee (\Diamond^{+}\beta \wedge (\Diamond^{-}\top \to \Diamond^{-}\neg\beta)))$	Red. Ax. β
	Inference rules	
(**MP**)	If φ and $\varphi \to \psi$, infer ψ	Modus Ponens *
(**Nec**$_s$)	If φ, infer $\Box\varphi$	Signed Necessitation *
(**Nec**$_D$)	If φ, infer $[D]\varphi$	Difference Necessitation *
(**Nec**$_{[adopt]}$)	If φ, infer $[adopt]\varphi$	$[adopt]$ Necessitation
(**RE**)	If φ and $\psi \leftrightarrow \chi$, infer $\varphi[\psi\backslash\chi]$	Replacement of Equivalents
(**US**)	If φ, infer $\varphi(\psi_1/p_1,\ldots,\psi_n/p_n)$	Universal Substitution *
(**IR**)	If $p \wedge \neg\langle D\rangle p \to \varphi$, infer φ where p does not occur in φ	D-rule*

Table 5.2: Axiomatization $L_{[\triangle\theta]}$, where $p,q \in \textsc{At}$, $\Box \in \{\boxplus, \boxminus\}$ and $\Diamond \in \{\Diamond^{+}, \Diamond^{-}\}$.

Definition 5.6 (Axiomatisation $L_{[\triangle\theta]}$)**.** The axiomatic system $L_{[\triangle\theta]}$ is comprised of all instances of propositional tautologies and the axioms and inference rules shown in Table 5.2. The axiomatic system $L_{\triangle\theta}$ is comprised of all instances of propositional tautologies, the Signed & Difference axioms and inference rules marked with a $*$ in Table 5.2.

The axiomatic system $L_{\triangle\theta}$ will also be referred to as *the static part of* $L_{[\triangle\theta]}$.

The logic $L_{[\triangle\theta]}$ is comprised of every formula that can be *generated* using these axioms and inference rules. Note that these formulas do not express validity on a model. For example, p is not in $L_{[\triangle\theta]}$ as there is no way to generate this formulas. On the other hand, $p \vee \neg p$ is in the logic, as it is a propositional tautology. If a formula is generated by the $L_{[\triangle\theta]}$, it is denoted as $\vdash_{L_{[\triangle\theta]}} \varphi$.

Example 5.7. The formula $\langle D\rangle(p \wedge q) \to \neg[D]\neg(p \wedge q)$ is generated from Difference Duality, the propositional tautology $(\chi_1 \leftrightarrow \chi_2) \to (\chi_1 \to \chi_2)$ and the inference rules modus ponens and universal substitution. Thus $\vdash_{L_{[\triangle\theta]}} \langle D\rangle(p \wedge q) \to \neg[D]\neg(p \wedge q)$.

We want to be able express that a formula is true in a logic whenever another set of formulas is true. This is where the notion of *deducibility* comes in.

Definition 5.8 (Deducibility, [5])**.** If $\Gamma \cup \{\varphi\}$ is a set of formulas in language $\mathcal{L}_{[+/-]}$, then φ is $L_{[\triangle\theta]}$-deducible from Γ, denoted as $\Gamma \vdash_{L_{[\triangle\theta]}} \varphi$, if either:

- $\vdash_{L_{[\triangle\theta]}} \varphi$;
- there exists $\psi_1,\ldots,\psi_n \in \Gamma$ such that $\vdash_{L_{[\triangle\theta]}} \psi_1 \wedge \cdots \wedge \psi_n \to \varphi$.

If φ is not $L_{[\triangle\theta]}$-deducible from Γ, it is denoted as $\Gamma \nvdash_{L_{[\triangle\theta]}} \varphi$.

For example, q is not in the logic $L_{[\triangle\theta]}$. However, it is deducible from $\{p, p \to q\}$, since $\vdash_{L_{[\triangle\theta]}} (p \wedge (p \to q)) \to q$ is a propositional tautology, which is in the logic $L_{[\triangle\theta]}$. Therefore, $p, p \to q \vdash q$.

The next two subsections are dedicated to proving soundness and completeness of the axiomatic system $L_{[\triangle\theta]}$ with respect to the class of balanced (signed threshold) models. These two concepts form the bridge between deducibility and validity.

5.4 Soundness

We have already seen that a formula can be satisfied at a specific agent given a specific valuation. But some formulas are satisfied at every agent and every valuation, depending on certain properties of the frame. This is where *validity* comes in.

Definition 5.9 (Validity, [5])**.**

- A formula φ is valid on a frame F, denoted $F \vDash \varphi$, iff $(F,V),x \vDash \varphi$ for all valuations V and all agents x in the frame.
- A formula φ is valid on a class of frames C, denoted $\vDash_C \varphi$, iff $F \vDash \varphi$ for all frames $F \in C$.
- A formula φ is a semantic consequence of set of formulas Γ over class of frames $\mathcal{F}$, denoted $\Gamma \vDash_{\mathcal{F}} \varphi$, iff $\vDash_{\mathcal{F}} \Gamma$ implies $\vDash_{\mathcal{F}} \varphi$.
- A formula φ is valid, denoted $\vDash \varphi$, iff it is valid on the class of all frames.

The soundness of a logic allows any formula generated by a logic to be proven as valid on a class of frames.

Definition 5.10 (Soundness, [39, 5])**.** An logic Λ is sound with respect to a class of frames $\mathcal{F}$, iff $\vdash_\Lambda \varphi$ implies $\vDash_{\mathcal{F}} \varphi$ for any formula φ.

Define $C_{\triangle\theta}$ to be the class of balanced signed threshold frames. In order to prove that $L_{\triangle\theta}$ is sound with respect to $C_{\triangle\theta}$, we show that all the axioms are valid and that the rules of proof preserve validity on the class of signed threshold frames [5]: if the assumptions are valid, the conclusion is valid as well.

The following lemma will be used to prove soundness:

Lemma 5.11 ($(Dual)^k$)**.** *Take any $\varphi \in \mathcal{L}_{[+/-]}$. Then*

$$\vDash \neg\Box^k\neg\varphi \iff \vDash \Diamond^k\varphi$$

where the order of positive and negative modalities between $\Box^k$ and $\Diamond^k$ is preserved.

Proof. Proof by Induction.

Base case Take $k = 1$, let M be an arbitrary model, and let x be an arbitrary agent such that $M,x \vDash \neg\boxplus\neg\varphi$. By definition of $\boxplus$ (Definition 5.1), this is true iff $M,x \vDash \neg\neg\oplus\neg\neg\varphi$. By propositional tautology $\neg\neg\varphi \leftrightarrow \varphi$, this is true iff $M,x \vDash \oplus\varphi$ is true.

Thus, $\vDash \neg\boxplus\neg\varphi \iff \vDash \oplus\varphi$. The proof for $\vDash \neg\boxminus\neg\varphi \iff \vDash \ominus\varphi$ is similar. We conclude that $\vDash \neg\Box\neg\varphi \iff \vDash \Diamond\varphi$ for $\Box \in \{\boxplus, \boxminus\}$ and $\Diamond \in \{\oplus, \ominus\}$.

Induction hypothesis Assume $\vDash \neg\Box^k\neg\varphi \iff \vDash \Diamond^k\varphi$ holds for some integer k, where the order of positive and negative modalities between $\Box^k$ and $\Diamond^k$ is preserved.

Induction step Let M be an arbitrary model, and let x be an arbitrary agent such that $M,x \vDash \neg\boxplus\Box^k\neg\varphi$. This holds iff $M,x \nvDash \boxplus\Box^k\neg\varphi$, which is true iff there exists some $y \in A$ such that xR^+y and $M,y \nvDash \Box^k\neg\varphi$, which is equivalent to $M,y \vDash \neg\Box^k\neg\varphi$. By Induction Hypothesis, this is true iff $M,y \vDash \Diamond^k\varphi$, where the order of positive and negative modalities is preserved. Since there exists y such that xR^+y and $M,y \vDash \Diamond^k\varphi$, it holds that $M,x \vDash \oplus\Diamond^k$. Since M and x were chosen arbitrarily, this result holds for all models.

Thus $\vDash \neg\boxplus\Box^k\neg\varphi \iff \vDash \oplus\Diamond^k\varphi$. The case for $\vDash \neg\boxminus\Box^k\neg\varphi \iff \vDash \ominus\Diamond^k\varphi$ is similar. We conclude that $\vDash \neg\Box^{k+1}\neg\varphi \iff \vDash \Diamond^{k+1}\varphi$, where the order of positive and negative modalities is preserved, holds.

□

Now we can prove the validity of all the signed and difference axioms.

Lemma 5.12. *The signed and difference axioms are all valid on the class of balanced signed threshold frames.*

(K_s) $\Box(p \to q) \to (\Box p \to \Box q)$ *is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash \boxplus(p \to q)$ and $M,x \vDash \boxplus p$. Take any y such that xR^+y. By definition of $\boxplus$, $M,y \vDash p \to q$ and $M,y \vDash p$, why directly implies $M,y \vDash q$. Since y is an arbitrary friend of x, $M,y \vDash q$ holds for all friends y, thus $M,x \vDash \boxplus q$. Assuming $\boxplus(p \to q)$ holds for x, then $M,x \vDash \boxplus p$ implies $M,x \vDash \boxplus q$. Since this holds for arbitrary V and x, the formula $\boxplus(p \to q) \to (\boxplus p \to \boxplus q)$ is valid on the class of signed threshold frames.
A similar proof holds for $\Box = \boxminus$. □

(K_D) $[D](p \to q) \to ([D]p \to [D]q)$ *is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash [D](p \to q)$ and $M,x \vDash [D]p$.

