Quantitative Models for Social Networks

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

José C. Oliveira

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Sophia Knight

August 2024

Acknowledgements

I am deeply grateful to my family for their unwavering support and encouragement throughout this journey. I am especially indebted to my supervisor, Dr. Sophia Knight, for her guidance, mentorship, friendship and for being optmistic when I was so pessimistic sometimes.

I would like to thank the Department of Computer Science, Swenson College of Science and Engineering, University of Minnesota Duluth, for providing a stimulating academic environment and essential resources. I am also grateful for the financial support provided by the University of Minnesota Duluth, through research and teaching assistantships.

I extend my sincere thanks my advisor from college, Dr. Mário S. Alvim, for his guidance, mentorship and friendship. I would also like to thank David Sullivan, Bernardo Amorim, Artur Gaspar, and Kaio Rosa for their insightful discussions and feedback.

I would like to express my sincere gratitude to my committee members, Dr. Sophia Knight, Dr. Andrew Sutton, and Dr. Alexis Elder, for their time, expertise, and valuable feedback.

I would like to thank my friends, David Sullivan and family, Jaymes "Seven" Andrews and family, Brody Wollner, Lucas Fraga, Ramon Gonzé, Erik Francisco, and João Lara, whose love and support have helped me navigate both the challenges and triumphs of this journey.

Finally, I am incredibly grateful for the support and guidance I have received throughout this journey.

Dedication

In memory of Pedro Rodrigues, whose friendship inspired me to persevere.

Abstract

Recently, social networks have begun to influence every aspect of our lives, with unprecedented, unanticipated consequences, especially in politics and public opinion. Research on social networks has studied opinions and their change over time, but to accurately model real people and their opinions and beliefs, we must include representations of *uncertainty* and *inconsistency* in formal models of social networks. This work presents models for social networks that includes representation of uncertainty, inconsistencies and dynamic trust between agents.

We developed a model for social networks using subjective logic, a probabilistic multi-agent logic with operations like trust-discount and belief fusion. We devise a desiderata of properties for an opinoin update function and explore update function with these operations. We found that no function with these operators satisfies our desiderata, but a function using cumulative belief fusion can represent interactions that can lead to polarization.

We expand the Alvim-Knight-Valencia model for social networks with dynamic influence allowing belief update function to change the influence over time. Using results from previous works in the AKV model, we define balance-conservative graphs, graphs where agents influence and are influenced by the same degree for every timestep. We conjecture that if there are two agents in a balanced graph that have different total influence, then the graph is not balanced-conservative.

Finally, we develop two Python tools for modeling and simulation with the AKV model and the Paraconsistent Gödel Modal Logic, a logic to reason about uncertainty and inconsistency. Using the AKV tool, we reproduced the results from previous works in the model. These tools are available publicly and under the MIT license, allowing researchers to freely use it and contribute.

Contents

List of Figures

1 Introduction

In Brazil, after the 2022 presidential elections, on January 8th, rioters gathered at the Three Powers Plaza, Brasilia, to attack the results of the election, to call for a military dictatorship, and to shut down the National Congress and the Supreme Federal Court, and to reinstate the previous president (Weterman 2023). Days before the attack, rioters were planning the attack on social networks (G1 2023). A research on Telegram chat groups identified individuals, and here we call them *influencers*, that played a central role in the distribution of content that led to the January 8th riots (Venâncio et al. 2024).

Figure 1.1: Rioters can be seen using their cellphones during the invasion at the National Congress of Brazil. Picture by Hugo Barreto from Rodrigues et al. (2023).

Given how social networks have begun to influence every aspect of our lives

especially in politics and public opinion, we need to understand how users get in consensus or polarize, as well develop guidelines to avoid polarization. Research on social networks has studied opinions and their change over time, but to accurately model real people and their opinions and beliefs, we must include representations of *uncertainty* and *inconsistencies* in formal models of social networks.

This work presents models for social networks with uncertainty, inconsistencies and dynamic trust between agents. This work is structured as follows:

- Chapter 2 – Background: In this work, we will use subjective logic, the Alvim-Knight-Valencia model, and paraconsistent Gödel modal logic. This chapter introduces key conceptes for each of these topics necessary for this work.

- Chapter 3 – Related Work: We review works on formal models for social networks including quantitative and logical approaches.

- Chapter 4 – Opinion Update in a Subjective Logic Model for Social Network: We develop a model for social networks with subjective logic, a probabilistic multi-agent logic that includes information about agent's beliefs and their uncertainty about a proposition. This logic also includes multinomial opinions, allowing to model topics that cannot be answered with only "yes" or "no".

  Here we devise a desiderata of properties for an opinion update function and explore update functions using the operators from subjective logic. We found that no update function with trust-discount and belief fusion operators satisfies our desiderata.

  On the other hand, we show that our update function with cumulative belief fusion has the potential to represent interaction that can lead to polarization when two agents on the same side repeatedly interact or when one agent is

transmitting more information than the other.

A preliminary research on subjective logic models for social networks was published by Oliveira (2022).

- Chapter 5 – Alvim-Knight Valencia model with Dynamic Influence: Given the mutability of relations in social network, we extended the AKV model, developed by Alvim, Knight, and Valencia (2019), that allows the update of influence graphs over time.

  Using results from Alvim, Amorim, et al. (2023), we define balance-conservative graphs, graphs that, for every timestep, every agent influences and is influenced by the same degree. We show examples of balance and non-balance-conservative graphs and conjecture on properties that a graph must have to be non-balanced-conservative. If these conjectures are proven to be true, then we have procedures to check if an influence graph can lead to polarization.

- Chapter 6 – A Python Tool for Social Network Simulations in the Alvim-Knight-Valencia Model: Given the use and continuous research on the AKV model, we developed a Python tool for the AKV model that implements the model and provides a catalog of initial configurations, influence graphs, and belief update function already defined in the literature. The tool is available publicly and under the MIT license, allowing researchers to freely use it and contribute. With our tool we reproduced the results from Alvim, Amorim, et al. (2023) and our simulations correspond to theirs.

- Chapter 7 – A Python Tool for Paraconsistent Gödel Modal Logic: The paraconsistent Gödel modal logic, or $\mathbf{KG}^2$, is a modal comparative logic equipped with Kripke models that can represent agents and belief with uncertainty and

inconsistencies. These characterists allow a more interesting and powerful way to model social networks. We implemented a Python tool for instantiation of model and evaluation of formulas in $\mathbf{KG}^2$. The tool is available publicly and under the MIT license, allowing researchers to freely use it and contribute.

2 Background

2.1 Subjective Logic

Alvim, Knight, and Valencia (2019) proposed to use Subjective Logic (Jøsang 2016) to reason about opinions and opinion update. Subjective logic is a logic which includes information both about agent's belief as well their *certainty* about their beliefs. Although there are researches about Subjective Logic's belief fusion (Jøsang, Wang, and Zhang 2017) and trust networks (Jøsang, Hayward, and Pope 2006), opinion update by interacting over time has not been explored yet.

Subjective Logic is a logic developed by Jøsang (2016) that extends probabilistic logic by adding *uncertainty* and *subjectivity*. In probabilistic logic, a uniform distribution does not express *"we don't know"*, because it says that we know that the distribution over the domain is uniform. Subjective logic can express *uncertainty* about a distribution. The *subjectivity* comes from the fact that we can assign an opinion about a proposition to an agent.

2.1.1 Opinion representation

The main object of subjective logic is the *opinion*. We represent an opinion by ω_X^A, where A is an agent, X is a random variable, and ω_X^A is A's opinion about X. This section presents the elementary definitions that compose an opinion. A *domain* is a state space consisting of a set of values which can be called states, events, outcomes,

hypotheses or propositions. The values are assumed to be exclusive and exhaustive. E.g. $\mathbb{X} = \{\text{global warming is happening, global warming is not happening}\}$.

Belief mass is distributed over a domain. Let $\mathbb{X}$ be a domain. The belief mass assigned to a value $x \in \mathbb{X}$ expresses support for x being TRUE. Belief mass is sub-additive, i.e. $\sum_{x \in \mathbb{X}} \mathbf{b}_X(x) \leq 1$. The sub-additivity is complemented by *uncertainty mass* u_X and it represents vacuity of evidence, i.e. the lack of support for the variable X to have any specific value.

Definition 2.1 (Belief mass distribution)**.** Let $\mathbb{X}$ be a domain of size $k \geq 2$, and let X be a variable over that domain. A belief mass distribution denoted $\mathbf{b}_X$ assigns belief mass to possible values of the variable X. Belief mass and uncertainty mass sum to one, i.e., $u_X + \sum_{x \in \mathbb{X}} \mathbf{b}_X(x) = 1$.

Opinions can be semantically different, depending on the situation they apply to. An *aleatory* opinion applies to a variable governed by a frequentist process, and representing the likelihood of values of the variable in any unknown past or future instance of the process. "*The bias of a coin is* $p = 0.6$" is an aleatory opinion. An *epistemic* opinion applies to a variable that is assumed to be non-frequentist, and that represents the likelihood of the variables in a specific unknown instance. "*Beatriz killed Evandro*" is an epistemic opinion. In an epistemic opinion, opposite/different pieces of evidence should cancel each other out. Therefore, it must be uncertainty-maximized.

Base rate distribution represents a *prior* probability distribution over a domain, i.e. when we do not consider any evidence and are completely uncertain.

Definition 2.2 (Base rate distribution)**.** Let $\mathbb{X}$ be a domain of size $k \geq 2$, and X a random variable in $\mathbb{X}$. The base rate distribution $\mathbf{a}_X$ assigns base rate probability to possible values of $X \in \mathbb{X}$, and is an additive probability distribution.

Definition 2.3 (Opinion)**.** Let $\mathbb{X}$ be a domain of size $k \geq 2$, and X a random variable in $\mathbb{X}$. An opinion over the random variable X is the ordered triple $\omega_X = (\mathbf{b}_X, u_X, \mathbf{a}_X)$ where

- $\mathbf{b}_X$ is a belief mass distribution over X,
- u_X is the uncertainty mass which represents the vacuity of evidence,
- $\mathbf{a}_X$ is a base rate distribution over $\mathbb{X}$.

The *projected probability distribution* of an opinion is a *posterior* probability distribution after updating the base rate distribution with the belief mass distribution. The more an opinion depends on the belief mass, the less it depends on the base rate. The projected probability distribution is defined by:

$$\mathbf{P}_X(x) = \mathbf{b}_X(x) + \mathbf{a}_X(x)u_X, \ \forall x \in \mathbb{X} \tag{2.1}$$

The definition of opinion is useful for our model because it is more expressive than the belief state of an agent over a proposition by Alvim, Knight, and Valencia (2019), which is similarly an opinion of $k = 2$ with no uncertainty mass.

Example 2.1. Let $\mathbb{X} = \{x, \overline{x}\}$ be a domain where x is "global warming is happening" and $\overline{x}$ is "global warming is *not* happening". Let X be a random variable in $\mathbb{X}$. An opinion about X must be epistemic, because it is about a fact in the present instance that is true or false. Let the base rate be uniform. With no evidence, an agent A will hold the opinion $\omega_X^A = ((0,0), 1, (0.5, 0.5))$ with $\mathbf{P}_X^A(x) = 0.5$, meaning that A is 50% sure that the global warming is happening, but their opinion is relying only on the base rate, with no evidence supporting either of the values.

After gathering evidence from newspapers, scientific studies and other people, A assigns a belief mass to x. If agent A agrees with x by 80%, A holds the opinion

$\omega_X^A = ((0.6, 0), 0.4, (0.5, 0.5))$ with $\mathbf{P}_X^A(x) = 0.8$, meaning that A is 80% sure that global warming is happening, and has evidence that corresponds to 60% of their mass. The uncertainty mass means that A is relying 40% on the base rate.

2.1.2 Mapping a opinion to a Beta PDF

This section shows the mapping of an opinion to a Beta PDF defined by Jøsang (2016). We show the evidence notation of the multivariate Beta PDF (or Dirichlet PDF) and show its equivalence with the opinion representation defined at Def. 2.3.

A multinomial probability density over a domain of cardinality k is expressed by the k-dimensional Beta PDF. Assume a domain $\mathbb{X}$ of cardinality k, and a random variable X over $\mathbb{X}$ with probability distribution $\mathbf{p}_X$. The Beta PDF can be used to represent *probability density* over $\mathbf{p}_X$.

The multivariate Beta PDF takes as a variable the k-dimensional probability distribution $\mathbf{p}_X$. The strength parameters for the k possible outcomes are represented as k positive real numbers $\alpha_X(x)$, each corresponding to one of the possible values $x \in \mathbb{X}$. The strength parameters represent evidence/observations of $X = x$ where $x \in \mathbb{X}$.

Definition 2.4 (Multivariate Beta Probability Density Function)**.** Let $\mathbb{X}$ be a domain consisting of k mutually disjoint values. Let α_X represent the strength vector over the values of $\mathbb{X}$, and let $\mathbf{p}_X$ denote the probability distribution over $\mathbb{X}$. With $\mathbf{p}_X$ as a k-dimensional variable, the multivariate Beta PDF denoted $\text{Beta}(\mathbf{p}_X, \alpha_X)$ is expressed as:

$$\text{Beta}(\mathbf{p}_X, \alpha_X) = \frac{\Gamma\left(\sum_{x \in \mathbb{X}} \alpha_X(x)\right)}{\prod_{x \in \mathbb{X}} \Gamma(\alpha_X(x))} \prod_{x \in \mathbb{X}} \mathbf{p}_X(x)^{(\alpha_X(x)-1)}, \text{ where } \alpha_X(x) \geq 0, \tag{2.2}$$

with the restrictions that $\mathbf{p}_X(x) \neq 0$ if $\alpha_X(x) < 1$. Γ is the gamma probability distribution.

Now assume that $x \in \mathbb{X}$ represents a frequentist event. Let $\mathbf{r}_X(x)$ denote the number of observations for x. The strength vector α_X can be expressed as a function of the observations $\mathbf{r}_X(x)$ and the base rate $\mathbf{a}_X$:

$$\alpha_X(x) = \mathbf{r}_X(x) + \mathbf{a}_X(x)W, \text{ where } \mathbf{r}_X(x) \geq 0 \; \forall x \in \mathbb{X}. \tag{2.3}$$

By expressing the strength vector α_X in terms of the evidence observation $\mathbf{r}_X$, the base rate $\mathbf{a}_X$, and the non-informative prior weight W, we get the representation of the Beta PDF denoted $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$. The exact definition of $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$ is described at Jøsang (2016).

Given a Beta PDF $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$, the expected distribution over $\mathbb{X}$ can be written as

$$\mathbf{E}_X(x) = \frac{\mathbf{r}_X(x) + \mathbf{a}_X(x)W}{W + \sum\limits_{x_j \in \mathbb{X}} \mathbf{r}_X(x_j)}, \; \forall x \in \mathbb{X}. \tag{2.4}$$

The Beta model translates observation evidence directly into a PDF over a k-component probability variable. The representation evidence, together with the base rate, can be used to determine subjective opinions. In other words, it is possible to define a bijective mapping between Beta PDFs and opinions.

The bijective mapping between ω_X and $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$ is based on the requirement for equality between the projected probability distribution $\mathbf{P}_X$ derived from ω_X and the expected probability distribution $\mathbf{E}_X$ derived from $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$. This means that the more evidence in favor of a particular outcome x, the greater the belief mass on the outcome. Furthermore, the more total evidence available, the less uncertainty mass.

Definition 2.5 (Mapping: Opinion $\leftrightarrow$ Beta PDF)**.** Let $\omega_X = (\mathbf{b}_X, u_X, \mathbf{a}_X)$ be an opinion and let $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X, \mathbf{a}_X)$ be a Beta PDF, both over the same random variable X over the domain $\mathbb{X}$. These are equivalent through the following mapping,

$$\forall x \in \mathbb{X}$$

$$\begin{cases} \mathbf{b}_X(x) &= \dfrac{\mathbf{r}_X(x)}{W + \sum\limits_{x_i \in \mathbb{X}} \mathbf{r}_X(x_i)} \\ u_X &= \dfrac{W}{W + \sum\limits_{x_i \in \mathbb{X}} \mathbf{r}_X(x_i)} \end{cases} \Leftrightarrow \begin{cases} \begin{cases} \mathbf{r}_X(x) = \dfrac{W\mathbf{b}_X(x)}{u_X} \\ 1 = u_X + \sum\limits_{x_i \in \mathbb{X}} \mathbf{b}_X(x_i) \end{cases} & \text{if } u_X \neq 0 \\ \begin{cases} \mathbf{r}_X(x) = \mathbf{b}_X(x) \cdot \infty \\ 1 = \sum\limits_{x_i \in \mathbb{X}} \mathbf{b}_X(x_i) \end{cases} & \text{if } u_X = 0 \end{cases} \tag{2.5}$$

This equivalence between opinions and Beta PDFs is very powerful because it makes it possible to determine opinions from statistical observations.