Now take any agent y such that $x \neq y$. Then $M,y \vDash p \to q$ and $M,y \vDash p$ hold at agent y, implying $M,y \vDash q$ as well. Since $M,y \vDash q$ for any y such that $x \neq y$, by definition of $[D]$, this implies that $M,x \vDash [D]q$ holds. Since this holds for arbitrary V and x, the formula $[D](p \to q) \to ([D]p \to [D]q)$ is valid on the class of signed threshold frames. □

$(Dual_s)$ *$\Diamond p \leftrightarrow \neg\Box\neg p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash \neg\boxplus\neg p$. This is true iff there exists some y such that xR^+y and $M,y \nvDash \neg p$, implying that $M,y \vDash p$. Agent x thus has a friend y such that $M,y \vDash p$, which is true iff $M,x \vDash \diamondplus p$. Since this holds for arbitrary V and x, the formula $\diamondplus p \leftrightarrow \neg\boxplus\neg p$ is valid on the class of signed threshold frames.
A similar proof holds for $\Box = \boxminus$, $\Diamond = \diamondminus$. □

$(Dual_D)$ *$\langle D\rangle p \leftrightarrow \neg[D]\neg p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash \neg[D]\neg p$. This holds iff there exists agent y such that $x \neq y$ and $M,y \nvDash \neg p$, implying that $M,y \vDash p$. Thus, there exist agent y such that $x \neq y$ and $M,y \vDash p$, which is true iff $M,x \vDash \langle D\rangle p$. Since this holds for arbitrary V and x, the formula $\langle D\rangle p \leftrightarrow \neg[D]\neg p$ is valid on the class of signed threshold frames. □

$(S_\pm)$ *$p \to \Box\Diamond p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash p$. Take any agent y such that xR^+y. By symmetry of R^+, also yR^+x, meaning that $M,y \vDash \diamondplus p$. Since $M,y \vDash \diamondplus p$ for any y such that xR^+y, we conclude that $M,x \vDash \boxplus\diamondplus p$. Since this holds for arbitrary V and x, the formula $p \to \boxplus\diamondplus p$ is valid on the class of signed threshold frames.
A similar proof holds for $\Box = \boxminus$, $\Diamond = \diamondminus$. □

(S_D) *$p \to [D]\langle D\rangle p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash p$. Take any agent y such that $x \neq y$. Then $M,y \vDash \langle D\rangle p$. Since this holds for any y such that $x \neq y$, this implies that $M,x \vDash [D]\langle D\rangle p$. Since this holds for arbitrary V and x, the formula $p \to [D]\langle D\rangle p$ is valid on the class of signed threshold frames. □

$(P4)$ *$\langle D\rangle\langle D\rangle p \to (p \vee \langle D\rangle p)$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash \langle D\rangle\langle D\rangle p$. By definition of $\langle D\rangle$, there exists agents y,z such that $x \neq y, y \neq z$ and $M,z \vDash p$. The cases $x = z$ and $x \neq z$ are distinguished. In case of the former, this implies that $M,x \vDash p$. In the latter case, there exists an agent $z \neq x$ such that $M,z \vDash p$, which implies that $M,x \vDash \langle D\rangle p$. Since this holds for arbitrary V and x, the formula $\langle D\rangle\langle D\rangle p \to (p \vee \langle D\rangle p)$ is valid on the class of signed threshold frames. □

(NO) *$(p \wedge \neg\langle D\rangle p) \to (\boxplus\neg\diamondminus p \wedge \boxminus\neg\diamondplus p)$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash p \wedge \neg\langle D\rangle p$. Then $M,x \vDash p$ and by semantics of $\langle D\rangle$, there does not exists $y \in A$ such that $x \neq y$ and $M,y \vDash p$. It follows that $V(p) = \{x\}$. Assume there exists $z \in A$ such that $(x,z) \in R^+$. By symmetry and non-overlapping property, this implies $(z,x) \notin R^-$. Then $V(p) = \{x\}$ implies that $(F,V),z \vDash \neg\diamondminus p$. Since this holds for any agent z where xR^+z, it holds for all of x friends. Thus, $(F,V),x \vDash \boxplus\neg\diamondminus p$. A similar proof holds for $M,x \vDash p \wedge \neg\langle D\rangle p$ implies $M,x \vDash \boxminus\neg\diamondplus p$. Since this holds for arbitrary V and x, the formula $(p \wedge \neg\langle D\rangle p) \to (\boxplus\neg\diamondminus p \wedge \boxminus\neg\diamondplus p)$ is valid on the class of signed threshold frames. □

$(T_\neg)$ *$\Diamond p \to \langle D\rangle p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash \diamondplus p$. Then there exists some agent y such that xR^+y and $M,y \vDash p$. Irreflexivity of R^+ implies that $x \neq y$, thus $M,x \vDash \langle D\rangle p$ holds. Since this holds for arbitrary V and x, the formula $\diamondplus p \to \langle D\rangle p$ is valid on the class of signed threshold frames.
A similar proof holds for $\Diamond = \diamondminus$. □

($B_\triangle$) *For any signed threshold frame F, $F \vDash (p \wedge \neg\langle D\rangle p) \to \varphi$ for all $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$ iff F is balanced.*

Proof. ($\Rightarrow$) Proof by contradiction. Let $F = (A, R^+, R^-)$ be an arbitrary signed threshold frame such that $F \vDash (p \wedge \neg\langle D\rangle p) \to \varphi$ for all $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$. Assume that there exists some cycle $x_0 \dots x_k = x_0$ with an odd number of negative edges in F. Let $V(p) = \{x_0\}$, which implies $(F,V), x_0 \vDash p \wedge \neg\langle D\rangle p$. Because agent x_0 is part of a negative cycle, it is implied that $(F,V), x_0 \vDash \Diamond^k p$ holds, where $\Diamond^k p \in (\oplus;\ominus)^{i,2j+1}_{p}$. From Lemma 5.11, it is inferred that $(F,V), x_0 \vDash \neg\Box^k\neg p$ where order of positive and negative modalities is preserved. Define $\varphi := \Box^k \neg p$, then $(F,V), x_0 \vDash \neg\varphi$ for some i, j such that $i + (2j+1) = k$ and some $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$. However, $(F,V), x_0 \vDash p \wedge \neg\langle D\rangle p$ implies $(F,V), x_0 \vDash \varphi$ is valid for all $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$. This is a contradiction, this it is not possible for a negative cycle to exist. The frame F is therefore balanced.

($\Leftarrow$) Proof by contraposition. Assume there exists some valuation V and some agent x such that $(F,V), x \nvDash (p \wedge \neg\langle D\rangle p) \to \varphi$ for all $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$. Then it holds that $(F,V), x \vDash p \wedge \neg\langle D\rangle p$, implying that $V(p) = \{x\}$, and $(F,V), x \vDash \neg\varphi$ for some $\varphi \in (\boxplus;\boxminus)^{i,2j+1}_{\neg p}$. Therefore $(F,V), x \vDash \neg\Box^k\neg p$ holds for $k = i + (2j+1)$. By Lemma 5.11, this implies $(F,V), x \vDash \Diamond^k p$ is true, where the order of positive and negative modalities is preserved. The formula φ was defined to contain an odd number of negative box modalities, by definition, thus $\Diamond^k p$ contains an odd number of negative diamond modalities. By Definition 5.4, $\exists x_k$ such that $xR^{*_1}x_1$, $x_1R^{*_2}x_2$, $\dots$, $x_{k-1}R^{*_k}x_k$ and $(F,V), x_k \vDash p$, where $*_i \in \{+,-\}$ and an odd number of R^{*_i} are R^-. Since $V(p) = \{x\}$, we have that $x_k = x$ and this path is a cycle with an odd number of negative edges. By Definition 2.4 and Definition 5.5, F is not balanced. □

Since the class of balanced signed threshold frames is a subclass of $C_{\triangle\theta}$, every axiom that is valid on the class of signed threshold frames is also valid on the class of balanced signed threshold frames. Thus all signed & difference axioms are valid on the class of balanced frames.

Next, the validity of all the reduction axioms will be proven.

Lemma 5.13 (Validity of Reduction Axioms). *The reduction axioms shown in Table 5.2 are all valid on the class of signed threshold models.*

(R_p) *$[adopt]p \leftrightarrow p$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M, x \vDash p$, and let M' be the model update of M. Since the model does not alter the network structure, $x \in V(p)$ iff $x \in V'(p)$. Thus, $M', x \vDash p$ also holds. By definition of $[adopt]$, $M', x \vDash p$ iff $M, x \vDash [adopt]p$.
Since this holds for arbitrary V and x, the formula $p \leftrightarrow [adopt]p$ is valid on the class of signed threshold frames. □

($R_\neg$) *$[adopt]\neg\varphi \leftrightarrow \neg[adopt]\varphi$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M, x \vDash \neg[adopt]\varphi$ and let M' be the model update of M. Then $M, x \vDash \neg[adopt]\varphi$ iff $M, x \nvDash [adopt]\varphi$, which holds iff $M', x \nvDash \varphi$. By definition of $\neg$, this is true iff $M', x \vDash \neg\varphi$. By definition of $[adopt]$, this is true if and only $M, x \vDash [adopt]\neg\varphi$.
Since this holds for arbitrary V and x, the formula $[adopt]\neg\varphi \leftrightarrow \neg[adopt]\varphi$ is valid on the class of signed diffusion frames. □

($R_\wedge$) *$[adopt](\varphi \wedge \psi) \leftrightarrow ([adopt]\varphi \wedge [adopt]\psi)$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M, x \vDash [adopt](\varphi \wedge \psi)$ and let M' be the model update of M. Then $M, x \vDash [adopt](\varphi \wedge \psi)$ iff $M', x \vDash \varphi \wedge \psi$. By definition of $\wedge$, this is true iff $M', x \vDash \varphi$ and $M', x \vDash \psi$. $M', x \vDash \varphi$ holds iff $M, x \vDash [adopt]\varphi$ holds, while $M', x \vDash \psi$ holds iff $M, x \vDash [adopt]\psi$ holds.
Since this holds for arbitrary V and x, $[adopt](\varphi \wedge \psi) \leftrightarrow ([adopt]\varphi \wedge [adopt]\psi)$ is valid on the class of signed threshold frames. □

($R_\Diamond$) *$[adopt]\Diamond\varphi \leftrightarrow \Diamond[adopt]\varphi$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M, x \vDash [adopt]\oplus\varphi$ and let M' be the model update of M. Then $M, x \vDash [adopt]\oplus\varphi$ iff $M', x \vDash \oplus\varphi$. By definition of $\oplus$, this is true iff there exists agent y such that xR^+y and $M', y \vDash \varphi$. By definition of $[adopt]$, $M', y \vDash \varphi$ iff $M, y \vDash [adopt]\varphi$, and since the model update does not change the network structure, xR^+y is also true for model M. Thus, $M, x \vDash \oplus[adopt]\varphi$.

Since this holds for arbitrary V and x, $[adopt]\oplus\varphi \leftrightarrow \oplus[adopt]\varphi$ is valid on the class of signed threshold frames. A similar proof holds for $\Diamond = \ominus$. □

$(R_{\langle D\rangle})$ *$[adopt]\langle D\rangle\varphi \leftrightarrow \langle D\rangle[adopt]\varphi$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash [adopt]\langle D\rangle\varphi$ and let M' be the model update of M. Then $M,x \vDash [adopt]\langle D\rangle\varphi$ iff $M',x \vDash \langle D\rangle\varphi$. By definition of $\langle D\rangle$, this is true iff there exists y such that $y \neq x$ and $M',y \vDash \varphi$. Then $M',y \vDash \varphi$ iff $M,y \vDash [adopt]\varphi$, and since the model update does not change the set of agents, $y \neq x$ is also true for model M. Thus, $M,x \vDash \langle D\rangle[adopt]\varphi$.
Since this holds for arbitrary V and x, $[adopt]\langle D\rangle\varphi \leftrightarrow \langle D\rangle[adopt]\varphi$ is valid on the class of signed threshold frames. □

(R_β) *$[adopt]\beta \leftrightarrow (\beta \vee (\oplus\beta \wedge (\ominus\top \rightarrow \ominus\neg\beta)))$ is valid on the class of signed threshold frames.*

Proof. Take any signed threshold frame F, valuation V and agent x such that $M,x \vDash [adopt]\beta$, and let M' be the model update of M. Then $M,x \vDash [adopt]\beta$ iff $M',x \vDash \beta$ iff $x \in V'(\beta)$. Define the set $\mathcal{V}$ to be

$$\mathcal{V} := \{x \in A : |N^+(x) \cap B| \geq 1\} \cap \left(\{x \in A : |N^-(x)| = 0\} \cup \{x \in A : |N^-(x) \cap B| < |N^-(x)|\}\right)$$

By Definition 5.3, $x \in V'(\beta)$ iff $x \in V(\beta)$ or $x \in \mathcal{V}$. In the former case, $x \in V(\beta)$ iff $M,x \vDash \beta$. Let us take a look at the latter case by evaluating each of the three sets that $\mathcal{V}$ composes of separately. Firstly, $x \in \{x \in A : |N^+(x) \cap B| \geq 1\}$ iff $N^+(x) \cap B \neq \emptyset$. Thus, there exists agent y such that xR^+y and $y \in V(\beta)$. This is exactly the definition of $\oplus\beta$, thus x is in this set iff $M,x \vDash \oplus\beta$. Secondly, $x \in \{x \in A : |N^-(x)| = 0\}$ iff $N^-(x) = \emptyset$ iff x has no enemies. Finally, $x \in \{x \in A : |N^-(x) \cap B| \leq |N^+(x)|\}$ iff there exists agent z such that $z \in N^-(x)$ and $z \notin B$. By syntactic definition of implication, x is in either the second or the third set iff x not in the second set implies that x is in the third set. Thus, if the union of these two sets holds iff whenever $\exists w$ such that xR^-w, then $\exists z$ such that xR^-z and $z \notin B$. This corresponds to the semantic definition of the formula $\ominus\top \rightarrow \ominus\neg\beta$. To conclude: $x \in \mathcal{V}$ iff $M,x \vDash \oplus\beta$ and $M,x \vDash \ominus\top \rightarrow \ominus\neg\beta$.
Since this holds for arbitrary V and x, $[adopt]\beta \leftrightarrow (\beta \vee (\oplus\beta \wedge (\ominus\top \rightarrow \ominus\neg\beta)))$ is valid on the class of signed threshold models. □

Thus, all reduction axioms are valid on the class of signed threshold models.

The validity of all formulas generated by $L_{[\triangle\theta]}$ can only be insured if all inference rules are validity-preserving.

Lemma 5.14. *The inference rules as stated in Table 5.2 are validity-preserving on the class of signed threshold frames.*

(MP) *If $\vDash \varphi$ and $\vDash \varphi \rightarrow \psi$, infer $\vDash \psi$.*

Proof. Take any $\varphi, \psi \in \mathcal{L}_{[+/-]}$ and assume $\vDash \varphi$ and $\vDash \varphi \rightarrow \psi$. Given arbitrary model M and agent x in the model, the assumption implies that $M,x \vDash \varphi$ and $M,x \vDash \varphi \rightarrow \psi$. By definition of the shorthand $\varphi \rightarrow \psi$, $M,x \vDash \psi$ or $M,x \vDash \neg\varphi$. Since $M,x \vDash \varphi$ holds, $M,x \vDash \neg\varphi$ is false, thus $M,x \vDash \psi$ must be true. Since M and x were chosen arbitrarily, this holds for any model. Thus, Modus Ponens is validity preserving. □

(Nec)$_s$ *If $\vDash \varphi$, infer $\vDash \Box\varphi$.*

Proof. Take any $\varphi \in \mathcal{L}_{[+/-]}$ and assume $\vDash \varphi$. Given arbitrary model M and agent x in the model, the assumption implies that $M,x \vDash \varphi$. Without loss of generality, take any agent y in the model such that xR^*y for $* \in \{+,-\}$. Then $M,y \vDash \varphi$ holds, by validity of φ. Since this holds for any neighbor y of x, we have $M,x \vDash \boxplus\varphi$ and $M,x \vDash \boxminus\varphi$. Since M and x were chosen arbitrarily, this holds for any model. Thus, Signed Necessitation is validity preserving. □

(Nec)$_D$ *If $\vDash \varphi$, infer $\vDash [D]\varphi$.*

Proof. Take any $\varphi \in \mathcal{L}_{[+/-]}$ and assume $\vDash \varphi$. Given arbitrary model M and agent x in the model, the assumption implies that $M,x \vDash \varphi$. Without loss of generality, take any agent $y \neq x$. Validity of φ implies that $M,y \vDash \varphi$ holds. Since this holds for any agent y such that $y \neq x$, we have that $M,x \vDash [D]\varphi$ holds for agent x. Since M and x were chosen arbitrarily, this holds for any model. Thus, Difference Necessitation is validity preserving. □

(Nec)$_{[adopt]}$ *If* $\vDash \varphi$*, infer* $\vDash [adopt]\varphi$.

Proof. Take any $\varphi \in \mathcal{L}_{[+/-]}$ and assume $\vDash \varphi$. Given arbitrary signed threshold model M, let M' be the updated signed threshold model of M. Since M' is also a signed threshold model, $M' \vDash \varphi$ holds, which implies $M \vDash [adopt]\varphi$. Since M and x were chosen arbitrarily, this holds for any model. Thus, $[adopt]$ Necessitation is validity preserving. □

(RE) *If* $\vDash \varphi$ *and* $\vDash \psi \leftrightarrow \chi$*, infer* $\vDash \varphi[\psi \backslash \chi]$.

Proof. Take any $\varphi, \psi, \chi \in \mathcal{L}_{[+/-]}$ such that formula ψ is in φ, and assume that $\vDash \varphi$ and $\vDash \psi \leftrightarrow \chi$. Given arbitrary model M and agent x in the model, the assumption implies that $M, x \vDash \varphi$ and $M, x \vDash \psi \leftrightarrow \chi$. Take $\varphi[\psi/\chi]$ to be the formula φ where every instance of ψ has been replaced with χ. Whenever $M, x \vDash \psi$ holds, also $M, x \vDash \chi$ holds and vice versa. Therefore, replacing ψ with χ in φ does not change the truth value of the formula. This implies that $M, x \vDash \varphi[\psi/\chi]$ holds. Since this holds for arbitrary model M and agent x, this holds for any model. Thus, Replacements of Equivalents is validity preserving. □

(US), [24] *If* $\vDash \varphi$*, infer* $\vDash \varphi(\psi_1/p_1, \ldots, \psi_n/p_n)$.

Proof. Take any $\varphi \in \mathcal{L}_{[+/-]}$ and let the atomic propositions of φ be given by the set $\{p_1, \ldots, p_n\}$. Assume $\vDash \varphi$, and assume $\nvDash \varphi(\psi_1/p_1, \ldots, \psi_n/p_n)$, where ψ_i/p_i is the substitution of proposition $p_i \in \{p_1, \ldots, p_n\}$ with arbitrary formula $\psi_i \in \mathcal{L}_{[+/-]}$. Then there exists a frame F, valuation V and agent x in the model such that $(F,V), x \nvDash \varphi(\psi_1/p_1, \ldots, \psi_n/p_n)$ and $(F,V), x \vDash \varphi$. Now take valuation V' that is defined as following: It is equal to V in all ways, except $(F,V'), y \vDash p_i$ iff $(F,V), y \vDash \psi_i$ for all $y \in A$. Then, for the original agent x and valuation V, where $\varphi(\psi_1/p_1, \ldots, \psi_n/p_n)$ is not satisfied, substituting p_i for ψ_i in the new valuation V' does not change the validity of the statement by definition of V'. We conclude $(F,V'), x \nvDash \varphi$, which is a contradiction with the validity of φ. Thus, Universal Substitution is validity preserving. □

(IR) *If* $\vDash p \wedge \neg\langle D\rangle p \to \varphi$ *where* p *does not occur in* φ*, infer* $\vDash \varphi$.

Proof. Proof by contradiction.
In order to prove the contrapositive, we first need to prove the following: *if a formula* φ *is satisfiable on frame* F*, then* $\varphi \wedge (p \wedge \neg\langle D\rangle p)$*, where* p *is an atomic proposition not in* φ*, is also satisfiable on* F. Given arbitrary signed threshold frame F, assume that there exists some V and x such that $(F,V), x \vDash \varphi$. Now take valuation V', equal to V except that $V(p) = \{x\}$. The satisfaction of φ is still the same with V', since p is not in φ, therefore $(F,V'), x \vDash \varphi$. Additionally, $p \wedge \neg\langle D\rangle p$ is true at x iff $V(p) = \{x\}$, which is true for valuation V'. Thus, $(F,V'), x \vDash \varphi \wedge (p \wedge \neg\langle D\rangle p)$, thus $\varphi \wedge (p \wedge \neg\langle D\rangle p)$ is satisfiable on frame F.