2.1.3 Trust opinions

To model influence that one agent has over another, the subjective logic has *trust opinion* (Jøsang 2016), an opinion that an agent has about another agent as a good source of information.

Definition 2.6 (Trust opinion)**.** Let $\mathbb{T}_B = \{t_B, \bar{t}_B\}$ be a *trust domain*, where t_B means "B is a good source of information" and $\bar{t}_B$ means "B is not a good source of information". Then $\omega_{t_B}^A$, or ω_B^A for short, is the (trust) opinion that A has about the trustworthiness of B as a source of information.

We use trust opinions because subjective logic has the trust-discounting operator ($\otimes$). The idea of trust-discounting is to generate a new opinion $\omega_X^{[A;B]}$ by taking belief mass of ω_X^B proportionally by the belief mass of the trust opinion ω_B^A. Below,

$\omega_X^{[A;B]}$ represents A's opinion about X after communicating with B, ω_B^A represents A's opinion about the trustworthiness of B, and ω_X^B represents B's opinion about X.

Definition 2.7 (Trust-discounting)**.** Let ω_B^A be a trust opinion that agent A has about B and ω_X^B be the opinion that agent B has about the random variable X. Then, the opinion that agent A has about the random variable X trust-discounted by agent B is

$$\omega_X^{[A;B]} = \omega_B^A \otimes \omega_X^B = \begin{cases} \mathbf{b}_X^{[A;B]}(x) & = \mathbf{P}_B^A \mathbf{b}_X^B(x) \\ u_X^{[A;B]} & = 1 - \mathbf{P}_B^A(t_B) \sum_{x \in \mathbb{X}} \mathbf{b}_X^B(x) \\ \mathbf{a}_X^{[A;B]}(x) & = \mathbf{a}_X^B(x) \end{cases} \tag{2.6}$$

Example 2.2. Let $\omega_B^A = ((0.5, 0.5), 0, \mathbf{a}_B^A)$ with $\mathbf{P}_B^A(t_B) = 0.5$ and $\omega_X^B = ((0.8, 0), 0.2, (0.5, 0.5))$ with $\mathbf{P}_X^A(x) = 0.9$. Here, A trusts B by 50% and B is 80% sure that x is true with 60% of their mass assigned to x.

$\mathbf{P}_B^A(t_B) = 0.5$. Then, $[A; B]$ will hold 50% of the belief mass of each value from B. Therefore, $\omega_X^{[A;B]} = ((0.4, 0), 0.6, (0.5, 0.5))$ with $\mathbf{P}_X^{[A;B]}(x) = 0.7$. By the opinion that A has about X by trusting B, A is 70% sure that x is true with 40% of their mass assigned to x.

Example 2.3. Let $\omega_B^A = ((0.5, 0.5), 0, \mathbf{a}_B^A)$ with $\mathbf{P}_B^A(t_B) = 0.5$ and $\omega_X^B = ((0.8, 0), 0.2, (0.5, 0.5))$ with $\mathbf{P}_X^A(x) = 0.9$. Here, A trusts B by 50% and B is 80% sure that x is true with 60% of their mass assigned to x.

$\mathbf{P}_B^A(t_B) = 0.5$. Then, $[A; B]$ will hold 50% of the belief mass of each value from B. Therefore, $\omega_X^{[A;B]} = ((0.4, 0), 0.6, (0.5, 0.5))$ with $\mathbf{P}_X^{[A;B]}(x) = 0.7$. By the opinion that A has about X by trusting B, A is 70% sure that x is true with 40% of their mass assigned to x.

2.1.4 Belief fusion

To model A's concurrent interactions with multiple other agents, we can use belief fusion (Jøsang 2016; Jøsang, Wang, and Zhang 2017). Belief fusion combines a set of opinions into a single opinion which then represents the opinion of the merged source. There is more than one possible definition for the belief fusion operator. They differ by their properties and applications.

Cumulative belief fusion The cumulative belief fusion is used when it is assumed that the amount of independent evidence increases by including more and more sources. The idea is to sum the amount of evidence of the opinions. E.g. a set of agents flips a coin a number of times and produces an opinion about the bias of the coin. An opinion produced by cumulative belief fusion represents all the experiments made by the agents.

Let $\mathbb{X}$ be a domain, and X be a random variable over $\mathbb{X}$. W.l.o.g., let A and B be agents, and ω_X^A and ω_X^B be opinions. The cumulative belief fusion between A and B is denoted $\omega_X^{(A \diamond B)}$ and it can be represented as a Beta PDF.

Definition 2.8 (Cumulative belief fusion)**.** Let $\mathbb{X}$ be a domain, and X be a random variable over $\mathbb{X}$. Let A and B be agents, and ω_X^A and ω_X^B be opinions. The cumulative belief fusion between ω_X^A and ω_X^B is denoted:

$$\omega_X^{(A \diamond B)} = \omega_X^A \oplus \omega_X^B \tag{2.7}$$

Let $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^A, \mathbf{a}_X^A)$ and $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^B, \mathbf{a}_X^B)$ be the Beta PDFs equivalent to ω_X^A and ω_X^B, respectively.

The opinion $\omega_X^{(A \diamond B)}$ is the opinion equivalent to $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^{(A \diamond B)}, \mathbf{a}_X^{(A \diamond B)})$ defined

as:

$$\begin{aligned} \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^{(A \diamond B)}, \mathbf{a}_X^{(A \diamond B)}) &= \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^A, \mathbf{a}_X^A) \oplus \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^B, \mathbf{a}_X^B) \\ &= \text{Beta}_X^{\text{e}}(\mathbf{p}_X, (\mathbf{r}_X^A + \mathbf{r}_X^B), \mathbf{a}_X^{(A \diamond B)}). \end{aligned} \tag{2.8}$$

More specifically, for each values $x \in \mathbb{X}$ the accumulated source evidence $\mathbf{r}^{(A \diamond B)}$ is computed as:

$$\mathbf{r}^{(A \diamond B)}(x) = \mathbf{r}_X^A(x) + \mathbf{r}_X^B(x), \ \forall x \in \mathbb{X}. \tag{2.9}$$

The fusion of base rates is the average between the base rate distributions.

The fusion of three or more opinions is defined at Jøsang, Wang, and Zhang (2017). It can be verified that the cumulative fusion operator is commutative, associative, and non-idempotent.

Averaging belief fusion Averaging belief fusion is used when including more sources does not mean that more evidence is supporting the conclusion. The idea is to take the average of the amount of evidence of the opinions. E.g. After observing the court proceedings, each member of a jury produces an opinion about the same evidences. The verdict is the fusion between those opinions.

Let $\mathbb{X}$ be a domain, and X be a random variable over $\mathbb{X}$. W.l.o.g., let A and B be agents, and ω_X^A and ω_X^B be opinions. The averaging belief fusion between A and B is denoted $\omega_X^{(A \underline{\diamond} B)} = \omega_X^A$ and it can be represented as a Beta PDF.

Definition 2.9 (Averaging belief fusion)**.** Let $\mathbb{X}$ be a domain, and X be a random variable over $\mathbb{X}$. Let A and B be agents, and ω_X^A and ω_X^B be opinions. The averaging belief fusion between ω_X^A and ω_X^B is denoted:

$$\omega_X^{(A \underline{\diamond} B)} = \omega_X^A \underline{\oplus} \omega_X^B \tag{2.10}$$

Let $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^A_X, \mathbf{a}^A_X)$ and $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^B_X, \mathbf{a}^B_X)$ be the Beta PDFs equivalent to ω^A_X and ω^B_X, respectively.

The opinion $\omega^{(A\underline{\diamond}B)}_X$ is the opinion equivalent to $\text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^{(A\underline{\diamond}B)}_X, \mathbf{a}^{(A\underline{\diamond}B)}_X)$ defined as:

$$\begin{aligned} \text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^{(A\underline{\diamond}B)}_X, \mathbf{a}^{(A\underline{\diamond}B)}_X) &= \text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^A_X, \mathbf{a}^A_X) \underline{\oplus} \text{Beta}^{\text{e}}_X(\mathbf{p}_X, \mathbf{r}^B_X, \mathbf{a}^B_X) \\ &= \text{Beta}^{\text{e}}_X(\mathbf{p}_X, (\mathbf{r}^A_X + \mathbf{r}^B_X)/2, \mathbf{a}^{(A\underline{\diamond}B)}_X). \end{aligned} \tag{2.11}$$

More specifically, for each values $x \in \mathbb{X}$ the accumulated source evidence $\mathbf{r}^{(A\diamond B)}$ is computed as:

$$\mathbf{r}^{(A\underline{\diamond}B)}(x) = \frac{\mathbf{r}^A_X(x) + \mathbf{r}^B_X(x)}{2}, \ \forall x \in \mathbb{X}. \tag{2.12}$$

The fusion of base rates is the average between the base rate distributions.

The fusion of three or more opinions is defined at Jøsang, Wang, and Zhang (2017). It can be verified that the averaging fusion operator is commutative, idempotent, but non-associative.

Weighted belief fusion Weighted belief fusion is used when we take the average of the amount of evidence of the opinions weighted by their lack of uncertainty. In particular, opinions with no belief mass are rejected. E.g. a group of medical doctors needs to decide on a diagnosis. Each of them has an opinion, but some of them are more certain (assigned more belief mass) than others. Those opinions must have more weight than the others upon fusion.

Let $\mathbb{X}$ be a domain, and X be a random variable over $\mathbb{X}$. W.l.o.g., let A and B be agents, and ω^A_X and ω^B_X be opinions. The averaging belief fusion between A and B is denoted $\omega^{(A\hat{\diamond}B)}_X = \omega^A_X$ and it can be represented as a Beta PDF.

Definition 2.10 (Weighted belief fusion)**.** Let $\mathbb{X}$ be a domain, and X be a random

variable over $\mathbb{X}$. Let A and B be agents, and ω_X^A and ω_X^B be opinions. The averaging belief fusion between ω_X^A and ω_X^B is denoted:

$$\omega_X^{(A\hat{\diamond}B)} = \omega_X^A \hat{\oplus} \omega_X^B \tag{2.13}$$

Let $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^A, \mathbf{a}_X^A)$ and $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^B, \mathbf{a}_X^B)$ be the Beta PDFs equivalent to ω_X^A and ω_X^B, respectively. Also, let c_X^A be the *confidence* that A has in their opinion. Formally:

$$c_X^A = 1 - u_X^A \tag{2.14}$$

The opinion $\omega_X^{(A\hat{\diamond}B)}$ is the opinion equivalent to $\text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^{(A\hat{\diamond}B)}, \mathbf{a}_X^{(A\hat{\diamond}B)})$ defined as:

$$\begin{aligned} \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^{(A\hat{\diamond}B)}, \mathbf{a}_X^{(A\hat{\diamond}B)}) &= \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^A, \mathbf{a}_X^A) \hat{\oplus} \text{Beta}_X^{\text{e}}(\mathbf{p}_X, \mathbf{r}_X^B, \mathbf{a}_X^B) \\ &= \text{Beta}_X^{\text{e}}(\mathbf{p}_X, (c_X^A \mathbf{r}_X^A + c_X^B \mathbf{r}_X^B)/2, \mathbf{a}_X^{(A\hat{\diamond}B)}). \end{aligned} \tag{2.15}$$

More specifically, for each values $x \in \mathbb{X}$ the accumulated source evidence $\mathbf{r}^{(A\hat{\diamond}B)}$ is computed as:

$$\mathbf{r}^{(A\hat{\diamond}B)}(x) = \frac{c_X^A \mathbf{r}_X^A(x) + c_X^B \mathbf{r}_X^B(x)}{c_X^A + c_X^B}, \ \forall x \in \mathbb{X}. \tag{2.16}$$

The fusion of base rates is the average between the base rate distributions.

The fusion of three or more opinions is defined at Jøsang, Wang, and Zhang (2017). It can be verified that the cumulative fusion operator is weighted, idempotent, but non-associative.

2.2 Alvim-Knight-Valencia model

In this chapter, we present the polarization model composed of "static" and "dynamic" elements. Then we present update functions, initial configurations, and influence graphs described in Alvim, Knight, and Valencia (2019),

The *static elements* represent a snapshot of a social network at a given point in time. They include the following components:

- A (finite) set

$$\mathcal{A} = \{0, 1, \cdots, n-1\} \tag{2.17}$$

  of $n \geq 1$ *agents*, representing the users of the social network.

- A proposition p representing a declarative sentence. For example, p could be the statement "*vaccines are safe*" or "*climate change is real and is caused by human activity*".

- A *belief configuration*

$$B : \mathcal{A} \to [0, 1] \tag{2.18}$$

  such that for each agent $i \in \mathcal{A}$, the value $B_i = B(i)$ represents the confidence of agent $i \in \mathcal{A}$ in the veracity of proposition p.

- A *polarization measure*

$$\rho : [0, 1]^{\mathcal{A}} \to \mathbb{R}^+ \tag{2.19}$$

  mapping belief configurations to the non-negative real-numbers. Given a belief configuration $B = (B_0, \cdots, B_{n-1})$, the value $\rho(B)$ indicates the polarization among all the agents in $\mathcal{A}$ given their beliefs $B_0, \cdots, B_{n-1}$. Polarization can be thought of as the opposite of consensus; a high polarization value indicates

disagreement between agents, while a lower polarization value represents a situation closer to consensus.

The *Esteban-Rey polarization measure*, from Esteban and Ray (1994), is a specific, widely-used family of polarization measures, originally developed for economic applications. For technical reasons, Esteban-Rey polarization measures depend on the division of the domain into a fixed number of "bins."

In Alvim, Amorim, et al. (2023), the Esteban and Ray (1994) polarization measure is used with k bins, as described below.

Definition 2.11 (Esteban-Ray Polarization measure)**.** Consider a (finite) discrete set $\mathcal{Y} = \{y_0, \cdots, y_{k-1}\}$ of size k, such that each $y_i \in \mathbb{R}$. Let $(\pi, y) = (\pi_0, \cdots, \pi_{k-1}, y_0, \cdots, y_{k-1})$ be a distribution on $\mathcal{Y}$ such that π_i is the frequency of the value $y_i \in \mathcal{Y}$ in the distribution. The *Esteban-Ray polarization measure* is defined as

$$\rho_{ER}(\pi, y) = K \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \pi_i^{1-\alpha} \pi_j |y_i - y_j|. \tag{2.20}$$

Definition 2.12 (k-bin polarization)**.** Let D_k be a discretization of the interval $[0, 1]$ into $k > 0$ consecutive non-overlapping, non-empty intervals *(bins)* $I_0, \cdots, I_{k-1}$.

Given a belief configuration B, let

$$\mathbf{bd}(B, D_k) = (\pi, y) = (\pi_0, \cdots, \pi_{k-1}, y_0, \cdots, y_{k-1}), \tag{2.21}$$

where each y_i is the mid-point of I_i, and π_i is the proportion of agents having their belief in I_i.

An Esteban-Ray polarization measure for belief configurations over D_k is a mapping $\rho : [0, 1]^{\mathcal{A}} \to \mathbb{R}^+$ such that for some Esteban-Ray polarization measure ρ_{ER}, we

have

$$\rho(B) = \rho_{ER}(\mathbf{bd}(B, D_k)) \tag{2.22}$$

for every belief configuration $B \in [0,1]^{\mathcal{A}}$.