Now, take any $\varphi \in \mathcal{L}_{[+/-]}$ and let the atomic propositions of φ be given by the set $\{p_1, \ldots, p_n\}$. Take any atomic proposition p such that $p \notin \{p_1, \ldots, p_n\}$. Note that p might need to be added to the set of atomic propositions if $\{p_1, \ldots, p_n\} = \mathrm{AT}$.
Assume that $\vDash p \wedge \neg\langle D\rangle p \to \varphi$. Take arbitrary frame F, valuation V and agent x in the model, then it holds that $(F,V), x \vDash p \wedge \neg\langle D\rangle p \to \varphi$ iff $(F,V), x \vDash \neg(p \wedge \neg\langle D\rangle p) \vee \varphi$ iff $(F,V), x \vDash \neg((p \wedge \neg\langle D\rangle p) \wedge \neg\varphi)$ iff $(F,V), x \nvDash (p \wedge \neg\langle D\rangle p) \wedge \neg\varphi$. Since this holds for any F,V and x, it holds for all F,V and x. Thus, there exists no F, V and x such that $(F,V), x \vDash (p \wedge \neg\langle D\rangle p) \wedge \neg\varphi$. Thus, $(p \wedge \neg\langle D\rangle p) \wedge \neg\varphi$ is not satisfiable on any frame. By contrapositive of the above, this means that for all frames, $\neg\varphi$ is not satisfiable. Thus, $(F,V), x \nvDash \neg\varphi \iff (F,V), x \vDash \varphi$ for all F, V and x. Therefore, $\vDash \varphi$. Thus, the D-rule is validity preserving. □

Finally, the soundness of $L_{\triangle\theta}$ is a direct result of Lemma 5.12, Lemma 5.13 and Lemma 5.14.

Theorem 5.15. *The axiomatisation* $L_{\triangle\theta}$ *is sound with respect to the class of balanced signed threshold models* $C_{\triangle\theta}$.

5.5 Completeness

Proving soundness is a straight-forward process of proving validity and validity-preserving. Proving completeness is less straight-forward. Completeness can be either *weak* or *strong*. In this section, only strong completeness is considered.

Definition 5.16 (Strong Completeness, [5])**.** A logic Λ is strongly complete with respect to class of frames $\mathcal{F}$ iff for any set of formulas $\Gamma \cup \{\varphi\}$: $\Gamma \vDash_{\mathcal{F}} \varphi$ implies $\Gamma \vdash_{\Lambda} \varphi$.

This definition says that if Γ semantically entails φ on $\mathcal{F}$, then φ is Λ-deducible from Γ.

In order to prove this, either one of two strategies is used. The first is proof by contraposition: Take any Γ and φ such that $\Gamma \nvdash_{\Lambda} \varphi$ and prove that $\Gamma \nvDash_{\mathcal{F}} \varphi$. The second strategy uses an important lemma that depends on the existence of *consistent* sets.

Definition 5.17 (Consistency, [5])**.** A set of formulas Γ is Λ-inconsistent if $\Gamma \vdash_\Lambda \bot$. It is called Λ-consistent otherwise.

It is not hard to see why any consistent set cannot include both a formula and the negation of that same formula. Take any φ and any Γ such that $\{\varphi, \neg\varphi\} \subseteq \Gamma$. Then $\vdash_\Lambda (\varphi \wedge \neg\varphi) \to \bot$ for any logic that extends propositional logic. A consistent set is then a set that 'does not contradict itself'. We can use this in the completeness lemma.

Lemma 5.18 (Completeness Lemma, [5])**.** *A logic Λ is strongly complete with respect to a class of frames $\mathcal{F}$ iff every Λ-consistent set of formulas is satisfiable on some frame $F \in \mathcal{F}$.*

Using this lemma, the second strategy to proving completeness, entails taking arbitrary Λ-consistent set, finding a frame that satisfies it and proving that this frame is in the desired class of frames. This is the strategy that will be used in this subsection to prove the strong completeness of the static part of $L_{[\triangle\theta]}$. The frame that satisfies this arbitrary set is called an *i-canonical frame*, where every state is a Λ-*nice* set. First, we will define Λ-nice sets. Then we prove that every consistent set can be extended to an Λ-nice set. We can then build the desired *i*-canonical frame. Afterwards, some Lemmas are given that show an *i*-canonical frame is a signed threshold frame, and that show a formula is true at a state iff the formula is in the Λ-nice set that defines the state. Finally, the completeness of $L_{\triangle\theta}$ and $L_{[\triangle\theta]}$ are proven. The main result is Theorem 5.30, which proves the strong completeness of $L_{[\triangle\theta]}$ with respect to balanced signed threshold frames.

This strategy is often used when proving completeness of a modal logic, but there is a key difference in this case. The inclusion of the difference operator, which depends not on a relationship that is seperately defined like R^+ and R^- are. Thus the *i*-canonical frame is slightly different from the more commonly used *canonical frame*. First some concepts from canonical models will be given, which will be connected similar concepts from *i*-canonical models.

Λ-*maximally consistent sets* are considered to be the states of a canonical frame.

Definition 5.19 (Λ-MCS, [5])**.** Let Λ be a logic. A set of formulas Γ is maximally Λ-consistent if Γ is Λ-consistent and for any formula $\varphi \notin \Gamma$, $\Gamma \cup \{\varphi\}$ is Λ-inconsistent. If Γ is a maximal Λ-consistent set of formulas, we say it is a Λ-MCS.

The reason these Λ-MCS's are so useful in building the desired frame, is because they have a number of interesting properties that connect the formulas in a Λ-MCS to their semantic definition. These properties are listed in the following lemma:

Lemma 5.20 (Properties of MCS, [5, 9, 4, 38])**.** *If Λ is a logic, Γ is a Λ-MCS and φ, ψ are formulas, then:*

(i) *Γ is closed under modus ponens: if $\varphi, \varphi \to \psi \in \Gamma$, then $\psi \in \Gamma$;*

(ii) $\Lambda \subseteq \Gamma$

(iii) *for all formulas φ: $\varphi \in \Gamma$ iff $\neg\varphi \notin \Gamma$;*

(iv) *for all formulas φ, ψ: $\varphi \vee \psi \in \Gamma$ iff $\varphi \in \Gamma$ or $\psi \in \Gamma$.*

(v) *for all formulas φ, ψ: $\varphi \wedge \psi \in \Gamma$ iff $\varphi \in \Gamma$ and $\psi \in \Gamma$.*

(vi) *Γ is deductively closed; $\Gamma \vdash_\Lambda \varphi$ implies $\varphi \in \Gamma$.*

Let us highlight the second part of this lemma, as it can be easily overlooked but is quite special. It tells us that any formula generated by Λ is included in every Λ-MCS. There can thus not be any maximally consistent set of propositional logic that does not include the formulas '$p \vee \neg p$', '$p \wedge q \to p$', '$p \wedge q \wedge r \wedge s \to p$', etcetera.

Lemma 5.18 takes any Λ-consistent set, which is not necessarily maximal. The next lemma fixes that issue;

Lemma 5.21 (Lindenbaum, [5, 38])**.** *If Γ is a Λ-consistent set of formulas, then there exists a Λ-MCS Γ' such that $\Gamma \subseteq \Gamma'$.*

Here is where the bridge is crossed between a canonical model and an *i*-canonical model. While the set of agents of a canonical models consists of Λ-MCS, the set of agents of an *i*-canonical model are Λ-nice sets, which are a special kind of Λ-MCS in the context of $L_{\triangle\theta}$.

Definition 5.22 (Λ-nice set, [12])**.** Let Λ be a logic that is an extension of $L_{\triangle\theta}$. A set of formulas Γ is called Λ-nice if Γ is Λ-MCS and $p \wedge \neg\langle D\rangle p \in \Gamma$ for some proposition letter p.

Similar to how every Λ-consistent set is a subset of a Λ-MCS, the following lemma proves that every Λ-consistent set is a subset of a Λ-nice set.

Lemma 5.23 (Nice Lindenbaum, [12])**.** *Let Λ be a logic that is an extension of $L_{\triangle\theta}$. Every Λ-consistent set can be extended to an Λ-nice set, possibly by adding new proposition letters to the language.*

Proof. Take any Λ-consistent set of formulas Γ, and take any proposition letter p such that $\Gamma \cup \{p \wedge \neg\langle D\rangle p\}$ is consistent.
By Lindebaum's lemma, there exists a Λ-MCS Γ' such that $\Gamma \subseteq \Gamma'$. By construction, $p \wedge \neg\langle D\rangle p \in \Sigma'$.
If no such p exists, add a new proposition letter q to be the language. Claim: $\Gamma \cup \{q \wedge \neg\langle D\rangle q\}$ is consistent.
Assume that $\Gamma \cup \{q \wedge \neg\langle D\rangle q\}$ is inconsistent, then there exists $\varphi_1, \ldots, \varphi_n \in \Gamma$ such that $\vdash \varphi_1 \wedge \cdots \wedge \varphi_n \wedge (q \wedge \neg\langle D\rangle q) \to \bot$.
By propositional tautology $((\chi_1 \wedge \chi_2) \to \chi_3) \leftrightarrow (\chi_2 \to (\chi_1 \to \chi_3))$, then also $\vdash (q \wedge \neg\langle D\rangle q) \to ((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \bot)$.
By rule (IR), this means $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \to \bot$, but this is a contradiction with consistency of Γ. Thus, $\Gamma \cup \{q \wedge \neg\langle D\rangle q\}$ is consistent.

By Lindenbaum's Lemma, there is a set Γ^+ such that $\Gamma \cup \{q \wedge \neg\langle D\rangle q\} \subseteq \Gamma^+$ and Γ^+ is a Λ-MCS. Since $q \wedge \neg\langle D\rangle q \in \Gamma^+$, Γ^+ is an $L_{\triangle\theta}$-nice extension of arbitrary $L_{\triangle\theta}$-consistent set Γ. □

Now we can extend any consistent set to a Λ-nice set. Before defining an i-canonical model, the following notation is introduced.

Notation (R-generated sets). Given some relationship $R \subseteq W \times W$ and set $\Gamma, \Gamma' \in W$. The relationship $R^n \subseteq W \times W$ is defined by: $(\Gamma, \Gamma') \in R^n$ iff $\exists \Gamma_1, \Gamma_2, \dots, \Gamma_{n-1}$ such that $(\Gamma, \Gamma_1), (\Gamma_1, \Gamma_2) \dots, (\Gamma_{n-1}, \Gamma') \in R$. If $(\Gamma, \Gamma') \in R^n$, it is denoted by $\Gamma R^n \Gamma'$. The set Γ' is called *R-generated by* Γ if there exists some $n \geq 0$ such that $(\Gamma, \Gamma') \in R^n$.

Now we are ready to define an i-canonical model.