The *dynamic elements* of the model formalize the evolution of agents' beliefs as they interact and communicate about their opinions over time. They include:

- A *time frame* $\mathcal{T} = \{0, 1, 2, \cdots\}$ representing the discrete passage of time.
- A *family of belief configurations*

  $$\{B^t : \mathcal{A} \to [0,1]\}_{t \in \mathcal{T}} \tag{2.23}$$

  such that each B^t is the belief configuration of agents in $\mathcal{A}$ with respect to proposition p at time step $t \in \mathcal{T}$.
- A *weighted directed graph*

  $$\mathcal{I} : \mathcal{A} \times \mathcal{A} \to [0,1]. \tag{2.24}$$

  The value $\mathcal{I}(i,j)$, written $\mathcal{I}_{i,j}$, represents the *direct influence* that agent i has on agent j, or the *weight* i carries with j.
- An *update function*

  $$\mu : (B^t, \mathcal{I}) \to B^{t+1} \tag{2.25}$$

  mapping belief configuration B^t to a new belief configuration at time $t+1$. This function models the evolution of agents' beliefs over time. In the AKV model, different possible update functions are considered, depending on the desired application, discussed next.

2.2.1 Update functions

Here we describe the update functions in the AKV work with the notation used in Alvim, Amorim, et al. (2023). Let B^t be a belief configuration at time $t \in \mathcal{T}$, and $\mathcal{I}$ be an influence graph. Let $\mathcal{A}_i = \{j \in \mathcal{A} \mid \mathcal{I}_{j,i} > 0\}$ be the set of *neighbors* of i.

- The *classic update function (authority non-confirmatory bias)* is the map $\mu^{AB} : (B^t, \mathcal{I}) \to B^{t+1}$ given by

$$B_i^{t+1} = B_i^t + \frac{1}{|A_i|} \sum_{j \in \mathcal{A}_i} \mathcal{I}_{j,i}(B_j^t - B_i^t), \tag{2.26}$$

  for every agent $i \in \mathcal{A}$.

- The *confirmation-bias update function* is the map $\mu^{CB} : (B^t, \mathcal{I}) \to B^{t+1}$ given by

$$B_i^{t+1} = B_i^t + \frac{1}{|A_i|} \sum_{j \in \mathcal{A}_i} \beta_{i,j}^t \mathcal{I}_{j,i}(B_j^t - B_i^t), \tag{2.27}$$

  for every agent $i \in \mathcal{A}$, and $\beta_{i,j}^t = 1 - |B_j^t - B_i^t|$ is the *confirmation-bias factor* of i with respect to j given their beliefs at time t. With this belief update, agent i discounts the beliefs of other agents more the more different they are from i's own beliefs.

2.2.2 Initial configurations

Here we describe the initial belief configurations used in the AKV work from Alvim, Amorim, et al. (2023).

- The *uniform* belief configuration represents a set of agents whose beliefs are as

varied as possible, all equally spaced in the interval $[0, 1]$: for every i

$$B_i^0 = {}^{i}\!/_{(n-1)} \tag{2.28}$$

- The *mildly polarized* belief configuration with agents evenly split into two groups with moderately dissimilar inter-group belief compared to intra-groups beliefs: for every i,

$$B_i^0 = \begin{cases} {}^{0.2i}\!/_{\lceil n/2 \rceil}, & \text{if } i < \lceil {}^{n}\!/_{2} \rceil \\ {}^{0.2(i-\lceil n/2 \rceil)}\!/_{(n-\lceil n/2 \rceil)} + 0.6 & \text{otherwise.} \end{cases} \tag{2.29}$$

- The *extremely polarized* belief configuration represents a situation in which half of the agents strongly believe the proposition, whereas half strongly disbelieve it: for every i,

$$B_i^0 = \begin{cases} {}^{0.2i}\!/_{\lceil n/2 \rceil}, & \text{if } i < \lceil {}^{n}\!/_{2} \rceil \\ {}^{0.2(i-\lceil n/2 \rceil)}\!/_{(n-\lceil n/2 \rceil)} + 0.8 & \text{otherwise.} \end{cases} \tag{2.30}$$

- The *tripolar* belief configuration with agents divided into three groups: for every i,

$$B_i^0 = \begin{cases} {}^{0.2i}\!/_{\lfloor n/3 \rfloor} & \text{if } i < \lfloor {}^{n}\!/_{3} \rfloor, \\ {}^{0.2(i-\lfloor n/3 \rfloor)}\!/_{(\lceil 2n/3 \rceil - \lfloor n/3 \rfloor)} + 0.4, & \text{if } \lfloor {}^{n}\!/_{3} \rfloor \leq i \leq \lceil {}^{2n}\!/_{3} \rceil, \\ {}^{0.2(i-\lfloor 2n/3 \rfloor)}\!/_{(n-\lfloor n/3 \rfloor)} + 0.8, & \text{otherwise.} \end{cases} \tag{2.31}$$

2.2.3 Influence graphs

Here we describe the influence graphs used in the AKV work from Alvim, Amorim, et al. (2023).

- The *circular* influence graph $\mathcal{I}^{circ}$ represents a social network in which agents can be organized in a circle in such a way each agent is only influenced by its predecessor and only influences its successor: for every i, j such that $i \neq j$,

$$\mathcal{I}^{circ}_{i,j} = \begin{cases} 0.5, & \text{if } (i+1) \bmod n = j, \\ 0, & \text{otherwise.} \end{cases} \tag{2.32}$$

- The *disconnected* influence graph $\mathcal{I}^{disc}$ representing a social network sharply divided into two groups in such a way that agents with the same group can considerably influence each other, just not all agents in the other group: for every i, j such that $i \neq j$,

$$\mathcal{I}^{disc}_{i,j} = \begin{cases} 0.5, & \text{if } i,j \text{ are both } < \lceil n/2 \rceil \text{ or both } \geq \lceil n/2 \rceil, \\ 0, & \text{otherwise.} \end{cases} \tag{2.33}$$

- The *faintly communicating* influence graph $\mathcal{I}^{faint}$ represents a social network divided into two groups that evolve mostly separately, with only faint communication between them. More precisely, agents from within the same group influence each other much more strongly than from different groups: for every i, j such that $i \neq j$,

$$\mathcal{I}^{faint}_{i,j} = \begin{cases} 0.5, & \text{if } i,j \text{ are both } < \lceil n/2 \rceil \text{ or both } \geq \lceil n/2 \rceil, \\ 0.1, & \text{otherwise.} \end{cases} \tag{2.34}$$

- The *unrelenting influencers* influence graph $\mathcal{I}^{unrel}$ represents a scenario in which two influencer agents (say, 0 and $n-1$) exert a significantly stronger influence on every other agent than those other agents have among themselves, and are uninfluenced by all other agents: for every i, j such that i,

$$\mathcal{I}^{unrel}_{i,j} = \begin{cases} 0.6, & \text{if } i = 0 \text{ and } j \neq n-1 \text{ or } i = n-1 \text{ and } j \neq 0, \\ 0, & \text{if } j = 0 \text{ or } j = n-1, \\ 0.1, & \text{if } 0 \neq i \neq n-1 \text{ and } 0 \neq j \neq n-1. \end{cases} \tag{2.35}$$

- The *malleable influencers* influence graph $\mathcal{I}^{malleable}$ represents a social network in which two agents have a strong influence on every other agent, and are weakly influenced by everyone else: for every i, j such that $i \neq j$,

$$\mathcal{I}^{malleable}_{i,j} = \begin{cases} 0.8, & \text{if } i = 0 \text{ and } j \neq n-1, \\ 0.4, & \text{if } i = n-1 \text{ and } j \neq 0, \\ 0.1, & \text{if } j = 0 \text{ or } j = n-1, \\ 0.1, & \text{if } 0 \neq i \neq n-1 \text{ and } 0 \neq j \neq n-1. \end{cases} \tag{2.36}$$

2.3 Paraconsistent Gödel Modal Logic

In this section, we present the Paraconsistent Gödel Modal Logic, introduced by Bílková, Frittella, and Kozhemiachenko (2022). Paraconsistent Gödel Modal Logic, or $\mathbf{KG}^2$, is a logic to reason about about graded, incomplete, and inconsistent information.

Truth-values assigned to a state are values in the interval $[0,1] \times [0,1]$ encoding positive and negative information the statement in the state. Such states can represent possible worlds or possible agents. If an agent assigns value $(0,1)$ to a

proposition p, it means they have no evidence for p and strong evidence against p, meaning they definitely do not believe p. If they assign value $(1, 0)$ to p, they have strong evidence for p and no evidence against p, representing strong belief in p. If they assign $(0.2, 0.8)$ to p, they have some evidence in favor of p but stronger evidence against p. More complicated situations can be represented as well: for example, a value of $(0.75, 0.75)$ for p would mean that the agent has fairly strong evidence both for p and against p, as could occur in a social network when someone receives information from different, possibly unreliable, agents, or has inconsistent beliefs for other reasons. Similarly, a value of $(0.2, 0.1)$ for p represents a high level of uncertainty about p, but more evidence against p than in favor of p. Thus, $\mathbf{KG}^2$ is valuable for representing nuanced information about evidence, strength of belief, consistency and inconsistency, and certainty and uncertainty, making it more interesting and powerful for modelling social networks than a simple spectrum of belief.

In this section we present the language and semantics of $\mathbf{KG}^2$ (Bílková, Frittella, and Kozhemiachenko 2022). Let Var be a countable set of propositional variables. The language $\mathsf{bi}\mathcal{L}^{\neg}_{\Box,\Diamond}$ of $\mathbf{KG}^2$ is defined by the following grammar:

$$\phi := p \in \mathsf{Var} \mid \neg\phi \mid (\phi \wedge \phi) \mid (\phi \vee \phi) \mid (\phi \to \phi) \mid (\phi \prec \phi) \mid \Box\phi \mid \Diamond\phi \tag{2.37}$$

Definition 2.13 (Bi-Gödel algebra)**.** A *bi-Gödel algebra*

$$[0,1]_{\mathsf{G}} = ([0,1], 0, 1, \wedge_{\mathsf{G}}, \vee_{\mathsf{G}}, \to_{\mathsf{G}}, \prec_{\mathsf{G}}) \tag{2.38}$$

is defined as follows: for all $a, b \in [0,1]$, the standard operators are given by:

$$a \wedge_{\mathsf{G}} b := \min(a,b), \qquad a \vee_{\mathsf{G}} b := \max(a,b)$$
$$a \to_{\mathsf{G}} b := \begin{cases} 1, & \text{if } a \leq b, \\ b, & \text{otherwise,} \end{cases} \qquad b \prec_{\mathsf{G}} a := \begin{cases} 0, & \text{if } b \leq a, \\ b, & \text{otherwise.} \end{cases} \tag{2.39}$$

$a \to_{\mathsf{G}} b$ is called Gödel implication, and $a \prec_{\mathsf{G}} b$ is called Gödel comimplication. If a is less true than b, then $a \to_{\mathsf{G}} b$ is completely true. Otherwise, $a \to_{\mathsf{G}} b$ is as true as b. Simlilarly, $b \prec_{\mathsf{G}} a$ is completely false when a is *more* true than b, otherwise $b \prec_{\mathsf{G}} a$ is as true as b.

The sources available to the agents as well the connections between these sources can be represented as states in a Kripke model and its accessibility relation, respectively.

Definition 2.14 (Frames)**.** A *crisp frame* is a tuple $\mathfrak{F} = \langle W, R \rangle$ with $W \neq \emptyset$ and $R \subseteq W \times W$. A *fuzzy frame* is a tuple $\mathfrak{F} = \langle W, R \rangle$ with $W \neq \emptyset$ and $R : W \times W \to [0,1]$.

In the context of social networks, the accessibility relation in a crisp frame represents the *availability* between agents, while in a fuzzy frame, it represents *degrees of trust*.

To reason with the information provided by the sources, the modalities $\Box$ (infima) and $\Diamond$ (suprema) are used. $\Box$ and $\Diamond$ can be viewed as two simple aggregation strategies: pessimistic and optimistic. These modalities allow an agent to optimistically or pessimistically, respectively, aggregate all the information they receive from the other agents, using their trust for each other agent, represented by an accessibility relation, and taking all the evidence both for and against the proposition into account.

Models with this language and these modalities are first defined in bi-Gödel modal logic, or **K**biG. This logic is not paraconsistent and only evaluates propositions with repect to information in favor.

Definition 2.15 (**K**biG models)**.** A **K**biG model is a tuple $\mathfrak{M} = \langle W, R, v\rangle$ with $\langle W, R\rangle$ being a (crisp or fuzzy) frame, and $v : \mathsf{Var} \times W \to [0, 1]$. A valuation v is an interpretation defined as follows:

$$v(\phi \circ \phi') = v(\phi, w) \circ_{\mathsf{G}} v(\phi', w) \text{ for every } \circ \in \{\wedge, \vee, \to, \prec\} \tag{2.40}$$

The interpretation of modal formulas on fuzzy frames is as follows:

$$v(\Box\phi) = \inf_{w' \in W} \{wRw' \to_{\mathsf{G}} v(\phi, w')\}, \quad v(\Diamond\phi, w) = \sup_{w' \in W} \{wRw' \wedge_{\mathsf{G}} v(\phi, w')\} \tag{2.41}$$

On crisp frames, the interpretation is simpler (here, $\inf(\emptyset) = 1$) and $\sup(\emptyset) = 0$:

$$v(\Box\phi) = \inf_{w' \in W} \{v(\phi, w') : wRw'\}, \quad v(\Diamond\phi, w) = \sup_{w' \in W} \{v(\phi, w') : wRw'\} \tag{2.42}$$

$\Diamond\phi$ represents the search for evidence from trusted sources that *supports* ϕ: $v(\Diamond\phi, t) > 0$ iff there must be a source t' to which t has positive degree of trust and that has at least some certainty in ϕ. $\Box\chi$ represents the search for evidence against χ given by trusted sources: $v(\Box\chi, t) < 1$ iff there is a source t' that gives to χ less certainty than t gives trust to t' (Bílková, Frittella, and Kozhemiachenko 2022).

On the other hand, Bílková, Frittella, and Kozhemiachenko (2022) define a $\mathbf{KG}^2$ model only for crisp frame. Recall that $\mathbf{KG}^2$'s truth-values encodes positive and negative information about a statement.

Definition 2.16 ($\mathbf{KG}^2$ models for crisp frames)**.** A $\mathbf{KG}^2$ model is a tuple $\mathfrak{M} = \langle W, R, v_1, v_2\rangle$ with $\langle W, R\rangle$ being a *crisp* frame, and $v_1, v_2 : \mathsf{Var} \times W \to [0,1]$. We interpret v_1 as support of truth and v_2 support of falsity. These valuations are extended to complex formulas as follows:

$$\begin{aligned}
v_1(\neg\phi, w) &= v_2(\phi, w)\\
v_1(\phi \wedge \phi', w) &= v_1(\phi, w) \wedge_{\mathsf{G}} v_1(\phi', w)\\
v_1(\phi \vee \phi', w) &= v_1(\phi, w) \vee_{\mathsf{G}} v_1(\phi', w)\\
v_1(\phi \to \phi', w) &= v_1(\phi, w) \to_G v_1(\phi', w)\\
v_1(\phi \prec \phi', w) &= v_1(\phi, w) \prec_G v_1(\phi', w)\\
v_1(\Box\phi, w) &= \inf\{v_1(\phi, w') : wRw'\}\\
v_1(\Diamond\phi, w) &= \sup\{v_1(\phi, w') : wRw'\}
\end{aligned} \tag{2.43}$$

$$\begin{aligned}
v_2(\neg\phi, w) &= v_1(\phi, w)\\
v_2(\phi \wedge \phi', w) &= v_2(\phi, w) \vee_{\mathsf{G}} v_2(\phi', w)\\
v_2(\phi \vee \phi', w) &= v_2(\phi, w) \wedge_{\mathsf{G}} v_2(\phi', w)\\
v_2(\phi \to \phi', w) &= v_2(\phi', w) \prec_G v_2(\phi, w)\\
v_2(\phi \prec \phi', w) &= v_2(\phi', w) \to_G v_2(\phi, w)\\
v_2(\Box\phi, w) &= \sup\{v_1(\phi, w') : wRw'\}\\
v_2(\Diamond\phi, w) &= \inf\{v_1(\phi, w') : wRw'\}
\end{aligned} \tag{2.44}$$

We say that $\phi \in \mathsf{biL}^{\neg}_{\Box,\Diamond}$ is $\mathbf{KG}^2$ valid on frame $\mathfrak{F}$ iff for any $w \in \mathfrak{F}$, it holds that $v_1(\phi, w) = 1$ and $v_2(\phi, w) = 0$ for any model $\mathfrak{M}$ on $\mathfrak{F}$.