Definition 5.24 (i-Canonical Frame / Model, [12]). An i-canonical model for a logic Λ extending $L_{\triangle\theta}$ in $\mathcal{L}_{+/-}$ is a tuple $M^C = (A^C, R_D^C, R_+^C, R_-^C, V^C)$ where

- $A^C = \{\Gamma \mid \exists n \geq 0 : \Sigma(R_D)^n \Gamma\}$, the set of Λ-nice sets R_D-generated by a single Λ-nice set Σ ;
- $R_D^C = \{(\Gamma, \Sigma) \mid \forall \varphi : \varphi \in \Sigma \implies \langle D\rangle \varphi \in \Gamma\} \subseteq A^C \times A^C$;
- $R_+^C = \{(\Gamma, \Sigma) \mid \forall \varphi : \varphi \in \Sigma \implies ⧫\varphi \in \Gamma\} \subseteq A^C \times A^C$;
- $R_-^C = \{(\Gamma, \Sigma) \mid \forall \varphi : \varphi \in \Sigma \implies \Diamond\varphi \in \Gamma\} \subseteq A^C \times A^C$;
- V^C is the valuation defined by $V^C(p) = \{\Gamma \in W \mid p \in \Gamma\}$, $V^C(\beta) = \{\Gamma \in W \mid \beta \in \Gamma\}$. This is called the *canonical valuation*.

R_D^C, R_+^C and R_-^C are called the *canonical relationships* (of respectively $\langle D\rangle$, ⧫ and ◊). An *i-canonical frame* $F^C = (A^C, R_D^C, R_+^C, R_-^C)$ is an i-canonical model without the valuation V^C.

The relationships R_D^C, R_+^C and R_-^C connect the $L_{\triangle\theta}$-nice sets to each other by their respective modalities, as was seen before in Lemma 5.20. For example, if p is satisfied at agent y in a signed threshold model and xR^+y holds, then ⧫p is satisfied at agent x. Compare this to: If $p \in \Gamma$ and $\Sigma R_+^C \Gamma$, then ⧫$p \in \Sigma$. This connection can be seen alternatively, by the following lemma, which shows the semantic connection of the dual operators.

Lemma 5.25 ([5]). *Take R_D^C, R_+^C, R_-^C as defined in Definition 5.24. These are equivalent to:*

- $R_D^C = \{(\Gamma, \Sigma) \mid \forall \varphi : [D]\varphi \in \Gamma \implies \varphi \in \Sigma\}$
- $R_+^C = \{(\Gamma, \Sigma) \mid \forall \varphi : ⊞\varphi \in \Gamma \implies \varphi \in \Sigma\}$
- $R_-^C = \{(\Gamma, \Sigma) \mid \forall \varphi : ⊟\varphi \in \Gamma \implies \varphi \in \Sigma\}$

For example, $⊟\varphi \in \Gamma$ iff $\varphi \in \Sigma$ in every set Σ such that $\Gamma R_-^C \Sigma$. This connects to the concept of $M, x \vDash ⊟\varphi$ iff $M, y \vDash \varphi$ for every agent y such that xR^-y. The connection between the modal operators and their canonical relationships is re-established. This connection is build into an i-canonical model on purpose, as the goal of building an i-canonical model is to eventually prove that it belongs to the class of signed threshold frames.

There are some issues arising. An i-canonical model is R_D-generated by a single set. This does however not reveal an obvious relationship between the states in A^C that does not involve the original set. Secondly, taking an arbitrary $L_{\triangle\theta}$-nice set does not imply that this set is in A^C, even if there is a canonical relationship between a set in A^C and this arbitrary set. Closure under R_D^C, R_+^C and R_-^C is required. The more glaring issue is that an i-canonical model consists of three relationships while a signed threshold model only includes two. These problems are fixed by the following lemmas.

Lemma 5.26. *Any i-canonical frame $F^C = (A^C, R_D^C, R_+^C, R_-^C)$ for logic $L_{\triangle\theta}$ is closed under R_D^C, R_+^C and R_-^C.*

Proof. Take any $L_{\triangle\theta}$-nice sets Γ, Σ such that $\Gamma \in A^C$ and $\Gamma R_D^C \Sigma$. $\Gamma \in A^C$ implies that $\Pi R_D^C \Gamma$ for some Π that generated A^C. Therefore $\Pi (R_D^C)^2 \Sigma$ implies that $\Sigma \in A^C$, and we conclude that A^C is closed under R_D^C.

Now take any $\Gamma \in A^C$ and $L_{\triangle\theta}$-nice set Σ such that $\Gamma R_+^C \Sigma$ and take any $\varphi \in \Sigma$. The definition of R_+^C tells us that ⧫$\varphi \in \Gamma$. By irreflexivity axiom $\Diamond p \to \langle D\rangle p$ and modus ponens closure (part (i) of Lemma 5.20), this implies that $\langle D\rangle\varphi \in \Gamma$. Therefore, $\Gamma R_D^C \Sigma$. Since A^C is closed under R_D^C, this implies that $\Sigma \in A^C$ and thus A^C is closed under R_+^C. The proof for closure under R_-^C is similar. □

The issue with three relationships in an i-canonical frame is fixed by Lemma 5.27, which tells us that the relationship R_D^C is equal to the inequality relationship, which is inherently included on any frame. This reveals the value of requiring every state of the i-canonical model to include a formula of the form '$p \wedge \neg\langle D\rangle p$'. This formula names a $L_{\triangle\theta}$-set. The relationship R_D^C then requires every neighboring set to not include the atomic proposition p. This allows us then to distinguish between states.

Lemma 5.27. *Let $F^C = (A^C, R_D^C, R_+^C, R_-^C)$ be an i-canonical frame for logic $L_{\triangle\theta}$, and let $\Gamma, \Sigma \in A^C$. Then:*

(i) $\Gamma R_D^C \Sigma \iff \Sigma R_D^C \Gamma$ *(symmetry);*

(ii) $\Gamma R_D^C \Sigma$ *or* $\Gamma = \Sigma$ *(pseudo-transitivity);*

(iii) $\Gamma R_D^C \Sigma \iff \Gamma \neq \Sigma$ *(inequality).*

Proof. (i) Assume $\Gamma R_D^C \Sigma$, and take any $\varphi \in \Gamma$. By symmetry axiom and and modus ponens closure, then $[D]\langle D\rangle\varphi \in \Gamma$, and by R_D^C this implies that $\langle D\rangle\varphi \in \Sigma$. Thus, for any φ: $\varphi \in \Gamma \implies \langle D\rangle\varphi \in \Sigma$. Thus $\Sigma R_D^C \Gamma$.

(ii) Take any two elements $\Gamma, \Sigma \in A^C$. A^C is R_D^C-generated by some $L_{\triangle\theta}$-nice set Π, thus $\Pi R_D^C \Gamma$ and $\Pi R_D^C \Sigma$. By part (i), also $\Gamma R_D^C \Pi$. Now take any $\varphi \in \Sigma$, then $\langle D\rangle\varphi \in \Pi$ and $\langle D\rangle\langle D\rangle\varphi \in \Gamma$. By pseudo-transitivity axiom, this implies $\varphi \vee \langle D\rangle\varphi \in \Gamma$ and Lemma 5.20 part (iv) implies that either $\varphi \in \Gamma$ or $\langle D\rangle\varphi \in \Gamma$.
Thus, $\varphi \in \Sigma \implies \varphi \in \Gamma$ or $\varphi \in \Sigma \implies \langle D\rangle\Gamma$. In the first case, this implies that $\Sigma \subseteq \Gamma$ but since Σ is $L_{\triangle\theta}$-MCS, this implies $\Gamma = \Sigma$. In the latter case, this is the definition of R_D^C, thus $\Gamma R_D^C \Sigma$.

(iii) Assume $\Gamma \neq \Sigma$. By part (ii), this implies that $\Gamma R_D^C \Sigma$. Now assume $\Gamma R_D^C \Sigma$. Both Γ and Σ are $L_{\triangle\theta}$-nice sets, so $p \wedge \neg\langle D\rangle p \in \Gamma$ for some p, thus $p \in \Gamma$ and by duality axiom and modus ponens closure, $[D]\neg p \in \Gamma$. Since $\Gamma R_D^C \Sigma$, Lemma 5.25 implies that $\neg p \in \Sigma$. Part (iii) of Lemma 5.20 implies that $p \notin \Sigma$. $p \in \Gamma$ but $p \notin \Sigma$ implies that $\Gamma \neq \Sigma$.
□

Finally, the connection between the semantic definitions of Definition 5.4 and the canonical relationships is made. This is done by the aptly-named truth lemma.

Lemma 5.28 (Truth Lemma for *i*-canonical models)**.** *For any formula $\varphi \in \mathcal{L}_{+/-}$ and every $L_{\triangle\theta}$-nice set Γ: $M^C, \Gamma \vDash \varphi$ if and only $\varphi \in \Gamma$, where M^C is an i-canonical model such that $\Gamma \in A^C$.*

Proof. By induction on the degree of φ.

Base Case $M^C, \Gamma \vDash p \iff \Gamma \in V^C(p) \iff p \in \Gamma$ by definition of the canonical valuation. The same holds for β.

Induction Hypothesis Assume $M^C, \Gamma \vDash \varphi \iff \varphi \in \Gamma$, and $M^C, \Gamma \vDash \psi \iff \psi \in \Gamma$.

Induction Step The cases $\neg\varphi$, $\varphi \wedge \psi$, $\oplus\varphi$, $\ominus\varphi$ and $\langle D\rangle\varphi$ are distinguished.

- **$\neg\varphi$:** Assume $M^C, \Gamma \vDash \neg\varphi$. This is true iff $M^C, \Gamma \nvDash \varphi$, and by Induction Hypothesis, iff $\varphi \notin \Gamma$. By part (iii) of Lemma 5.20, this is true iff $\neg\varphi \in \Gamma$.

- **$\varphi \wedge \psi$:** Assume $M^C, \Gamma \vDash \varphi \wedge \psi$. This is true iff $M^C, \Gamma \vDash \varphi$ and $M^C, \Gamma \vDash \psi$ iff $\varphi \in \Gamma$ and $\psi \in \Gamma$ by Induction Hypothesis. This holds iff $\varphi \wedge \psi \in \Gamma$, by Lemma 5.20 part (v).

- $\oplus\varphi$ Assume $M^C, \Gamma \vDash \oplus\varphi$. This is true iff there exists $\Sigma \in A^C$ such that $\Gamma R_+^C \Sigma$ and $M^C, \Sigma \vDash \varphi$. The Induction Hypothesis tells us this is true iff $\exists\Sigma$ such that $\Gamma R_+^C \Sigma$ and $\varphi \in \Sigma$.
  ($\Rightarrow$) Assume there exists $\Sigma \in A^C$ such that $\Gamma R_+^C \Sigma$ and $\varphi \in \Sigma$. By definition of R_+^C, this implies that $\oplus\varphi \in \Gamma$.
  ($\Leftarrow$) Assume $\oplus\varphi \in \Gamma$. Take $\Sigma' = \{\varphi\} \cup \{\psi \mid \boxplus\psi \in \Gamma\}$ and assume Σ' is inconsistent. Then $\exists\psi_1, \ldots, \psi_n \in \Sigma'$ such that $\vdash \psi_1 \wedge \cdots \wedge \psi_n \to \neg\varphi$. By Signed Necessitation, $\vdash \boxplus(\psi_1 \wedge \cdots \wedge \psi_n \to \neg\varphi)$ and by Signed K-distribution $\vdash (\boxplus(\psi_1 \wedge \cdots \wedge \psi_n)) \to \boxplus\neg\varphi)$. By propositional tautologies $(\chi_1 \to \chi_2) \to ((\chi_2 \to \chi_3) \to (\chi_1 \to \chi_3))$, and normal modal logic theorem $(\Box\psi_1 \wedge \ldots \Box\psi_n) \to \Box(\psi_1 \wedge \ldots \psi_n)$, this implies $\vdash (\boxplus\psi_1 \wedge \cdots \wedge \boxplus\psi_n) \to \boxplus\neg\varphi$. Since $\boxplus\psi_1, \ldots, \boxplus\psi_n \in \Gamma$ by design of Σ', then $\boxplus\psi_1 \wedge \cdots \wedge \boxplus\psi_n \in \Gamma$. By deductive closure, this means $\boxplus\neg\varphi \in \Gamma$ and by duality $\neg\oplus\varphi \in \Gamma$. This is a contradiction to $\oplus\varphi \in \Gamma$ and part (iii) of Lemma 5.20. Therefore, Σ' is consistent. By Lemma 5.23, there exists a nice set Σ such that $\Sigma' \subseteq \Sigma$. Since $\varphi \in \Sigma'$, also $\varphi \in \Sigma$. Additionally, for all ψ we have that $\boxplus\psi \in \Gamma \implies \psi \in \Sigma$, which implies that $\Gamma R_+^C \Sigma$. Lemma 5.26 tells us then that $\Sigma \in A^C$.
  Thus, there exists $L_{\triangle\theta}$-nice $\Sigma \in A^C$ such that $\Gamma R_+^C \Sigma$ and $\varphi \in \Sigma$.

- $\ominus\varphi$ Similar to the proof for $\oplus\varphi$.

- $\langle D\rangle\varphi$ Similar to the proof for $\oplus\varphi$: $M^C, \Gamma \vDash \langle D\rangle\varphi \iff \exists\Sigma \in A^C$ such that $\Gamma R_D^C \Sigma$ and $\varphi \in \Sigma$. By definition of R_D^C, this implies that $\langle D\rangle\varphi \in \Gamma$. On the other hand, assuming $\langle D\rangle\varphi \in \Gamma$, define $\Sigma' = \{\varphi \wedge \neg\langle D\rangle\varphi\}$. It is consistent, thus Nice Lindenbaum implies that there exists $L_{\triangle\theta}$-nice set Σ such that $\Sigma' \subseteq \Sigma$ and $\varphi \in \Sigma$. Note that $\neg\langle D\rangle\varphi \in \Sigma$, thus $\Sigma \neq \Gamma$ and by Lemma 5.27, this implies that $\Gamma R_D^C \Sigma$.

□

After establishing the truth lemma, we are ready to prove completeness

Theorem 5.29 (Completeness of $L_{\triangle\theta}$)**.** *The logic $L_{\triangle\theta}$ is strongly complete with respect to the class of balanced signed threshold frames.*

Proof. Take $F^C = (A^C, R_D^C, R_+^C, R_-^C)$ to be an *i*-canonical frame of logic $L_{\triangle\theta}$. First, we prove that the *i*-canonical frame satisfies arbitrary consistent set Γ.
Given any $L_{\triangle\theta}$-consistent set Γ, then there exists $L_{\triangle\theta}$-nice set Γ' such that $\Gamma \subseteq \Gamma'$. Consider an *i*-canonical model $M^C = (F^C, V^C)$ such that $\Gamma' \in A^C$, where F^C is the *i*-canonical frame as defined above and V^C is the canonical valuation. For every $\varphi \in \Gamma$, $\varphi \in \Gamma'$. The truth lemma (Lemma 5.28) then implies that $M^C, \Gamma' \vDash \varphi$ for all $\varphi \in \Gamma$. Thus $M^C, \Gamma' \vDash \Gamma$. Thus

there exists a valuation V^C and a state Γ' in the model such that $F^C, V^C \models \Gamma$, meaning Γ is satisfiable on the i-canonical frame.

Now, we need to prove that the i-canonical frame is in the class of signed threshold frames. Note that though a signed threshold frame does not have a relationship R_D, Lemma 5.27 part (iii) proves that the relationship R_D^C is equal to real inequality. This relationship is naturally applied on any frame.

Symmetry Take any $\Gamma, \Sigma \in A^C$ such that $\Gamma R_+^C \Sigma$. Now take any $\varphi \in \Gamma$, then by signed symmetry $\boxplus\Diamond\!\!\!+\varphi \in \Gamma$, and by definition of R_+^C, $\Diamond\!\!\!+\varphi \in \Sigma$. Thus, $\varphi \in \Gamma \implies \Diamond\!\!\!+\varphi \in \Sigma$ implies that $\Sigma R_+^C \Gamma$. Since this holds for any Γ, Σ with $\Gamma R_+^C \Sigma$, the relationship R_+^C is symmetric. A similar proof is given for R_-^C.

Irreflexivity Take any $\Gamma \in A^C$, and assume $\Gamma R_+^C \Gamma$. Then for any $\varphi \in \Gamma$, also $\Diamond\!\!\!+\varphi \in \Gamma$. The irreflexivity axiom and modus ponens closure then says that $\langle D \rangle \varphi \in \Gamma$. Since this holds for any $\varphi \in \Gamma$, this implies that $\Gamma R_D^C \Gamma$. But by Lemma 5.27 part (iii), R_D^C is inequality, so this would imply that $\Gamma \neq \Gamma$. This is a contradiction, thus not $\Gamma R_+^C \Gamma$, and the relationship R_+^C is irreflexive. A similar proof is given for R_-^C.

Non-overlapping Take any $\Gamma, \Sigma \in A^C$ such that $\Gamma R_+^C \Sigma$. Since Γ is an $L_{\triangle\theta}$-nice set, it contains a formula of the form $p \wedge \neg\langle D \rangle p$, thus $p \in \Gamma$. The non-overlapping axiom and Lemma 5.20 part (v) then tells us $\boxplus\neg\Diamond\!\!\!-p \in \Gamma$ and $\boxminus\neg\Diamond\!\!\!+p \in \Gamma$. The formula $\boxplus\neg\Diamond\!\!\!-p \in \Gamma$ and $\Gamma R_+^C \Sigma$ then implies that $\neg\Diamond\!\!\!-p \in \Sigma$. Thus, $\Diamond\!\!\!-p \notin \Sigma$. Since there exists a formula φ such that $\varphi \in \Gamma$, but not $\Diamond\!\!\!-\varphi \in \Sigma$, this proves that not $\Sigma R_-^C \Gamma$. Symmetry of R_-^C then tells us that $(\Gamma, \Sigma) \notin R_-^C$. A similar proof holds for $(\Gamma, \Sigma) \in R_-^C \implies (\Gamma, \Sigma) \notin R_+^C$, thus $\Gamma R_+^C \Sigma \iff \Gamma R_-^C \Sigma$, the relationships R_+^C and R_-^C are non-overlapping.

Balanced Take any $\Gamma_0 \in A^C$. Since Γ_0 is an $L_{\triangle\theta}$-nice set, it contains a formula of the form $p \wedge \neg\langle D \rangle p$ for some propositional atom p. By the balance axiom, this implies that $\varphi \in \Gamma_0$ for all $\varphi \in (\boxplus;\boxminus)_{\neg p}^{i,2j+1}$. Now take any $\Gamma_1, \ldots, \Gamma_n \in A^C$ such that $\Gamma_0 R_{*_1}^C \Gamma_1, \Gamma_1 R_{*_2}^C \Gamma_2, \ldots, \Gamma_{n-1} R_{*_n}^C \Gamma_n$ for $*_i \in \{+,-\}$, where an odd number of $R_{*_i}^C$ are R_-^C. In other words, take any path with an odd number of negative edges starting in Γ_0. By definition of $(\boxplus;\boxminus)_{\neg p}^{i,2j+1}$ and Lemma 5.25, this implies that $\neg p \in \Gamma_n$. But $p \wedge \neg\langle D \rangle p \in \Gamma_0$ implies that $\neg p \notin \Gamma_0$. Thus, $\Gamma_0 \neq \Gamma_n$, and the path is not a cycle.
Thus any path with an odd number of negative edges starting from Γ_0 is not a cycle, from which we conclude that Γ_0 is not in a negative cycle. Since this holds for arbitrary $\Gamma \in A^C$, this implies there are no negative cycles in the model, which means that the i-canonical model is balanced.

Thus, an i-canonical frame is a tuple (A^C, R_+^C, R_-^C) where the relationships R_+^C and R_-^C are symmetric, irreflexive and non-overlapping. It also also shown that the i-canonical frame is balanced. Therefore, an i-canonical frame is in the class of balanced signed threshold frames. Therefore, $L_{\triangle\theta}$ is strongly complete with respect to $C_{\triangle\theta}$.

□

Theorem 5.30 (Completeness of $L_{[\triangle\theta]}$). *The logic $L_{[\triangle\theta]}$ is strongly complete with respect to the class of balanced signed threshold frames.*

Proof. Every formula in $\mathcal{L}_{[+/-]}$ can be translated to a formula in $\mathcal{L}_{+/-}$ using the reduction axioms (Table 5.2). The soundness of these reduction axioms (Lemma 5.13) proves that this translation preserves the validity, and therefore the meaning, of a formula. By Theorem 5.29 and this translation, the logic $L_{[\triangle\theta]}$ is complete with respect to the class of balanced signed threshold frames. □

6 Conclusion

This section concludes this thesis. Section 6.1 gives a summary of the most important results of Section 4 and Section 5. Section 6.2 discusses the limitations of the research and possible solutions. Finally, Section 6.3 details possible ventures to expand on signed threshold graphs and $L_{[\triangle\theta]}$.