The work of Bílková, Frittella, and Kozhemiachenko (2022) defines a model $\mathbf{KG}^2$ only for crisp frames. In the necessity of modeling social networks with the trust between agents within the range of $[0,1]$, we use an extension of this model applied in fuzzy frames. In Bílková, Frittella, and Kozhemiachenko (2023), a model for $\mathbf{KG}^2$

for fuzzy frames is presented, but with a bi-relation between agents. They define two relations R^+ and R^- to be used in evaluations of modalities on v_1 and v_2. The relation R^+ represents degree of trust in afirmation and R^- represents degree of denial by a given source. In this work we use our own definition of $\mathbf{K}\mathsf{G}^2$ models which uses only one accessiblity relation that represents degree of trust, as defined in the original work. Also, to interpret the modalities on v_2, we negate ($\neg$) the respective definitions on v_1.

Definition 2.17 ($\mathbf{K}\mathsf{G}^2$ models for fuzzy frames)**.** A $\mathbf{K}\mathsf{G}^2$ is a tuple $\mathfrak{M} = \langle W, R, v_1, v_2\rangle$ with $\langle W, R\rangle$ being a *fuzzy* frame, and $v_1, v_2 : \mathsf{Var} \times W \to [0, 1]$. The valuations which we interpret as support of truth and support of falsity, respectively, are extended to complex formulas as expected:

$$\begin{aligned}
v_1(\neg\phi, w) &= v_2(\phi, w)\\
v_1(\phi \wedge \phi', w) &= v_1(\phi, w) \wedge_{\mathsf{G}} v_1(\phi', w)\\
v_1(\phi \vee \phi', w) &= v_1(\phi, w) \vee_{\mathsf{G}} v_1(\phi', w)\\
v_1(\phi \to \phi', w) &= v_1(\phi, w) \to_G v_1(\phi', w)\\
v_1(\phi \prec \phi', w) &= v_1(\phi, w) \prec_G v_1(\phi', w)\\
v_1(\Box\phi, w) &= \inf_{w' \in W}\{wRw' \to_{\mathsf{G}} v_1(\phi, w')\}\\
v_1(\Diamond\phi, w) &= \sup_{w' \in W}\{wRw' \wedge_{\mathsf{G}} v_1(\phi, w')\}
\end{aligned} \tag{2.45}$$

$$
\begin{aligned}
v_2(\neg\phi, w) &= v_1(\phi, w) \\
v_2(\phi \wedge \phi', w) &= v_2(\phi, w) \vee_{\mathsf{G}} v_2(\phi', w) \\
v_2(\phi \vee \phi', w) &= v_2(\phi, w) \wedge_{\mathsf{G}} v_2(\phi', w) \\
v_2(\phi \to \phi', w) &= v_2(\phi', w) \prec_G v_2(\phi, w) \\
v_2(\phi \prec \phi', w) &= v_2(\phi', w) \to_G v_2(\phi, w) \\
v_2(\Box\phi, w) &= \sup_{w' \in W}\{wRw' \wedge_{\mathsf{G}} v_2(\phi, w')\} \\
v_2(\Diamond\phi, w) &= \inf_{w' \in W}\{wRw' \to_{\mathsf{G}} v_2(\phi, w')\}
\end{aligned}
\tag{2.46}
$$

We say that $\phi \in \mathsf{bi}\mathcal{L}^{\neg}_{\Box,\Diamond}$ is $\mathbf{K}\mathsf{G}^2$ valid on frame $\mathfrak{F}$ iff for any $w \in \mathfrak{F}$, it holds that $v_1(\phi, w) = 1$ and $v_2(\phi, w) = 0$ for any model $\mathfrak{M}$ on $\mathfrak{F}$.

3 Related Work

Research about the dynamics of opinions in social networks using formal models and logic-based approaches is fairly new. Most of the papers collected by this work North America and Europe. Recent political events like Brexit, United States Donald Trump's election showed that social media platforms had a role with respect to opinion-making and the spread of fake news (J. Jackson 2017).

3.1 Logical approaches to social networks

This section describes some of the approaches using logic systems to describe social networks.

Baltag and Smets (2008a) and Baltag and Smets (2008b) worked with dynamic epistemic logic to build probabilistic models with build revision and update. They axiomatized the their logic for belief revision and the update mechanism, demonstrated that they can "simulate" many different belief-revision policies. Using dynamic epistemic logic, Baltag, Christoff, et al. (2019a) created a threshold model which agents change the behavior when the proportion of supporters changes. Seligman, Liu, and Girard (2013) use dynamic epistemic logic to investigate conceptual issues about communication between agents in a social network and broadcast of information to group of agents.

With modal logic, Christoff and Hansen (2015) and Christoff, Hansen, and Proietti (2016) developed a tableau system to represent reasoning in social networks.

Their framework has agents leaning about other agents and differentiate between private opinion and public opinion. Other works uses information flow, dynamic epistemic logic and 2-colorable networks to reason about social networks. (Baltag, Christoff, et al. 2019b)

Liu, Seligman, and Girard (2014) use ideas from doxastic logic and dynamic epistemic logic to qualitatively model influence and belief changes in social networks. They aim to reason about how people change their belief when influenced by others. By a model called 'threshold influence', they capture many aspects of social belief change.

Hunter (2017) introduces a logic of belief updates over social networks where closer agents in the social network are more trusted and thus more influential. They develop a formal model to capture the perceived areas of expertise of each agent, and they introduce a logical operator to determine how beliefs change following reported information. They prove a number of formal properties, and demonstrate that our approach can actually model a wide range of practical trust problems involving social agents.

Delgrande, Renne, and Sack (2019) developed a logic for qualitative probability and a operator to compare probability of a sequence of events. Fagin, Halpern, and Megiddo (1990) developed a logic for reasoning about probabilities with Nilsson's probabilistic logic and Dempster-Shafer belief fusion.

Ditmarsch, Lang, and Saffidine (2013) use epistemic logic and Stackelberg games to develop a framework for strategic voting when a voter may lack knowledge about other votes.

3.2 Other formal models with multi-agent systems

This section describes some other approaches using formal modals with a multi-agent system to describe social networks.

Sîrbu et al. (2019) developed a formal model that introduces algorithmic bias to the condition for two agents to interact. Algorithmic bias increases opinion fragmentation and decreases convergence speed of agents belief. Li et al. (2013) developed a variation of Deffuant-Weishbuch model to find in what conditions clustering patterns emerges in social networks. They use simulations and applied data to validate results. Gargiulo and Gandica (2017) investigates the role of homophily and in social networks when a society faces a dominant message. They show that, contrary to the common idea, homophily helps consensus formation. Using DeGroot rule and Markov chain, Golub and M. O. Jackson (2010) develop a formal model in which the opinions in a large society converge to the truth if and only if the influence of the most influential agent vanishes as the society grows.

Alvim, Knight, and Valencia (2019) develop a preliminary formal model for social networks, and a measure of the level of polarization in these social networks based on the classic measure of polarization developed by Esteban and Ray (1994). Their model includes some elements like belief over a proposition, influence by an agent over another, rational and biased belief updates, and different initial scenarios for influence graphs. By simulations, they consider evolution over time, and the implications of these results for real-world social networks. In their model, polarization eventually disappears if the influence between agents is a strongly-connected graph. They also determine the speed of convergence of belief states and set conditions that imply a polarized society (Alvim, Amorim, et al. 2021; Alvim, Amorim, et al. 2023).

4 Opinion Update in a Subjective Logic Model for Social Networks

In this chapter, we introduce a model for social networks with subjective logic. With subjective logic, we use a multinomial opinion representation that allows expressing certainty and uncertainty about a proposition. In addition we use operators of subjective logic to express the relationship between agents and belief update.

This chapter is structed as follows:

- Section 4.1 – Model definition: Describe our model for social networks with subjective logic.
- Section 4.2 – Update function with trust discount and belief fusion: We define a desiderata of properties for an update function in social networks to represent rational reasoning and an update function using operators from subjective logic. We show that our new update function does not satisfies the desiderata.
- Section 4.3 – Analysis of the update function with cumulative belief fusion: We analyze the update function with cumulative belief fusion operator. We show that this update function is promising for modeling belief updates in social networks. This function can represent a radicalization phenomenon that happens when agents on the same side repeatedly comunicate or when one agent is transmitting more evidence than the other.

- Section 4.4 – Update function with Alvim-Knight-Valencia update function and Subjective Logic - We define a new update function that uses the features of multinomial representation of opinions and uncertainty and the classic update from Alvim, Knight, and Valencia (2019). We show that this update satisfies our desiderata.

4.1 Model definition

We describe how to use subjective logic to model a social network. Existing notions from subjective logic work well for modeling the relevant aspects of agents' opinions about an issue in a social network, except for the update function. Our goal is to expand subjective logic with an appropriate update function.

4.1.1 Static elements of the model

The *static elements* represent a snapshot of a social network at a given point in time. They include the following components:

- A (finite) set $\mathcal{A} = \{A_0, A_1, \ldots, A_{n-1}\}$ of $n \geq 1$ *agents*. An agent is a user in a social network

- A *domain* of k disjoint events $\mathbb{X} = \{x_0, x_1, \ldots, x_{k-1}\}$ and a random variable X over $\mathbb{X}$. A domain is a generalization of a proposition from binary logic by having multiple truth values for a proposition. A proposition from binary logic would have a domain of size 2. It can represent topics that cannot be answered with just a YES or NO. In the examples in this work, we use a domain of size two for simplicity, and the random variable X represents opinions about a single issue.

- A set of *opinions* $\{\omega_X^{A_j}\}_{A_i \in \mathcal{A}}$, one for each agent, about a single random variable X.

4.1.2 Dynamic elements of the model

The *dynamic elements* of the model formalize the evolution of agents' beliefs as they interact and communicate about their opinions over time. They include:

- A set of *trust opinions* $\{\omega_{A_j}^{A_i}\}_{A_i, A_j \in \mathcal{A}}$ over the domain $\mathbb{T}_{A_j} = \{t_{A_j}, \bar{t}_{a_j}\}$ where $A_i, A_j \in \mathcal{A}$ and $i \neq j$. Each trust opinion represents how much A_i trusts A_j as a source of evidence. We consider trust opinion as dogmatic, i.e. the uncertainty mass is $i_{T_j} = 0$ for all trust opinions. Therefore, $\mathbf{P}_{A_j}^{A_i}(t_{A_j}) = \mathbf{b}_{A_j}^{A_i}(t_{A_j})$. Throughout this paper, we represent a trust opinion only by $\mathbf{P}_{A_j}^{A_i}(t_{A_j})$.

- A *time frame* $\mathcal{T} = \{0, 1, 2, \ldots, t_{max}\}$ representing the discrete passage of time.

- An *update function* f such that

$$\omega_X^{A_i[t+1]} = f(\omega_X^{A_i[t]}, \{\omega_{A_j}^{A_i}\}_{A_j \in \mathcal{A}}, \{\omega_X^{A_j[t]}\}_{A_j \in \mathcal{A}}). \tag{4.1}$$

  An update function f takes A_i's opinions at time t and updates it using all other agents' opinions and trust opinions to them. Here we consider that agents have complete trust in themselves. Otherwise, we can define this function with A_i's opinion included in the set of opinions $\{\omega_{A_j}^{A_i}\}_{A_j \in \mathcal{A}}$.

4.2 Update function with trust discount and belief fusion

Our goal is to model agents' opinions in social networks and to develop an update function for opinions after agents interact. This section describes our proposal for an update function for two agents with belief fusion and trust opinions.

Definition 4.1 (Desiderata for the properties of an update function)**.** Let $\mathbb{X} = \{x, \overline{x}\}$ be a domain, X a random variable in $\mathbb{X}$, and A and B agents holding opinions $\omega_X^{A[t]}$ and $\omega_X^{B[t]}$ at time t. Let ω_B^A be a trust opinion. We need an update function $\omega_X^{A[t+1]} = f(\omega_X^{A[t]}, \omega_B^A, \omega_X^{B[t]})$ satisfying the following properties:

1. *Weak convergence*: $\mathbf{P}_X^{A[t]} \leq \mathbf{P}_X^{A[t+1]} \leq \mathbf{P}_X^{B[t]}$ or $\mathbf{P}_X^{A[t]} \geq \mathbf{P}_X^{A[t+1]} \geq \mathbf{P}_X^{B[t]}$. A cannot move further from B upon update.

2. *Markovian behaviour*: $\omega_X^{A[t+1]}$ must not depend on the opinions from time less than t. Each step of the update must be independent.

3. *Idempotence*: If $\omega_X^{A[t]} = \omega_X^{B[t]}$, then $\omega_X^{A[t]} = \omega_X^{A[t+1]}$. No matter the trust, if A and B have the same opinion, A must not change their opinion.

4. *Non-increasing uncertainty*: we have two versions of this desideratum.

   (a) *Non-increasing group uncertainty:* The sum of the uncertainty of all agents must not increase over time.

   (b) *Non-increasing individual uncertainty:* Each agent's individual uncertainty never increases.

These desiderata are our interpretation of rationality when updating opinions. The *weak convergence* property says that when an agent receives new information,

the belief is updated to a new value closer to the new information they receive, and they do not become more extreme. The *Markovian behavior* property says that our current belief contains all relevant information, so past beliefs are not taken into account to compute a new one. The *idempotence* property says that, if the new information an agent receives is identical to their current belief, then the belief remains unchanged. The *non-increasing uncertainty* property says that at least some agents do not become totally uncertain as they receive new information and communicate. We aim for one of the two versions of this property because we hope to avoid the situation where all agents become totally uncertain in the limit.

The intuition behind our planned update function is to fuse agent A's current opinion with all the opinions that A can gather by trusting other agents. Define a *dogmatic opinion* as an opinion with no uncertainty, i.e. $u_X = 0$. For this update function, we are not considering situations with dogmatic opinions, because it means the agent has an infinite amount of evidence and the belief fusion operators remove non-dogmatic opinions when at least one is present.

Definition 4.2 (Update function for 2 agents with belief fusion and trust)**.** Let $\omega_X^{A[t]}$ and $\omega_X^{B[t]}$ be non-dogmatic opinions. Let $\oplus$ be a belief fusion operator. Define the update function for $\omega_X^{A[t+1]}$ as

$$\omega_X^{A[t+1]} = \omega_X^{A[t]} \oplus (\omega_B^A \otimes \omega_X^{B[t]}) \tag{4.2}$$

We call $(\omega_B^A \otimes \omega_X^{B[t]})$ the opinion that A *will* learn by interacting with B. $\omega_X^{A[t+1]}$ is the opinion that A holds after merging their previous opinion $(\omega_X^{A[t]})$ with the opinion that A learned $(\omega_B^A \otimes \omega_X^{B[t]})$.

For this update function, we are not considering situations with dogmatic opinions, i.e. $u_X = 0$, because it means the agent has an infinite amount of evidence.

By definition, the update function for $\omega_X^{A[t+1]}$ is Markovian. Note that, if $\oplus$ is the averaging or weighted belief fusion operator, and $\omega_X^{A[t]} = \omega_B^A \otimes \omega_X^{B[t]}$, then $\omega_X^{A[t]} = \omega_X^{A[t+1]}$. This is different from our desired idempotence property, which does not depend on trust opinion.

Example 4.1. For brevity, we write the value of $\mathbf{P}_B^A(t_B)$ where it should be ω_B^A.

$$\begin{aligned} \omega_X^{A[t+1]} &= ((0,0),1,(0.5,0.5)) \oplus (0.5 \otimes ((0.8,0),0.2,(0.5,0.5))) \\ &= ((0,0),1,(0.5,0.5)) \oplus ((0.4,0),0.6,(0.5,0.5)) \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} \omega_X^{A[t]} &= ((0,0),1,(0.5,0.5)), & \mathbf{P}_X^{A[t]}(x) &= 0.5 \\ \omega_B^A &= ((0.5,0.5),0,\mathbf{a}_B^A), & \mathbf{P}_B^A(t_B) &= 0.5 \\ \omega_X^{B[t]} &= ((0.8,0),0.2,(0.5,0.5))) & \mathbf{P}_X^{A[t]}(x) &= 0.9 \end{aligned} \tag{4.4}$$

The final result depends on the choice of belief fusion operator.

Example 4.2. Now consider these two agents updating their opinion over time.

$$\begin{aligned} \omega_X^{A[0]} &= ((0.2,0),0.8,(0.5,0.5)) \quad \mathbf{P}_X^{A[0]}(x) = 0.6 \quad \mathbf{P}_B^A(t_B) = 0.5 \\ \omega_X^{B[0]} &= ((0.8,0),0.2,(0.5,0.5)) \quad \mathbf{P}_X^{B[0]}(x) = 0.9 \quad \mathbf{P}_A^B(t_A) = 0.5 \end{aligned} \tag{4.5}$$

Here, agent A is 60% sure about x and B is 90% about x. Both trust each other by 50%. The evolution of $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ is shown in Fig. 4.1.

We expected that the update function would satisfy the weak convergence property by $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ converging to some value between $\mathbf{P}_X^{A[0]}(x) = 0.6$ and $\mathbf{P}_X^{B[0]}(x) = 0.9$. But because evidence keeps accumulating over time, with cumulative belief fusion, $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ converge to 1.

For averaging and weighted belief fusion, $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ converge to 0.5, violating weak convergence. With epistemic opinions, increasing uncertainty over

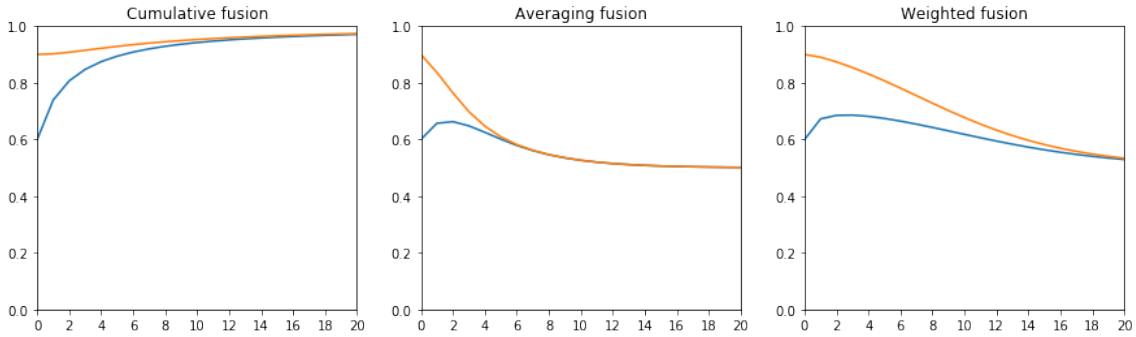


Figure 4.1: $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ updated 20 times as in Example 4.2.

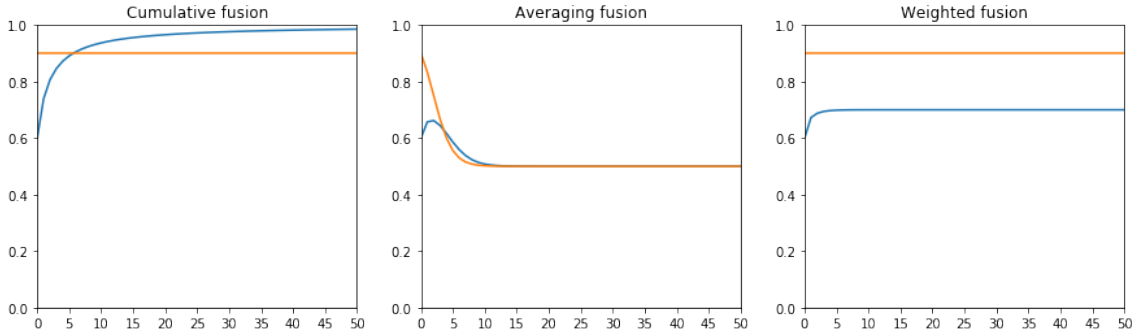


Figure 4.2: $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ updated 20 times as in Example 4.3.

time is expected, but the same happens with aleatory opinions.

Example 4.3. In this case, agent A trusts B by 0.5, but B does not trust A.

$$\begin{aligned} \omega_X^{A[0]} &= ((0.2, 0), 0.8, (0.5, 0.5)) \qquad \mathbf{P}_X^{A[0]}(x) = 0.6 \qquad \mathbf{P}_B^A(t_B) = 0.5 \\ \omega_X^{B[0]} &= ((0.8, 0), 0.2, (0.5, 0.5)) \qquad \mathbf{P}_X^{B[0]}(x) = 0.9 \qquad \mathbf{P}_A^B(t_A) = 0 \end{aligned} \tag{4.6}$$

Here, agent A is 60% and B 90% sure about x. A trusts B by 50%, but B does not trust A. The evolution of $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ is shown in Fig. 4.2.

For cumulative belief fusion, B does not change their opinion because $\omega_B^A \otimes \omega_X^{B[t]}$ has no belief mass, therefore, B cannot learn anything. However A keeps learning from B over time, and $\mathbf{P}_X^A(x)$ converges to 1.

The weighted belief fusion case shows how the idempotency works for its operator

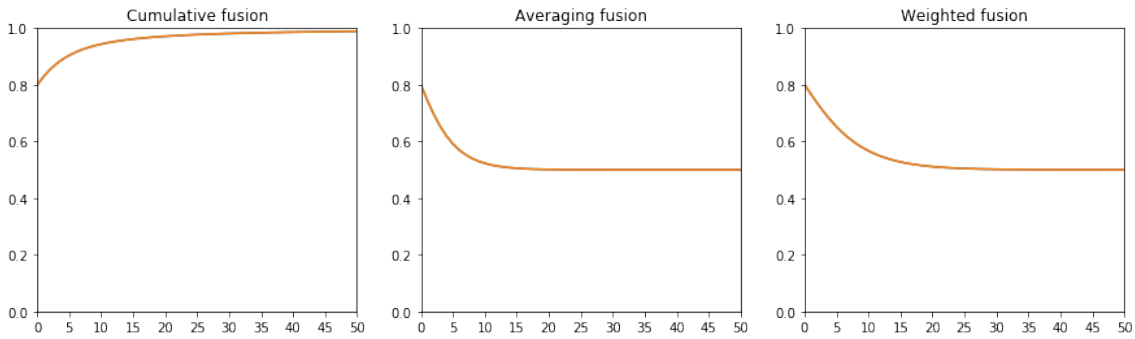


Figure 4.3: $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ updated 20 times as in Example 4.4.

and how it is different from the idempotence property that we are aiming for. When $\omega_X^A = ((0.4, 0), 0.6, (0.5, 0.5))$ with $\mathbf{P}_X^A(x) = 0.7$, $\omega_X^A = \omega_B^A \otimes \omega_X^B$. Therefore, $\mathbf{P}_X^A(x)$ converges to 0.7.

Example 4.4. In this case, A and B have the same opinion and the same trust.

$$\begin{aligned} \omega_X^{A[0]} &= ((0.6, 0), 0.4, (0.5, 0.5)) \qquad \mathbf{P}_X^{A[0]}(x) = 0.8 \qquad \mathbf{P}_B^A(t_B) = 0.5 \\ \omega_X^{B[0]} &= ((0.6, 0), 0.4, (0.5, 0.5)) \qquad \mathbf{P}_X^{B[0]}(x) = 0.8 \qquad \mathbf{P}_A^B(t_A) = 0.5 \end{aligned} \tag{4.7}$$

Here, A and B have the same opinion. They are 80% sure about x. Both trust each other by 50%. The evolution of $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ is shown in Fig. 4.3.

Even starting with the same opinion, agents A and B do not keep the same opinion over time. With cumulative belief fusion, A and B keep accumulating evidence and $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ converge to 1. With averaging or weighted belief fusion, uncertainty keeps increasing, and $\mathbf{P}_X^A(x)$ and $\mathbf{P}_X^B(x)$ converge to 0.5.

To sum up, these examples show that an update function with trust discount and belief fusion does not satisfies our desiderata. In the next section, we propose a update function using an earlier work on social networks that satisfies our desiderata.

4.3 Analysis of the update function with cumulative belief fusion

In this section, we analyze the cumulative belief fusion operator, which we believe in practice provides some interesting outcomes in simple simulations of social networks, even though it does not always weakly converge, as seen above. We will show that the cumulative belief function is promising for modeling belief updates in social networks.

Recalling the definition of cumulative belief fusion, it is assumed that the amount of evidence increases by including more sources. An opinion in subjective logic can be translated to a Beta PDF using the amount of evidence for each state in the domain as parameters. Say that $\mathbf{r}_X^A(x)$ are $\mathbf{r}_X^A(x)$ the amount of evidence supporting the state x for agents A and B respectively. The cumulative fusion operator gets the sum $\mathbf{r}_X^A(x) + \mathbf{r}_X^A(x)$ to be the amount of evidence of the merged opinion and translates it back to a subjective logic opinion. More details are presented in the appendix. For repeated interactions using the cumulative belief fusion, we interpret repeated evidence contained in an opinion as reinforcement of the belief in the state.

We did many simulations using the two-agent update function with cumulative fusion and we found three cases for initial epistemic opinions with the projected probability $\mathbf{P}_X^A(x)$ ranging from 0 to 1.

Suppose that $\mathbb{X} = \{x, \overline{x}\}$ and all opinions are epistemic. Recall that the update function for two agents with trust discount and belief fusion is

$$\omega_X^{A[t+1]} = \omega_X^{A[t]} \oplus (\omega_B^A \otimes \omega_X^{B[t]}) \tag{4.8}$$

where A and B are agents, $\omega_X^{A[t]}$ and $\omega_X^{B[t]}$ are non-dogmatic epistemic opinions, and

ω_B^A is a trust opinion. Also, recall that $\omega_B^A \otimes \omega_X^{B[t]}$ is the opinion that A learns after trusting B about X. The update function behaves like these three cases:

1. *Consensus*: This case happens when both agents are agreeing or disagreeing at the same time, not necessarily with the same projected probability, i.e.

$$\begin{gathered}\mathbf{P}_X^{A[0]}(x) < 0.5 \text{ and } \mathbf{P}_B^A(t_b) \otimes \mathbf{P}_X^{B[0]}(x) < 0.5 \text{ or} \\ \mathbf{P}_X^{A[0]}(x) > 0.5 \text{ and } \mathbf{P}_B^A(t_b) \otimes \mathbf{P}_X^{B[0]}(x) > 0.5\end{gathered} \tag{4.9}$$

   When these agents interact, they will accumulate evidence about the same outcome in each interaction. The fusion leads $\mathbf{b}_X^A(x)$ and $\mathbf{b}_X^B(x)$ to converge to 0 or 1, depending on what they are agreeing upon, and both with uncertainty mass to 0. Increasing the trust discount or the uncertainty will increase the speed of convergence.

   This represents a situation when two agents agree that x is TRUE, with different levels of projected probabilities. That leads them to be completely certain about the proposition.

   Example 4.5. Let A and B be initially in consensus agreeing that x is TRUE as follows.

$$\begin{aligned}\omega_X^{A[0]} &= ((0.2, 0.0), 0.8, (0.5, 0.5)) \qquad \mathbf{P}_X^{A[0]}x = 0.6 \qquad \mathbf{P}_B^A(t_b) = 0.5 \\ \omega_X^{B[0]} &= ((0.4, 0.0), 0.6, (0.5, 0.5)) \qquad \mathbf{P}_X^{B[0]}x = 0.7 \qquad \mathbf{P}_A^B(t_b) = 0.5\end{aligned} \tag{4.10}$$

   A is 60% sure about x and B is 70% sure about x. Both trust each other by 50%. On the limit of the interactions between A and B, both will be sure by 100% that x is true.

$$\mathbf{P}_X^{A[t]}(x) = \mathbf{P}_X^{B[t]}(x) \xrightarrow{t\to\infty} 1 \tag{4.11}$$

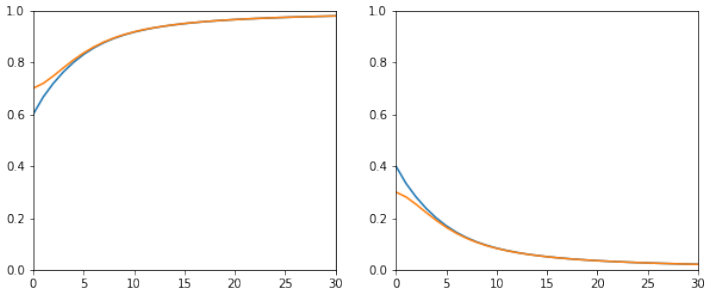


Figure 4.4: On the left, the evolution of $\mathbf{P}_X^{A[t]}(x)$ and $\mathbf{P}_X^{A[t]}(x)$ for the Example 4.5. On the right, the similar case for when both agents disagree that x is TRUE.

Similarly if both disagree that x is TRUE, i.e. that $\overline{x}$ is TRUE, $\mathbf{P}_X^{A[t]}(x)$ and $\mathbf{P}_X^{B[t]}(x)$ will converge to 0.

2. *Balanced opposite*: This case happens when both agents are learning the exact opposite opinion that A already had, i.e.

$$\mathbf{P}_X^{A[0]}(x) = 1 - \mathbf{P}_B^A(t_b) \otimes \mathbf{P}_X^{B[0]}(x), \ \forall x \in \mathbb{X} \tag{4.12}$$

Because contrary pieces of evidence cancel each other, the fusion leads $\mathbf{b}_X^{A[0]}(x)$ to be a vacuous opinion. The speed of convergence is defined by the trust opinion. The more an agent trusts another, the faster the convergence.

This represents a situation when two agents support opposite views but with the same projected probability. That leads them to be completely indecisive about the proposition.

Example 4.6. Let A and B have balanced opposite opinions.

$$\begin{aligned} \omega_X^{A[0]} &= ((0, 0.4), 0.6, (0.5, 0.5)) \qquad \mathbf{P}_X^{A[0]}(x) = 0.3 \qquad \mathbf{P}_B^A(t_B) = 1 \\ \omega_X^{B[0]} &= ((0.4, 0), 0.6, (0.5, 0.5)) \qquad \mathbf{P}_X^{B[0]}(x) = 0.7 \qquad \mathbf{P}_A^B(t_A) = 1 \end{aligned} \tag{4.13}$$

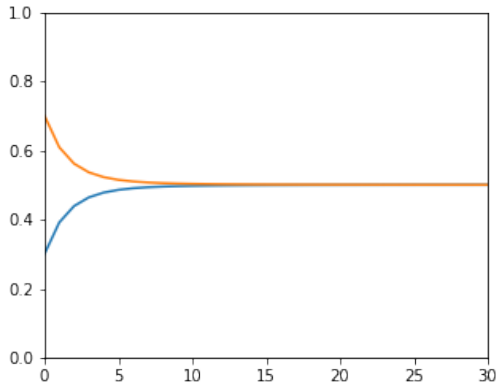


Figure 4.5: Evolution of $\mathbf{P}_X^{A[t]}(x)$ and $\mathbf{P}_X^{A[t]}(x)$ for the Example 4.6.

A is 30% sure about x and B is 70% sure about x. Both trust each other by 100%. On the limit of interactions between A and B, both will be sure by 50% that x is true.

$$\mathbf{P}_X^{A[t]}(x) = \mathbf{P}_X^{B[t]}(x) \xrightarrow{t\to\infty} 0.5 \tag{4.14}$$

3. *Unbalanced opposite*: This happens when there are opposite beliefs, but they don't have the same projected probability in their respective views. In other words, both are in conflict but one agent is more radical than the other.

$$\begin{gathered}(\mathbf{P}_X^{A[0]}(x) < 0.5 \text{ and } \mathbf{P}_B^A(t_b) \oplus \mathbf{P}_X^B > 0.5 \\ \text{or } \mathbf{P}_X^{A[0]}(x) > 0.5 \text{ and } \mathbf{P}_B^A(t_b) \oplus \mathbf{P}_X^B < 0.5) \\ \text{and } \mathbf{P}_X^{A[0]} \neq 1 - \mathbf{P}_B^A(t_b)\ \mathbf{P}_X^{B[0]}(x)\end{gathered} \tag{4.15}$$

In this case, contrary pieces of evidence will cancel each other, but one agent will transmit more evidence than the other upon interaction. The agents that transmit more evidence will win in the limit. Increasing the trust or becoming more radical will increase the speed of convergence.