6.1 Summary

In Section 1, the following research questions have been posed: *Is it possible to combine a signed graph and a threshold model, to obtain a model of friends and enemies that influence each others decision to adopt a behavior?* and *Can the logics* $\boldsymbol{pnl}_n$ *and* L_θ *be combined to form a meaningful logic?*. These questions were answered in Section 4 and Section 5.

A signed threshold model is an irreflexive threshold model where a distinction is made between positive and negative relationships between agents. Agents play a coordination game with their friends, where they are rewarded for making the same choice as a friend. Simultaneously, agents play an anti-coordination game with their enemies, where they are punished for making the same choice as an enemy. The choice of whether to adopt or not is then a result of playing both these games with several friends and enemies at once.

Several results follow. Firstly, an agent can get the highest possible payoff when all friends have made the same choice as them and all their enemies have made the opposite choice. This leads to the second result: If a signed threshold graph is balanced, every agent achieves maximal payoff iff the set of adopted agents is equal to one of the two friend groups that make up a balanced graph. This results actually characterizes signed balanced graphs, as the third results states: a signed threshold graph is balanced iff there exists a configuration of the set of adopted agents where every agent has maximal payoff.

A simplification of the coordination and anti-coordination games are given: an agent gets positive payoff for coordination with a friend on a choice, a receives a negative payoff for coordinating on that same choice with an enemy. The agent receives no payoff when making any other choice. This allows for a concrete update rule to depict which agents adopt. Depending on whether an agent has more friends or enemies, and whether an agent has more adopted friends or adopted enemies, this leads to several different scenarios where the agent does not always take the threshold into account when deciding.

A new phenomenon is that an agent can adopt a behavior even when no neighbors have adopted it. Clusters and cascades as defined in threshold models are thus not applicable. Instead, a cluster of friends is defined, which blocks adoption of a behavior. This still allows for some condition on when the optimal configuration in a balanced graph can be achieved.

The logic $L_{[\triangle\theta]}$, is defined. Its language includes an atomic proposition for 'has adopted', positive and negative diamond modalities, a dynamic adopt modality and the difference operator. The semantics are defined on a signed threshold model, composed of a signed threshold frame and a valuation function. An agent updates whenever one of their friends has adopted and not all of their enemies have adopted, if they have any enemies at all. Other possible update rules are discussed.

After defining the semantics of $L_{[\triangle\theta]}$, an axiomatisation is defined that is split up into three parts: signed & difference axioms, reduction axioms and inference rules. Soundness is proven through validity of all axioms and validity-preserving of all inference rules. The reduction axioms allow the language to be translated to a static language that does not include the $[adopt]$ modality. The signed & difference axioms include axioms for the previously undefinable properties of irreflexivity, balance and non-overlapping edges. The inclusion of K-distribution, duality, and necessitation for both the diamond operators and the difference operator, allows us to prove completeness with respect to balanced signed threshold frames of the static part of the axiomatisation. Using the reduction axioms, it is proven that $L_{[\triangle\theta]}$ is complete with respect to the class of balanced signed threshold frames.

6.2 Discussion

Naturally, every research has its limitations. Signed threshold models and $L_{[\triangle\theta]}$ are no exemptions in this regard.

The limitations of signed graphs naturally apply to signed threshold graphs as well. One of these limitations is that balance is quite a fragile concept. A balanced network of a million nodes would become unbalanced by just one edge changing from positive to negative or vice versa. Instead of requiring balance, one could require balance up to a certain degree (Section 2.2).

Another limitation inherented from signed graphs is the discrepancy between mathematical theory and psychological reality. In practice, relationships between people take on many varying forms. Categorizing these as either positive or negative simplifies the relationships too much, some would say [2]. Neutral or indifferent relationships exist. Furthermore, the intensity of a friendship or enmity is not expressed in a signed graph.

But trying to add this notion of relationship intensity to a signed graph results in new problems. As Flament states: "One might think that a valued algebraic graph is necessary to represent psycho-social reality, if it is to take into account the degree of intensity of interpersonal relationships. But in fact it then seems hardly possible to define the balance of a graph, not for mathematical but for psychological reasons. If the relationship AB is +3, the relationship BC is –4, what should the AC relationship be in order that the triangle be balanced?" [14]. This is an open question, within psychology and mathematics.

If we let go of the desire for balance in a signed graph, this quantification of relationship intensity is useful with analysing behavior adoption. The more intense a friendship is, the more likely you are to be influenced by them, which could be represented by a weight in the calculation of the update rule. In this situation, one close friend adopting the behavior would have the same influence on an agent as multiple distant colleagues adopting the behavior. This would correspond more closely to real-life social influence, and could be useful when analysing a situation where a person is influenced by only one or two people in their life.

Another situation that this model cannot depict, is that of one-sided influence. Though the assumption of non-overlapping relationships is natural, and the assumption of irreflexivity would only make sure agents do not have a hand in influencing themselves, there are benefits to not requiring symmetry in a signed threshold graph. Previously, it was dismissed because a signed graph that was not symmetric would not be balanced. If we focus not on balance, but rather on the flow of a behavior through a network, then asymmetric relationship are very relevant. Every person is constantly receiving influence from sources that they do not have a similarly big influence on. Companies exert their influence through advertisements, while celebrities and aptly-named ‘influencers’ mere existence influences people to buy products [26]. Taking it a step further, governments influence their people, through policy changes, regulation of products and propaganda. This impacts the products people use, what people do and even what people think. This influence is not double-edged: though democratic societies allow the people of a country to vote, they have no direct impact on policy. In all of these examples, the companies, celebrities and government do not have any direct relationship with a majority of the people that they influence. This is a situation where there is no symmetric relationship, but there is a behavior being adopted throughout the network. Especially in the case of politics, negative relationships would be essential to include in the model.

A signed threshold graph also assumes that once an agent adopts, they cannot rethink that choice and unadopt the behavior. This assumption was inherited directly from threshold models, where there is no need to unadopt, as the assumption reinforces itself: an agent adopts if their threshold is reached and this threshold cannot be ‘un-reached’, since their neighbors also cannot unadopt. Thus, there is no need to reconsider. As an added benefit, it prevents the initial adopters from unadopting and breaking the cascade before it has even started. However, in the context of signed threshold graphs, it makes perfect sense for an agent to unadopt sometimes. For example, if an agent adopts, and then some of their enemies adopt afterwards, they might regret their choice and want to unadopt. Adopting the behavior then does not lead to the highest possible payoff anymore.

The final assumption inherited from threshold models, is that of a uniform threshold. If the threshold instead is a function from the agents to the interval $[0,1]$, every agent would have their own threshold. While in the context of a computer program that can easily calculate the results in such a threshold model, it becomes harder to conduct a formal analysis. My strong suspicion is that there would be too much variables in the network to draw any meaningful conclusion.

That brings us to the limitations of $L_{[\triangle\theta]}$, where the biggest limitation is that there is no threshold. It is therefore not possible for this logic to describe signed threshold graphs. While the update rule is still of the form ‘an agent adopts depending on whether enough friends and enemies have adopted’, the threshold is not a fraction like threshold models and signed threshold graphs have. The original research goal was to find a logic that would describe a signed threshold graph, and this goal was not achieved. The update rules from threshold models and signed threshold cannot be expressed in this logic, as the global saisfaction from L_θ is missing.

When venturing away from modal logic into hybrid logic, progress has already been made on logic of balance. In [29, 28], $L_{\mathbf{PNL}}$ has been extended with nominals and hybrid operators. This logic axiomatizes non-overlapping and balance, and is proven to be sound and complete with respect to the class of balanced signed hybrid frames. In light of this hybrid extension, a complete logic that axiomatizes balance is not new. Some of the results of this thesis are thus not entirely novel.

Overall, the focus on this thesis was on the property of balance. When starting the research, the conjecture was posed that balanced graphs and maximal payoff were connected in a sense. This speculation eventually resulted in Section 4.2. Reflecting on this thesis, I think the focus on balance has limited the research more than I had predicted.

As stated in this subsection, if we forego the desire to define balance and analyse balanced graphs, the possibility for a graph to be non-symmetric and have weighted edges opens up. In light of behavior diffusion, these are non-trivial limitations. At the same time, $L_{[\triangle\theta]}$ was constructed in such a way that balance could be defined, with the downside that the logic does not describe a signed threshold graph. The next subsection gives some details on another language that would allow for a logic to describe signed threshold graphs. It was decided not to choose this language, as the belief that balance could not be defined was quickly established. This is another unfortunate instance where the focus on balance has hindered this thesis in exploring the research questions.

6.3 Further research

First, I will discuss some solutions to limitations posed in the previous subsection. Afterwards, some possible extensions of signed threshold graphs and $L_{[\triangle\theta]}$ are listed.

There are two obvious approaches to incorporating relationship intensity to signed threshold graphs: weighted edges and strong and weak ties. For weighted edges, edge xy between agents x and y would receive weight w_{xy}. The payoff agent x would receive from agent y would be $w_{xy}*$‘payoff from agent y’, where the payoff from agent y is determined by the (anti-)coordination game matrix. The closer agents x and y are, the more w_{xy} would increase, allowing close friends and family members to have a bigger influence on an agent. The second approach is that of strong and weak ties as explained by

Granovetter [17]. This influential paper mentions both diffusion processes, balance and clustering [11]. Adding strong and weak ties to signed threshold graphs is therefore a natural idea.

Other ideas for further research would be to incorporate assymetric relationships to signed threshold graphs, incorporating a non-uniform threshold and to research agents being able to unadopt. When considering agents unadopting, the following conjecture is posed:

Conjecture. Given a signed threshold graph with model update rule such that agents can unadopt, the diffusion sequence will eventually reach a state where either no changes occur or the sequence oscillates between two models. Most diffusion sequences of signed threshold graphs will end in oscillation.

The final proposal for further research on signed threshold graphs is on the network structure. Instead of considering behavior, take the set B to represent a political opinion or religion. Differences in opinion on these matters can change relationships, even between close family members. Incorporating some element of the network structure changing to accommodate this would be an interesting area of research. For example, take a signed threshold graph that is not balanced, therefore not every agent has optimal payoff. Now take one positive edge xy such that agents x has adopted but agent y has not. This can be a motivation to change the positive relationship to a negative one, possibly making the graph balanced. This would be even more effective if different relationship strengths are incorporated as well.