- If A is sure that x is TRUE more than B is sure that x is FALSE discounted by the trust, i.e.

$$\mathbf{P}_X^{A[t]}(x) > 1 - \mathbf{P}_B^A(t_b) \otimes \mathbf{P}_X^{B[t]}(x) \tag{4.16}$$

  then, A and B will eventually both agree that x is TRUE in some degree, i.e.

$$\mathbf{P}_X^{A[t]}(x) = \mathbf{P}_X^{B[t]}(x) \xrightarrow{t\to\infty} p \in (0.5, 1] \tag{4.17}$$

- If A is sure that x is FALSE more than B is sure that x is TRUE discounted by the trust, i.e.

$$\mathbf{P}_X^{A[t]}(x) < 1 - \mathbf{P}_B^A(t_b) \otimes \mathbf{P}_X^{B[t]}(x) \tag{4.18}$$

  then, A and B will eventually both agree that x is FALSE in some degree, i.e.

$$\mathbf{P}_X^{A[t]}(x) = \mathbf{P}_X^{B[t]}(x) \xrightarrow{t\to\infty} p \in [0, 0.5) \tag{4.19}$$

The unbalanced opposite is the most interesting case. It can represent some bias when conflicting agents are interacting. The behavior is different from any update function described by Alvim, Knight, and Valencia (2019), After experiments, we found that

- If the agents are too far, they will radicalize, i.e. $\mathbf{P}_X^{A[t]}(x) \xrightarrow{t\to\infty} 0$ or $\mathbf{P}_X^{A[t]}(x) \xrightarrow{t\to\infty} 1$
- Otherwise, they will converge to some value at the winning outcome. $[0, 0.5)$ if A is sure that x is FALSE more than B is sure that x is TRUE discounted by the trust, or $(0.5, 1]$ if A is sure that x is TRUE more than B is sure that

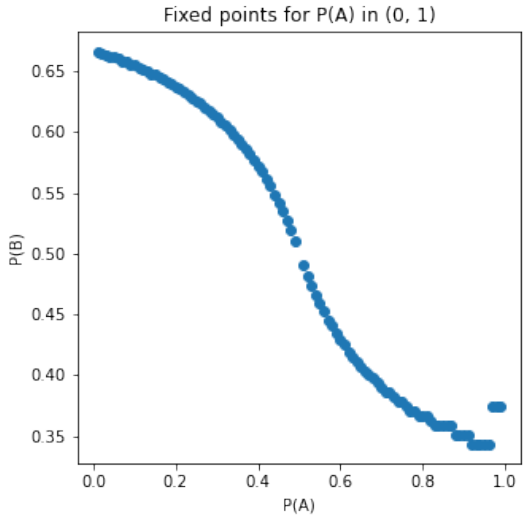


Figure 4.6: Fixed point for $\mathbf{P}_X^{A[0]}(x)$ for each $\mathbf{P}_X^{B[t]}(x)$

x is FALSE discounted by the trust.

This behavior shows that we have a fixed-point for $\mathbf{P}_X^{A[0]}(x)$ for each $\mathbf{P}_X^{B[t]}(x)$ such that $\mathbf{P}_X^{B[0]}(x) = \mathbf{P}_X^{B[t]}(x)$ when t goes to infinity, in other words, for every initial A's opinion there is an initial B's opinion such that A's and B's opinions will both eventually converge to B's initial opinion, when A and B communicate repeatedly. Fig. 4.6 shows the results from the experiments. We found the fixed-points by a binary search for each $\mathbf{P}_X^{A[0]}(x)$.

The experiments show that the fixed-point function has a curve similar to the logistic function. Each fixed-point represents the boundary between the region where two agents will radicalize and the region where they will eventually agree on some less radical point. We believe that the radicalization phenomenon happens because the agents are transmitting evidence in an *unstable* way, when the proportion of evidence x vs. $\overline{x}$ increases over time. They do not radicalize when the agents are transmitting evidence in a *stable* way, when the proportion of evidence x vs. $\overline{x}$ is the

same over time.

4.4 Update function with Alvim-Knight-Valencia update function and Subjective Logic

In the previous chapter, we showed through some examples that our update function defined in Def. 4.2 does not satisfy the desired properties. This section shows an alternative update function with elements from Alvim, Knight, and Valencia 2019.

One of our goals is to represent belief about multiple issues. We can update each state of the belief mass distribution using a definition similar to the classical belief update function from Alvim, Knight, and Valencia (2019). Let $\mathcal{A} = \{A_1, \cdots, A_n\}$ be a set of n agents. Let p be a proposition. Let $\mathrm{Bel}_p^t : A \to [0,1]$ be the belief state of an agent at the time t. Let $\mathrm{In} : \mathcal{A} \times \mathcal{A} \to [0,1]$ be the influence graph s.t. $\mathrm{In}(A_i, A_j)$ represents the influence of agent A_i on A_j, and $\mathrm{In}(A_i, A_j) = 1$ if $i = j$.

Definition 4.3 (Update function from Alvim, Knight, and Valencia (2019))**.** The belief of agent A_i given the interaction with all other agents in A at time t is defined as

$$\mathrm{Bel}_p^{t+1}(A_i) = \frac{1}{n} \sum_{A_j \in \mathcal{A}} \left(\mathrm{Bel_p^t(A_i)} + \mathrm{In(A_j, A_i)} \left(\mathrm{Bel_p^t(A_j)} - \mathrm{Bel_p^t(A_i)} \right) \right) \tag{4.20}$$

In particular, the update function for two agents A and B is defined as

$$\mathrm{Bel}_p^{t+1}(A) = \frac{\mathrm{Bel}_p^t(A)}{2} + \frac{\mathrm{Bel}_p^t(A) + \mathrm{In}(B, A)(\mathrm{Bel}_p^t(B) - \mathrm{Bel}_p^t(A)}{2} \tag{4.21}$$

The idea behind this update function is to move the belief state of A to the belief state of B proportionally by the influence that B has on A. So, the new belief state is the average between the belief state that A has with the belief that A has when influenced by B. We define a new update function with subjective logic's opinion and the Def. 4.3.

Definition 4.4 (Update function for 2 agents with constant group uncertainty)**.** Let $\mathbb{X}$ be a domain of size $k \geq 2$, X a random variable in $\mathbb{X}$, A and B agents holding opinions $\omega_X^{A[t]}$ and $\omega_X^{B[t]}$ at time t, and ω_B^A a trust opinion. Define the update function for $\omega_X^{A[t+1]}$ as

$$\omega_X^{A[t+1]} = \begin{cases} \mathbf{b}_X^{A[t+1]} &= \dfrac{\mathbf{b}_X^{A[t]}}{2} + \dfrac{\mathbf{b}_X^{A[t]} + \mathbf{P}_B^A(t_B)(\mathbf{b}_X^{B[t]} - \mathbf{b}_X^{A[t]})}{2} \\ u_X^{A[t+1]} &= \dfrac{u_X^{A[t]}}{2} + \dfrac{u_X^{A[t]} + \mathbf{P}_B^A(t_A)(u_X^{B[t]} - u_X^{A[t]})}{2} \\ \mathbf{a}_X^{A[t+1]} &= \mathbf{a}_X^{A[t]} \end{cases} \tag{4.22}$$

Example 4.7. Let $\mathbb{X} = \{x_1, x_2, x_3\}$ be the domain. Agents A and B have the following opinions and trust opinions:

$$\begin{aligned} \omega_X^{A[0]} &= ((0, 0.4, 0.6), 0, \mathbf{a}_X^{A[0]}) \qquad \mathbf{P}_B^A(t_B) = 0.5 \\ \omega_X^{B[0]} &= ((0.4, 0.4, 0.2), 0, \mathbf{a}_X^{B[0]}) \qquad \mathbf{P}_A^B(t_A) = 0.5 \end{aligned} \tag{4.23}$$

After one step, A and B have the following opinions:

$$\begin{aligned} \omega_X^{A[0]} &= ((0.1, 0.4, 0.5), 0, \mathbf{a}_X^{A[0]}) \\ \omega_X^{B[0]} &= ((0.3, 0.4, 0.3), 0, \mathbf{a}_X^{B[0]}) \end{aligned} \tag{4.24}$$

Note that, for x_1 and x_3, the belief masses does not move further from each agent

upon update. Also, $\mathbf{b}_X^{A[0]}(x_2) = \mathbf{b}_X^{B[0]}(x_2) = \mathbf{b}_X^{A[1]}(x_2) = \mathbf{b}_X^{A[1]}(x_2) = 0.4$. Because A and B have dogmatic opinions, $\mathbf{b}_X = \mathbf{P}_X$.

Theorem 4.1. *If $\mathbb{X}$ is a domain of size $k = 2$, and $\omega_t^{A[0]}$ and $\omega_t^{A[0]}$ are dogmatic opinions, then the update function for 2 agents with constant group uncertainty from Def. 4.4 corresponds to the Def. 4.3 for 2 agents.*

Proof. Since $\omega_X^{A[0]}$ (and $\omega_X^{B[0]}$) is a dogmatic opinion, $\omega_X^{A[1]}$ is defined as

$$\omega_X^{A[1]} = \begin{cases} \mathbf{b}_X^{A[1]} & = \dfrac{\mathbf{b}_X^{A[0]}}{2} + \dfrac{\mathbf{b}_X^{A[0]} + \mathbf{P}_B^A(t_B)(\mathbf{b}_X^{B[0]} - \mathbf{b}_X^{A[0]})}{2} \\ u_X^{A[1]} & = 0 \\ \mathbf{a}_X^{A[1]} & = \mathbf{a}_X^{A[0]} \end{cases} \tag{4.25}$$

Also, $\mathbf{b}_X^{A[t]} = \mathbf{P}_X^{A[t]}$ for any t.

We want to map the $\mathbf{P}_X^{A[1]}(x)$ to $\mathrm{Bel}_p^{t+1}(A)$, because we consider that A's take decisions using $\mathbf{P}_X^{A[1]}$ in subjective logic and $\mathrm{Bel}_p^{t+1}(A)$ in Alvim, Knight, and Valencia (2019).

If $x = p$, the belief mass $\mathbf{b}_X^{A[0]}$ corresponds to the belief state $\mathrm{Bel}_p^0(A)$, the belief mass $\mathbf{b}_X^{B[0]}$ corresponds to the belief state $\mathrm{Bel}_p^0(B)$, and the posterior probability $\mathbf{P}_B^A(t_B)$ from the trust opinion ω_B^A corresponds to the influence $\mathrm{In}(B, A)$, then $\mathbf{b}_X^{A[t]}(x)$ and $\mathrm{Bel}_p^{t+1}(A)$ have the same definition and $\mathbf{P}_X^{A[1]} = \mathrm{Bel}_p^{t+1}(A)$. □

It is easy to extend the update function to arbitrarily many agents, using the same Def. 4.3. It is possible to extend the update function for domains of size $k > 2$, but with no clear correspondence for belief state.

Theorem 4.2. *If $\mathbb{X}$ is a domain of size $k = 2$, the update function from Def. 4.4 satisfies the desiderata from Def. 4.1: 1. weak convergence, 2. Markovian behaviour, 3. idempotence, and 4.(a) non-increasing group uncertainty.*

Proof. 1. Weak convergence: Suppose that $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{B[t]}(x)$, then

$$\mathbf{b}_X^{A[t+1]}(x) = \mathbf{b}_X^{A[t]}(x) + \left(\frac{\mathbf{P}_B^A(t_B)(\mathbf{b}_X^{B[t]}(x) - \mathbf{b}_X^{B[t]}(x))}{2} \right). \tag{4.26}$$

Since $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{B[t]}(x)$ and $\mathbf{P}_B^A(t_B) \geq 0$, $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{A[t+1]}(x)$. Now,

$$\mathbf{b}_X^{A[t+1]}(x) = \mathbf{b}_X^{B[t]}(x) - \left(\mathbf{b}_X^{B[t]}(x) - \mathbf{b}_X^{A[t]} - \frac{\mathbf{P}_B^A(t_B)(\mathbf{b}_X^{B[t]}(x) - \mathbf{b}_X^{B[t]}(x))}{2} \right). \tag{4.27}$$

Since $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{B[t]}(x)$ and $\mathbf{P}_B^A(t_B) \leq 1$, $\mathbf{b}_X^{A[t+1]}(x) \leq \mathbf{b}_X^{B[t]}(x)$. Therefore, $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{A[t+1]}(x) \leq \mathbf{b}_X^{B[t]}(x)$. With a similar proof, we can show that $u_X^{A[t]} \leq u_X^{A[t+1]} \leq u_X^{B[t]}$.

Now, we need to show that $\mathbf{P}_X^{A[t]}(x) \leq \mathbf{P}_X^{A[t+1]}(x) \leq \mathbf{P}_X^{B[t]}(x)$. By definition of projected probability,

$$\mathbf{b}_X^{A[t]}(x) + u_X^{A[t]}\mathbf{a}_X^{A[t]}(x) \leq \mathbf{b}_X^{A[t+1]}(x) + u_X^{A[t+1]}\mathbf{a}_X^{A[t]}(x) \leq \mathbf{b}_X^{B[t]}(x) + u_X^{B[t]}\mathbf{a}_X^{B[t]}(x). \tag{4.28}$$

Since $\mathbf{b}_X^{A[t]}(x) \leq \mathbf{b}_X^{A[t+1]}(x) \leq \mathbf{b}_X^{B[t]}(x)$ and $u_X^{A[t]} \leq u_X^{A[t+1]} \leq u_X^{B[t]}$,

$$\mathbf{P}_X^{A[t]}(x) \leq \mathbf{P}_X^{A[t+1]}(x) \leq \mathbf{P}_X^{B[t]}(x). \tag{4.29}$$

The proof for $\mathbf{P}_X^{A[t]}(x) \geq \mathbf{P}_X^{A[t+1]}(x) \geq \mathbf{P}_X^{B[t]}(x)$ is similar.

2. Markovian behavior: The definition does not uses opinions from time before t. We also consider the agents does not have a history of acquired evidence over time.

3. Idempotence: Suppose that $\omega_X^{A[t]} = \omega_X^{B[t]}$, then

$$\begin{aligned} \mathbf{b}_X^{A[t+1]}(x) &= \frac{\mathbf{b}_X^{A[t]}(x) + \mathbf{b}_X^{A[t]}(x) + \mathbf{P}_B^A(\mathbf{b}_X^{A[t]}(x) - \mathbf{b}_X^{A[t]}(x))}{2} \\ &= \frac{\mathbf{b}_X^{A[t]}(x) + \mathbf{b}_X^{A[t]}(x)}{2} = \mathbf{b}_X^{A[t]}(x) \end{aligned} \tag{4.30}$$

4.(a) non-increasing group uncertainty i.e. $u_X^{A[t]} + u_X^{B[t]} = u_X^{A[t+1]} + u_X^{B[t+1]}$.

$$\begin{aligned} u_X^{A[t]} + u_X^{B[t]} &= u_X^{A[t+1]} + u_X^{B[t+1]} \\ &= \frac{u_X^{A[t]}}{2} + \frac{u_X^{A[t]} + \mathbf{P}_B^A(t_A)(u_X^{B[t]} - u_X^{A[t]})}{2} \\ &\quad + \frac{u_X^{B[t]}}{2} + \frac{u_X^{B[t]} + \mathbf{P}_A^B(t_B)(u_X^{A[t]} - u_X^{B[t]})}{2} \\ &= u_X^{A[t]} + u_X^{B[t]} \end{aligned} \tag{4.31}$$

□

To sum up, in this section we first presented the belief update function from Alvim, Knight, and Valencia (2019). Then we proposed a subjective logic belief update function, and in Theorem 4.1 we showed that our subjective logic belief update function corresponds to the belief update function from Alvim, Knight, and Valencia (2019). Next in Theorem 4.2 we proved that our new belief update function satisfies the four desiderata from Def. 4.1. This provides us with several valuable pieces of information: first, it is possible to define at least one belief update function in subjective logic satisfying the four desiderata. Second, we can define a belief update function in subjective logic that meaningfully corresponds to the belief update function from Alvim, Knight, and Valencia (2019), even though subjective logic is in fact more expressive than the formalism used in Alvim, Knight, and Valencia (2019), which does not provide for uncertainty. This shows that subjective logic is promising for modeling social networks, since it can already capture the information used in

Alvim, Knight, and Valencia (2019), and also is equipped with an additional method for reasoning about agents' uncertainty.

5 Dynamic Influence on the Alvim-Knight-Valencia Model

The Alvim-Knight-Valencia (AKV) model (Alvim, Knight, and Valencia 2019) is a model for belief update and group polarization in a social network. The AKV model defines updates function that changes only the agents' belief over time, but not the influence between agents. Aiming to represent the mutating influence between agents in social networks, we define an AKV that implements dynamic influence graphs.

This chapter is structured as follows:

- Section 5.1 – Model definition: Describe our model for social networks with dynamic influence.
- Section 5.2 – Balanced influence conservation: Using the conditions for polarization demonstrated by Alvim, Amorim, et al. (2023) on the AKV model, we define *balance-conservation*, a property of influence graphs that agents influence and are influenced by the same degree. We introduce a non-exhaustive list balance-conservative and non-balance-conservative graphs.
- Section 5.3 – Conjectures on general graphs: Based on our examples from the previous section, we devise two conjectures that, if proven to be true, it shows procedures to test the balance-conservation of influence graphs without simulations.

5.1 Model definition

The Alvim-Knight-Valencia (AKV) model (Alvim, Knight, and Valencia 2019) is a model for belief update and group polarization in a social network, including agents' initial beliefs, and an influence graph. Here we define an alternative model, called *Alvim-Knight-Valencia model with dynamic influence* where the influence between agents can also change over time given an update function.

The *static elements* represent a snapshot of a social network at a given point in time. They include the following components:

- A (finite) set

$$\mathcal{A} = \{0, 1, \cdots, n-1\} \tag{5.1}$$

  of $n \geq 1$ *agents*, representing the users of the social network.

- A proposition p representing a declarative sentence. For example, p could be the statement "*vaccines are safe*" or "*climate change is real and is caused by human activity*".

- A *belief configuration*

$$B : \mathcal{A} \to [0, 1] \tag{5.2}$$

  such that for each agent $i \in \mathcal{A}$, the value $B_i = B(i)$ represents the confidence of agent $i \in \mathcal{A}$ in the veracity of proposition p.

- A *weighted directed graph*

$$\mathcal{I} : \mathcal{A} \times \mathcal{A} \to [0, 1]. \tag{5.3}$$

  The value $\mathcal{I}(i, j)$, written $\mathcal{I}_{i,j}$, represents the *direct influence* that agent i has

on agent j, or the *weight* i carries with j.

The *dynamic elements* of the model formalize the evolution of agents' beliefs as they interact and communicate about their opinions over time. They include:

- A *time frame* $\mathcal{T} = \{0, 1, 2, \cdots\}$ representing the discrete passage of time.

- An *update function*

$$\mu : (B^t, \mathcal{I}^t) \to (B^{t+1}, \mathcal{I}^{t+1}) \tag{5.4}$$

  mapping belief configuration B^t and the influence graph I^{t+1} to a new belief configuration and influence graph at time $t + 1$. This function models the evolution of agents' beliefs about the propostion *and about each other* over time. Updating the value $I^t_{i,j}$ means that agent i is changing their belief about the proposition "I, agent i, am influenced by agent j".

- A *polarization measure*

$$\rho : [0, 1]^{\mathcal{A}} \to \mathbb{R}^+ \tag{5.5}$$

  mapping belief configurations to the non-negative real-numbers. Given a belief configuration $B = (B_0, \cdots, B_{n-1})$, the value $\rho(B)$ indicates the polarization among all the agents in $\mathcal{A}$ given their beliefs $B_0, \cdots, B_{n-1}$. Polarization can be thought of as the opposite of consensus; a high polarization value indicates disagreement between agents, while a lower polarization value represents a situation closer to consensus.

This model differs from the original AKV model by making the influence graph part of the static elements. Besides, the update function, where it did not change

the influence graph, now needs to output the update influence graph. The model with dynamic influence is equivalent to the AKV model when the update function does not change the influence graph.

5.2 Balanced influence conservation

One of the contributions of Alvim, Amorim, et al. (2023) is the conditions for polarization happen under the confirmation bias update function and the classical update function.

Theorem 5.1 (Conditions for polarization)**.** *Suppose that* $\lim_{t\to\infty} \rho(B^t) \neq 0$*. Then either:*

1. $\mathcal{I}$ *is not balanced;*

2. $\mathcal{I}$ *is not weakly connected;*

3. $B_i^0 = 0$ *or* $B_i^0 = 1$ *for every* $i \in \mathcal{A}$*; or*

4. *for some borderline value* v, $\lim_{t\to\infty} B_i^t = v$ *for each* $i \in \mathcal{A}$.

Given that our model have a dynamic influence graph, we have an influence graph $\mathcal{I}^0$ can be balanced and weakly connected, but it some point in the future, it might not be. That implies that a model can polarize on the limit and not meet any of the conditions for polarization.

In this chapter, we define an update function that uses the classic update function (Alvim, Knight, and Valencia 2019) on both the belief configuration and the influence graph and analyze models that do not (or that does) meet conditions for polarization and how it behaves over time. In particular, we study the conservation of balanced influence over time.

First, we define a balanced influence graph given a point of time:

Definition 5.1 (Balanced influence)**.** An influence graph is balanced if each agent is influenced by the same degree that influence others. Formally, given a timestep t, an influence graph is balanced if, for every $k \in \mathcal{A}$,

$$\sum_{i \in \mathcal{A}} \mathcal{I}^t_{k,i} = \sum_{i \in \mathcal{A}} \mathcal{I}^t_{i,k}. \tag{5.6}$$

A clique influence graph where all agents influence each other by the same degree is balanced, but a graph with an unrelenting influencer that is not influenced by any other agent in the network is not balanced because this influencer has a net positive influence.

Since influence graphs can change over time, we define balanced influence conservation:

Definition 5.2 (Balance-conservation)**.** An influence graph conserves its balance if it is balanced in every timestep. Formally, given a time frame $\mathcal{T} = \{0, 1, 2, \cdots\}$ and a sequence of influence graphs $\{\mathcal{I}^0, \mathcal{I}^1, \mathcal{I}^2, \cdots\}$, an influence graph $\mathcal{I}^0$ is balance-conservative if, for every $t \in \mathcal{T}$, $\mathcal{I}^t$ is balanced.

To apply this defintion we need to define how an influence graph evolves with an update function. Throught this chapter we investigate influence graphs with the following update function for AKV models with dynamic influence.

Definition 5.3 (Classic update function for models with dynamic influence)**.** Let B^t and I^t be a belief configuration and an influence graph at time $t \in \mathcal{T}$, respectively. The classic update function is the map $\mu : (B^t, \mathcal{I}^t) \to (B^{t+1}, \mathcal{I}^{t+1})$ with B^{t+1} given by

$$B^{t+1}_i = B^t_i + \frac{1}{|A_i|} \sum_{j \in \mathcal{A}_i} \mathcal{I}_{j,i}(B^t_j - B^t_i), \tag{5.7}$$

for every agent $i \in \mathcal{A}$, and $\mathcal{I}^{t+1}$ given by

$$\mathcal{I}_{k,i}^{i+1} = \begin{cases} 1, & \text{if } k = i, \\ \mathcal{I}_{k,i} + \dfrac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} \mathcal{I}_{j,i}^{t} (\mathcal{I}_{k,j}^{t} - \mathcal{I}_{k,i}^{t}), & \text{otherwise,} \end{cases} \tag{5.8}$$

for every agents $k, i \in \mathcal{A}$.

Recall that the influence $\mathcal{I}_{k,i}$ is the influence the agent k has on i. Note that we are using the classic update function, usually applied to a belief value, on each influence value. To explain the update method, we discuss the closely related notion of trust, rather than influence. When agent i decides how much they will trust agent k at the next timestep, they consider how much all their neighbors currently trust k, and how much they trust their neighbors, in order to update this value. As an example, say that Alice is deciding how much she trusts Bob. She consults her neighbors, Charlie and David, and considers how much they each trust Bob, and how much she trusts each of them, in order to decide how much she should trust Bob.

Agent i is updating the influence that agent k has on them (agent i). Agent i will use the influence that k has on i's neighbors to update this value.

Furthermore, we assume that the influence agents have on themselves is always 1, i.e. they never doubt their ability to hold a belief.

Now we investigate the balance conservation with this update function. We show examples of balanced influence graphs that conserve and does not conserve its balance. Finally, we conjecture a condition in which a balanced influence graph must meet to be balance-conservative.

5.2.1 Balance-conservative influence graphs

Clique influence graphs The first set of graphs we investigate the set of influence graphs where every agent influence each other by the same degree.

Definition 5.4 (Clique influence graph)**.** Let $i, j \in \mathcal{A}$ be agents, we define the clique influence graph as follows:

$$\mathcal{I}_{k,i} = \begin{cases} 1, & \text{if } k = i, \text{ otherwise} \\ C, & \text{if } k \neq 1 \end{cases} \tag{5.9}$$

for some constant C such that $0 \leq C \leq 1$ representing the influence between different agents. Alternatively, we can define $\mathcal{I}$ as:

$$\mathcal{I} = \begin{bmatrix} 1 & C & \cdots & C & C \\ C & 1 & \cdots & C & C \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C & C & \cdots & 1 & C \\ C & C & \cdots & C & 1 \end{bmatrix}. \tag{5.10}$$

Example 5.1. Let $\mathcal{I}^0$ be a clique influence graph with 4 agents and $C = 0.5$. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 0.56 & 0.56 & 0.56 \\ 0.56 & 1 & 0.56 & 0.56 \\ 0.56 & 0.56 & 1 & 0.56 \\ 0.56 & 0.56 & 0.56 & 1 \end{Bmatrix}. \tag{5.11}$$

.

Both influece graphs are balanced.

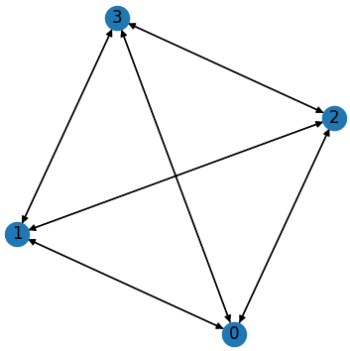


Figure 5.1: Influence graph $\mathcal{I}^0$ from Example 5.1. Self influence and influence 0 are omitted.

Now we prove that clique influece graphs are balance-conservative. First, we need to show that a clique influence graph is balanced.

Lemma 5.1. *Clique influence graphs are balanced.*

Proof. Let $\mathcal{I}$ be a clique influence graph. Given an agent $k \in \mathcal{A}$. Considering that agent k influences themselves by 1 and influences every other agent by C, the sum of the influence that k has on all agents is

$$\sum_{i \in \mathcal{A}} \mathcal{I}_{k,i} = 1 + C(|A| - 1). \tag{5.12}$$

Now, considering that agent k is influenced by themselves by 1 and is influenced by every other agent by C, the sum of the influence that all agents have on k is

$$\sum_{i \in \mathcal{A}} \mathcal{I}_{i,k} = 1 + C(|A| - 1). \tag{5.13}$$

Therefore, the clique influence graphs are balanced. □

Then, we show that updating a clique influence graph generates a clique influence graph.

Lemma 5.2. *Let $\mathcal{I}^t$ be a clique influence graph. Then, after one update with the classic update function for models with dynamic influence (Definition 5.3), $\mathcal{I}^{t+1}$ is also a clique influence graph.*

Proof. We know that if $k = i$, then $\mathcal{I}^{t+1}_{k,i} = 1$ for every $k, i \in \mathcal{A}$. To show that $\mathcal{I}^{t+1}$ is also a clique influence graph, we need to show that there exists a constant C' such that $\mathcal{I}^{t+1}_{k.i} = C'$ if $k \neq i$ for every $k, i \in \mathcal{A}$.

If $k \neq i$, then:

$$\begin{aligned}
\mathcal{I}^{t+1}_{k,i} &= \mathcal{I}^t_{k,i} + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}^t_{j,i}(\mathcal{I}^t_{k,j} - \mathcal{I}^t_{k,i}) \\
&= C + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}^t_{j,i}(\mathcal{I}^t_{k,j} - C) \\
&= C + \tfrac{1}{|\mathcal{A}|}\left(1\cdot(C-C) + \textstyle\sum_{\substack{j\in\mathcal{A}\\ j\neq i}} C\cdot(\mathcal{I}^t_{k,j} - C)\right) \\
&= C + \tfrac{1}{|\mathcal{A}|}\left(C\cdot(1-C) + \textstyle\sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j\neq k}} C\cdot(C-C)\right) \\
&= C + \frac{1}{|\mathcal{A}|}C(1-C) = C'.
\end{aligned} \tag{5.14}$$

Thus, we can define $\mathcal{I}^{t+1}$ as

$$\mathcal{I}^{t+1}k, i = \begin{cases} 1, & \text{if } i = j, \\ C + \dfrac{1}{|\mathcal{A}|}C(1-C), & \text{if } i \neq j. \end{cases} \tag{5.15}$$

Therefore, $\mathcal{I}^{t+1}$ is a clique influence graph. □

Finally, we use those lemmas to show that clique influence graphs are balance-conservative. Recall that a influence graph is balance-conservative if the influence graph is balanced in every timestep.

Theorem 5.2. *Clique influence graphs are balance-conservative under the classic update function.*

Proof. The Lemma 5.2 shows that a clique influence graph always generate a clique influence graph after update with the classic update function. And the Lemma 5.1 shows that every clique influence graph is balanced. Therefore, clique influence graphs are balance-conservative. □

Bipartite influence graph with equal-sized groups The second class of graphs is the set of influence graphs where agents are separeted in two equal-sized groups. In the following definition we set the influence degree C as the influence between agents of different groups and C' as the influence between agents of the same group. This graph is similar to the faintly communicating influence graph from Alvim, Knight, and Valencia (2019).

Definition 5.5 (Bipartite groups influence graph with equal-sized)**.** Let $i, j \in \mathcal{A}$ be agents. Let $2n = |A|$. We define the bipartite influence graph with equal-sized groups as follows:

$$\mathcal{I}_{k,1} = \begin{cases} 1, & \text{if } k = i, \text{ otherwise} \\ C, & \text{if } k < n \text{ and } i \geq n \text{ or } k \geq n \text{ and } i < n \\ C' & \text{if } k < n \text{ and } i < n \text{ or } k \geq n \text{ and } i \geq n \end{cases} \tag{5.16}$$

for some constants C, C' such that $0 \leq C \leq 1$ representing the influence between agents of different groups and from the same group, respectively. Alternatively, we can define $\mathcal{I}$ as:

$$\mathcal{I} = \begin{bmatrix} 1 & C' & \cdots & C & C \\ C' & 1 & \cdots & C & C \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C & C & \cdots & 1 & C' \\ C & C & \cdots & C' & 1 \end{bmatrix}. \tag{5.17}$$

Example 5.2. Let $\mathcal{I}$ be a bipartite with equal-sized groups influence graph with 4 agents, $C = 0.5$, and $C = 0$. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0 & 0.3 & 0.3 \\ 0 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 0.04 & 0.33 & 0.33 \\ 0.04 & 1 & 0.33 & 0.33 \\ 0.33 & 0.33 & 1 & 0.04 \\ 0.33 & 0.33 & 0.04 & 1 \end{Bmatrix}. \tag{5.18}$$

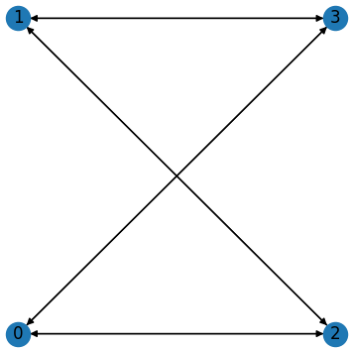


Figure 5.2: Influence graph $\mathcal{I}^0$ from Example 5.2. Self influence and influence 0 are omitted.

Both influence graphs are balanced.

As we demonstrated for clique influence graphs, we prove that bipartite influence graphs with equal-sized groups are balanced.