When discussing $L_{[\triangle\theta]}$, a couple of questions come to mind for further research. The first is that of decidability: is there an algorithm which is capable of deciding, given an arbitrary formula, whether it belongs to the logic or not? [8] When considering why $L_{[\triangle\theta]}$ might not be decidable, the balance axiom is most likely the cause, as this axiom captures an infinite number of cycles. However, $(\boxplus;\boxminus)^{n,m}_{\neg p}$ is 'recursively enumerable, comparable to the countable set of propositional tautologies' [28]. This notion of recursively enumerability is closely connected to algorithms and thus decidability. The interested reader is referred to [8] and [5] for more information on decidability. For now, decidability of $L_{[\triangle\theta]}$ is an open question.

It would not be very complicated to extend $L_{[\triangle\theta]}$ with axioms that define completely connectedness of a network and weak balance. See, for instance, [29]. Another possibility to extend $L_{[\triangle\theta]}$ is to include epistemic operators, inspired by [4]. After defining threshold models and its update rule, the paper goes into the knowledge that agents have over their network. In fact, there are some underlying assumptions on knowledge already present when discussing threshold models and signed threshold models. Agents in a threshold model only have knowledge of their direct neighbors. Agents in a signed threshold graph are able to adopt without any of their neighbors adopting, thus there is a common knowledge within the network of the existence of this behavior, an agent does not need to *learn* of the behavior first. Including an epistemic operator into the language $\mathcal{L}_{[+/-]}$ and allowing agents to have greater knowledge over the network then only their direct neighbors. It would also allow agents to predict the behavior adoption of their neighbors, therefore the agents adopt at an earlier point in the diffusion sequence.

Another extension of $L_{[\triangle\theta]}$ is to include multiple update rules. Construct n update rules as defined in Section 5.2. Every agent would then adhere to one of these update rules. This divides the agents into *n types*. The syntax of the language is then extended to include $[adopt]_1$, $[adopt]_2$, ..., $[adopt]_n$. The reduction axioms for all these different adoption rules would be the same as in Table 5.2, except for the β reduction axiom, which would be replaced by n reduction axiomas. Semantically, $[adopt]_i\varphi$ is satisfied at agent x if agent x is of type i and φ is true in the model update of a signed threshold model. The signed threshold model would include a function $t : A \to \{1,\dots,n\}$, assiging every agent to a type.

Finally, there is an alternative way to fuse $L_{\mathbf{PNL}}$ and L_θ, by replacing N_{xy} with the atomic propositions N^+_{xy} and N^-_{xy} to distinguish between positive and negative relationships between agents. The proposed language would then be

$$\varphi = \beta_x \mid N^+_{xy} \mid N^-_{xy} \mid \neg\varphi \mid \varphi \wedge \varphi \mid [adopt]\varphi$$

where x and y are agents in the network (Section 3.4). Call this language $\mathcal{L}_{[\pm]}$. A proposal for the reduction axiom of β would look like

$$[adopt]\beta_x \leftrightarrow \beta_x \vee \beta_{N^\pm(x)\geq\theta}$$

where

$$\beta_{N^\pm(x)\geq\theta} = \bigvee_{\left\{\substack{G^+\subseteq N^+(x),G^-\subseteq N^-(x):\\ |G^+|-|G^-|\geq\theta(|N^+(x)|-|N^-(x)|)}\right\}} \left(\bigwedge_{y\in G^+\cup G^-}\beta_y \vee \bigwedge_{y\notin G^+\cup G^-}\neg\beta_y\right)$$

This logic would then be a logic that expresses the update rule, and the model on which the semantics is defined would be a signed threshold graph. Because β_x, N^+_{xy} and N^-_{xy} are atomic propositions that can be globally evaluated, it is possible to talk about a specific agent and thus to count the number of agents. The network axioms of non-overlapping and irreflexivity would be respectively expressed by the formulas $N^+_{xy} \leftrightarrow N^-_{xy}$ and $\neg N^+_{xx} \wedge \neg N^-_{xx}$. However, I pose the following conjecture:

Conjecture. The property of balance is not definable in a logic with the language $\mathcal{L}_{[\pm]}$.

When constructing $L_{[\theta\triangle]}$, there was a strong desire to implement balance into the logic. This was the main reason to include positive and negative modalities instead of positive and negative propositional network atoms. Nevertheless, this other language $\mathcal{L}_{[\pm]}$ and a possible axiomatisation would be interesting areas for further research.

References

[1] Robert P Abelson and Milton J Rosenberg. Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science*, 3(1):1–13, 1958.

[2] Bo Anderson. Cognitive balance theory and social network analysis: Remarks on some fundamental theoretical matters. In *Perspectives on social network research*, pages 453–469. Elsevier, 1979.

[3] Yacine Atif, Kanna Al-Falahi, Tshering Wangchuk, and Birgitta Lindström. A fuzzy logic approach to influence maximization in social networks. *Journal of Ambient Intelligence and Humanized Computing*, 11(6):2435–2451, 2020.

[4] Alexandru Baltag, Zoé Christoff, Rasmus K Rendsvig, and Sonja Smets. Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, 107:489–531, 2019.

[5] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*, volume 53. Cambridge University Press, 2001.

[6] Elizabeth Blair. When celebrities show up to protest, the media follows — but so does the backlash. *National Public Radio (NPR)*. URL https://www.npr.org/2024/03/01/1232497188/israel-hamas-celebrity-activism. referenced at 13-03-2025.

[7] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider's theory. *Psychological review*, 63 (5):277–293, 1956.

[8] Alexander Chagrov and Michael Zakharyaschev. *Modal logic*. Oxford University Press, 1997.

[9] Chen Chung Chang and H Jerome Keisler. *Model theory*, volume 73. Elsevier, 1990.

[10] Zoé Christoff. *Dynamic logics of networks: information flow and the spread of opinion*. PhD thesis, University of Amsterdam, 2016.

[11] James A Davis. Clustering and structural balance in graphs. *Human relations*, 20(2):181–187, 1967.

[12] Maarten De Rijke. The modal logic of inequality. *The Journal of Symbolic Logic*, 57(2):566–584, 1992.

[13] David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a Highly Connected World*, volume 1. Cambridge University Press, 2010.

[14] Claude Flament, Maurice Pinard, Raymond Breton, and Fernand Fontaine. *Applications of graph theory to group structure*. Prentice-Hall, 1963.

[15] Craig Garthwaite and Timothy J Moore. Can celebrity endorsements affect political outcomes? evidence from the 2008 us democratic presidential primary. *The journal of law, economics, & organization*, 29(2):355–384, 2013.

[16] Robert H. Gass. Social influence, sociology of. In James D. Wright, editor, *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pages 348–354. Elsevier, Oxford, second edition, 2015. ISBN 978-0-08-097087-5.

[17] Mark Granovetter. The strength of weak ties. *American journal of sociology*, 78(6):1360–1380, 1973.

[18] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.

[19] Frank Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953.

[20] Frank Harary. On local balance and n-balance in signed graphs. *Michigan Mathematical Journal*, 3(1):37–41, 1955.

[21] Frank Harary. On the measurement of structural balance. *Behavioral Science*, 4(4):316–323, 1959.

[22] Rainer Hegselmann. Thomas C. Schelling and James M. Sakoda: The intellectual, technical, and social history of a model. *Journal of Artificial Societies and Social Simulation*, 20(3), 2017.

[23] Fritz Heider. Attitudes and cognitive organization. *The Journal of psychology*, 21(1):107–112, 1946.

[24] GE Hughes. *A New Introduction to Modal Logic*. Routledge, 1996.

[25] Fenrong Liu, Jeremy Seligman, and Patrick Girard. Logical dynamics of belief change in the community. *Synthese*, 191: 2403–2431, 2014.

[26] Paris Martineau. The wired guide to influencers. *WIRED*. URL https://www.wired.com/story/what-is-an-influencer/. accessed at 23-03-2025.

[27] Michael Moore. An international application of heider's balance theory. *European Journal of Social Psychology*, 8(3), 1978.

[28] Mina Young Pedersen. *Malicious Agents and the Power of the Few: On the Logic of Abnormality in Social Networks*. PhD thesis, The University of Bergen, 2024.

[29] Mina Young Pedersen, Sonja Smets, and Thomas Ågotnes. Modal logics and group polarization. *Journal of Logic and Computation*, 31(8):2240–2269, 2021.

[30] Graham Priest. *An introduction to non-classical logic: From if to is*. Cambridge University Press, 2008.

[31] Ji Ruan and Michael Thielscher. A logic for knowledge flow in social networks. In *AI 2011: Advances in Artificial Intelligence: 24th Australasian Joint Conference, Perth, Australia, December 5-8, 2011. Proceedings 24*, pages 511–520. Springer, 2011.

[32] James M Sakoda. The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, 1(1):119–132, 1971.

[33] Thomas C Schelling. Models of segregation. *The American economic review*, 59(2):488–493, 1969.

[34] Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[35] Jeremy Seligman, Fenrong Liu, and Patrick Girard. Logic in the community. In *Indian conference on logic and its applications*, pages 178–188. Springer, 2011.

[36] Kat Tenbarge. Social media fuels boycotts against mcdonald's and starbucks over israel-hamas war. *NBC News*. URL https://www.nbcnews.com/tech/social-media/social-media-fuels-boycotts-mcdonalds-starbucks-israel-hamas-war-rcna125121. referenced at 12-03-2025.

[37] Wiebe Van der Hoek, Louwe Kuijer, and Yì Wáng. Logics of allies and enemies: A formal approach to the dynamics of social balance theory. In *Proc. of IJCAI 2020*, volume 2021, pages 210–216, 2020.

[38] Hans Van Ditmarsch, Wiebe van Der Hoek, and Barteld Kooi. *Dynamic epistemic logic*, volume 337. Springer Science & Business Media, 2007.

[39] Zuojun Xiong and Thomas Ågotnes. On the logic of balance in social networks. *Journal of Logic, Language and Information*, 29(1):53–75, 2020.

[40] Thomas Zaslavsky. A mathematical bibliography of signed and gain graphs and allied areas. *The Electronic Journal of Combinatorics*, pages DS8–Dec, 2012.

[41] Liang Zhen and Jeremy Seligman. A logical model of the dynamics of peer pressure. *Electronic Notes in Theoretical Computer Science*, 278:275–288, 2011.