Lemma 5.3. *Bipartite with equal-sized groups influence graphs are balanced.*

Proof. Let $\mathcal{I}$ be a bipartite influence graph with equal-sized groups. Given an agent $k \in \mathcal{A}$, considerint that agent k influences themselves by 1, influences agents of the other group by C, and influences agents of the same group by C', the sum of the influence that k has on all agents is

$$\sum_{i \in \mathcal{A}} \mathcal{I}_{k,i} = 1 + C\frac{|\mathcal{A}|}{2} + C'\left(\frac{|\mathcal{A}|}{2} - 1\right). \tag{5.19}$$

Now, considering that agent k is influenced by themselves by 1, is influenced by agents of the other group by C, and is influenced by agents of the same group by C', the sum of the influence that all agents have on k is

$$\sum_{i \in \mathcal{A}} \mathcal{I}_{i,k} = 1 + C\frac{|\mathcal{A}|}{2} + C'\left(\frac{|\mathcal{A}|}{2} - 1\right). \tag{5.20}$$

Therefore, the bipartite influence graphs with equal-sized groups are balanced. □

Then, we show that updating a bipartite influence graph with equal-sized groups generates a biparite influence graph.

Lemma 5.4. *Let $\mathcal{I}^t$ be a bipartite groups influence graph with equal-sized. Then, after one update with the classic update function for models with dynamic influence (Definition 5.3), $\mathcal{I}^{t+1}$ is also a bipartite groups influence graph with equal-sized.*

Proof. We know that if $k = i$, then $\mathcal{I}^{t+1}_{k,i} = 1$ for every $k, i \in \mathcal{A}$. To show that $\mathcal{I}^{t+1}$ is also a bipartite groups influence graph with equal-sized, we need to show that there exist constants C, C' such that $\mathcal{I}^{t+1}$ bipartite influence graph with equal-sized groups.

If $k < n$ and $i \geq n$, then:

$$\begin{aligned}
\mathcal{I}_{k,i}^{t+1} &= \mathcal{I}_{k,i}^{t} + \tfrac{1}{|\mathcal{A}|} \textstyle\sum_{j \in \mathcal{A}} \mathcal{I}_{j,i}^{t} (\mathcal{I}_{k,j}^{t} - \mathcal{I}_{k,i}^{t}) \\
&= C + \tfrac{1}{|\mathcal{A}|} \textstyle\sum_{j \in \mathcal{A}} \mathcal{I}_{j,i}^{t} (\mathcal{I}_{k,j}^{i} + C) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(1 \cdot (C - C) + \textstyle\sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j < n}} C(\mathcal{I}_{k,j}^{t} - C) + \sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j \geq n}} C'(C - C) \right) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(C(1 - C) + \textstyle\sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j < k \\ j \neq k}} C(C' - C) \right) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(C(1 - C) + \left(\tfrac{|\mathcal{A}|}{2} - 1 \right) (C(C' - C)) \right).
\end{aligned} \tag{5.21}$$

If $k \geq n$ and $i < n$, then:

$$\begin{aligned}
\mathcal{I}_{k,i}^{t+1} &= \mathcal{I}_{k,i}^{t} + \tfrac{1}{|\mathcal{A}|} \textstyle\sum_{j \in \mathcal{A}} \mathcal{I}_{j,i}^{t} (\mathcal{I}_{k,j}^{t} - \mathcal{I}_{k,i}^{t}) \\
&= C + \tfrac{1}{|\mathcal{A}|} \textstyle\sum_{j \in \mathcal{A}} \mathcal{I}_{j,i}^{t} (\mathcal{I}_{k,j}^{t} + C) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(1 \cdot (C - C) + \textstyle\sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j < n}} C'(C - C) + \sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j \geq n}} C(\mathcal{I}_{k,j}^{t} - C) \right) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(C(1 - C) + \textstyle\sum_{\substack{j \in \mathcal{A} \\ j \neq i \\ j \geq k \\ j \neq k}} C(C' - C) \right) \\
&= C + \tfrac{1}{|\mathcal{A}|} \left(C(1 - C) + \left(\tfrac{|\mathcal{A}|}{2} - 1 \right) C(C' - C) \right).
\end{aligned} \tag{5.22}$$

If $k < n$ and $i < n$, then:

$$\begin{aligned}
\mathcal{I}_{k,i}^{t+1} &= \mathcal{I}_{k,i}^{t} + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}_{j,i}^{t}(\mathcal{I}_{k,j}^{t} - \mathcal{I}_{k,i}^{t}) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}_{j,i}^{t}(\mathcal{I}_{k,j}^{t} + C') \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(1\cdot(C'-C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j<n}} C'(\mathcal{I}_{k,j}^{t} - C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j\geq n}} C(C-C')\right) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(C'(1-C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j<n\\ j\neq k}} C'(C'-C') + \tfrac{|\mathcal{A}|}{2}C(C-C')\right) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(C'(1-C') + \tfrac{|\mathcal{A}|}{2}C(C-C')\right).
\end{aligned} \tag{5.23}$$

If $k \geq n$ and $i \geq n$, then:

$$\begin{aligned}
\mathcal{I}_{k,i}^{t+1} &= \mathcal{I}_{k,i}^{t} + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}_{j,i}^{t}(\mathcal{I}_{k,j}^{t} - \mathcal{I}_{k,i}^{t}) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\textstyle\sum_{j\in\mathcal{A}} \mathcal{I}_{j,i}^{t}(\mathcal{I}_{k,j}^{t} + C') \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(1\cdot(C'-C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j<n}} C'(C-C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j\geq n}} C(\mathcal{I}_{k,j}^{t} - C')\right) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(\tfrac{|\mathcal{A}|}{2}C(C-C') + C'(1-C') + \sum_{\substack{j\in\mathcal{A}\\ j\neq i\\ j\geq n\\ j\neq k}} C'(C'-C')\right) \\
&= C' + \tfrac{1}{|\mathcal{A}|}\left(C'(1-C') + \tfrac{|\mathcal{A}|}{2}C(C-C')\right).
\end{aligned} \tag{5.24}$$

Thus, we can define $\mathcal{I}^{t+1}$ as

$$\mathcal{I}_{k,1} = \begin{cases} 1, & \text{if } k = i, \text{ otherwise} \\ C + \frac{1}{|\mathcal{A}|}\left(C(1-C) + \left(\frac{|\mathcal{A}|}{2} - 1\right) C(C' - C)\right), & \text{if } k < n \text{ and } i \geq n \text{ or} \\ & \quad k \geq n \text{ and } i < n \\ C' + \frac{1}{|\mathcal{A}|}\left(C'(1-C') + \frac{|\mathcal{A}|}{2} C(C - C')\right) & \text{if } k < n \text{ and } i < n \text{ or} \\ & \quad k \geq n \text{ and } i \geq n \end{cases} \tag{5.25}$$

□

Finally, we use those lemmas to show that bipartite influence graphs with equal-sized groups are balance-conservative. Recall that a influence graph is balance-conservative if the influence graph is balanced in every timestep.

Theorem 5.3. *Bipartite influence graphs with equal-sized groups are balanced influence conservative under the classic update function.*

Proof. The Lemma 5.4 shows that a clique influence graph always generate a clique influence graph after update with the classic update function. And the Lemma 5.3 shows that every clique influence graph is balanced. Therefore, clique influence graphs are balance-conservative. □

In this section, we showed that clique and bipartite influence graphs with equal-sized groups are balanced and balance-conservative. Next, we show some graphs that are not balance-conservative.

5.2.2 Non-balanced-conservative influence graphs

In this section we show some graphs that are balanced, but are not balance-conservative.

Star influence graphs Star influence graphs are graphs that have a central agent conected to every other agent by the same influence degree, but the other agents are note connected to each other. Star graphs are denoted S_n or the complete bipartite graph $K_{1,n}$ where n is the number of agents connected to the central agent.

Example 5.3. Let $\mathcal{I}^0$ be a star influence graph with 4 agents, and $\mathcal{I}^1$ be the influence graph generated from $\mathcal{I}^0$ with the classic update function. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are:

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 0.56 & 0.56 & 0.56 \\ 0.44 & 1 & 0.06 & 0.06 \\ 0.44 & 0.06 & 1 & 0.06 \\ 0.44 & 0.06 & 0.06 & 1 \end{Bmatrix}. \tag{5.26}$$

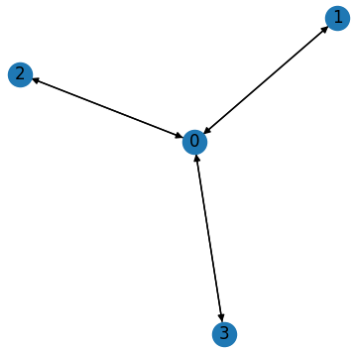


Figure 5.3: Influence graph $\mathcal{I}^0$ from Example 5.3. Self influence and influence 0 are omitted.

$\mathcal{I}^0$ is balanced. On $\mathcal{I}^1$, agent 0, the center of the graph, is influencing more than is being influenced. Therefore $\mathcal{I}^1$, is not balanced and this star influence graph is not balance-conservative. We conjecture that star graphs S_n for any $n \geq 2$ are not balance-conservative.

Wheel influence graphs Wheel influence graphs also have a central agent, but all other agents are connected in a cycle. Wheel graphs are denoted W_n where n

is the total number of agents. A wheel influence graph with 4 agents is a clique influence graph. Here we consider graphs with 5 or more agents.

Example 5.4. Let $\mathcal{I}^0$ be a wheel influence graph with 5 agents, and $\mathcal{I}^1$ be the influence graph generated from $\mathcal{I}^0$ with the classic update function. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are:

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0.5 & 0 & 0.5 \\ 1 & 0.5 & 1 & 0.5 & 0 \\ 1 & 0 & 0.5 & 1 & 0.5 \\ 1 & 0.5 & 0 & 0.5 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.6 & 1 & 0.6 & 0.3 & 0.6 \\ 0.6 & 0.6 & 1 & 0.6 & 0.3 \\ 0.6 & 0.3 & 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.3 & 0.6 & 1 \end{Bmatrix}. \tag{5.27}$$

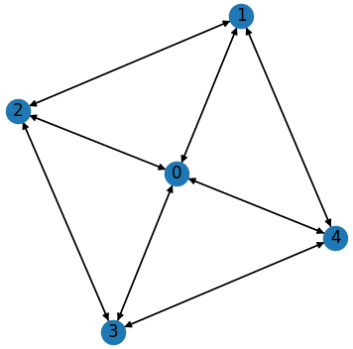


Figure 5.4: Influence graph $\mathcal{I}^0$ from Example 5.4. Self influence and influence 0 are omitted.

$\mathcal{I}^0$ is balanced. On $\mathcal{I}^1$, agent 0, the center of the graph, is influencing more than is being influenced. Therefore $\mathcal{I}^1$, is not balanced and this wheel influence graph is not balance-conservative. We conjecture that wheel graphs W_n for any $n \geq 5$ are not balance-conservative.

Friendship influence graphs Friendship influence graphs are graphs that have a central agent that is also a common agent to copies of cycle graphs of 3 agents. Star

graphs are denoted F_n where n is the number of cycles.

Example 5.5. Let $\mathcal{I}^0$ be a friendship influence graph with 5 agents, and $\mathcal{I}^1$ be the influence graph generated from $\mathcal{I}^0$ with the classic update function. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are:

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 1 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 0.55 & 0.55 & 0.55 & 0.55 \\ 0.45 & 1 & 0.55 & 0.05 & 0.55 \\ 0.45 & 0.55 & 1 & 0.05 & 0.05 \\ 0.45 & 0.05 & 0.05 & 1 & 0.55 \\ 0.45 & 0.05 & 0.05 & 0.55 & 1 \end{Bmatrix}. \tag{5.28}$$

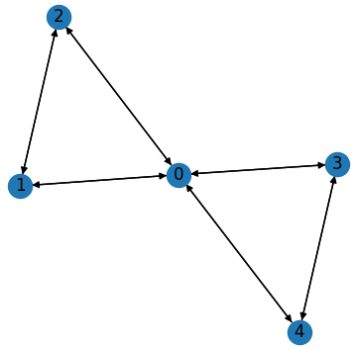


Figure 5.5: Influence graph $\mathcal{I}^0$ from Example 5.5. Self influence and influence 0 are omitted.

$\mathcal{I}^0$ is balanced. On $\mathcal{I}^1$, agent 0, the center of the graph, is influencing more than is being influenced. Therefore $\mathcal{I}^1$, is not balanced and this friendship influence graph is not balance-conservative. We conjecture that friendship graphs F_n for any $n \geq 2$ are not balance-conservative.

Bipartite influence graph with different-sized groups Bipartite influence graphs with different sized-groups do not have a central agent, but have a group trans-

miting more information than the other. Bipartited graphs with different-groups are denoted $K_{n,m}$ where n and m are the number of agents on each group and $n \neq m$.

Example 5.6. Let $\mathcal{I}^0$ be a bipartite influence graph with different-sized groups with 5 agents, and $\mathcal{I}^1$ be the influence graph generated from $\mathcal{I}^0$ with the classic update function. Then $\mathcal{I}^0$ and $\mathcal{I}^1$ are:

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0.5 & 0.5 \\ 0 & 0 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 0 \\ 0.5 & 0.5 & 0.5 & 0 & 1 \end{Bmatrix} \text{ and } \mathcal{I}^1 = \begin{Bmatrix} 1 & 0.1 & 0.1 & 0.45 & 0.45 \\ 0.1 & 1 & 0.1 & 0.45 & 0.45 \\ 0.1 & 0.1 & 1 & 0.45 & 0.45 \\ 0.5 & 0.5 & 0.5 & 1 & 0.15 \\ 0.5 & 0.5 & 0.5 & 0.15 & 1 \end{Bmatrix}. \quad (5.29)$$

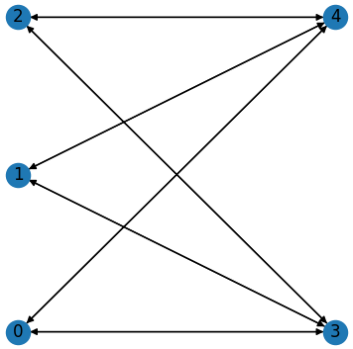


Figure 5.6: Influence graph $\mathcal{I}^0$ from Example 5.6. Self influence and influence 0 are omitted.

$\mathcal{I}^0$ is balanced. On $\mathcal{I}^1$, agents 3 and 4, the *smaller group* is influencing more than being influenced. We conjecure that every bipartite influence graph with different-sized groups are not balance-conservative.

5.3 Conjectures on general influence graphs

In this section we show that an example of a balanced influence graph with a induced non-balanced-conservative subgraph.

Example 5.7. Let $\mathcal{I}^0$ be the influence graph described below, and $\mathcal{I}^1$ be the influence graph generated from $\mathcal{I}^0$ with the classic update function.

$$\mathcal{I}^0 = \begin{Bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 1 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0.5 & 1 & 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0 & 0 & 1 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0.5 & 1 \end{Bmatrix} \tag{5.30}$$

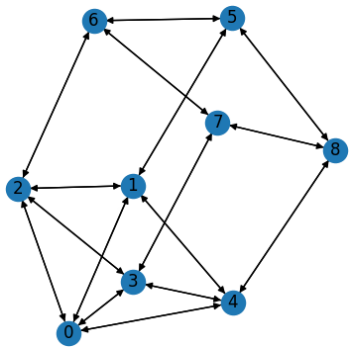


Figure 5.7: Influence graph $\mathcal{I}^0$ from Example 5.7. Self influence and influence 0 are omitted.

This graph has a induced wheel subgraph formed by the agents 0 to 4. On $\mathcal{I}^1$,

agent 0, the center of the graph, is influencing more than is being influenced. Therefore $\mathcal{I}^1$, is not balanced and this wheel influence graph is not balance-conservative.

This example leads us to to the following conjecture:

Conjecture 5.1. *Every balanced influence graph with a non-balanced-conservative induced subgraph is non-balance-conservative.*

If this and the previous conjectures are proven to be true, then searches for star, wheel, friendship or different group sized bipartite induced subgraphs on balanced influence graphs can show than a influence graph in balanced non-conservative.

In every example of non-balanced-conservative influence graphs presented in this chapter, there is at least one agent such that the sum of the influence to other agents is not equal to some other agent. This leads us to this more general conjecture:

Conjecture 5.2. *Let $\mathcal{I}$ be a balanced influence graph. $\mathcal{I}$ is balance-conservative iff every agent has the same sum of influence to the other agents.*

If this conjecture is proven to be true, then, not only it would be possible to show whether any influence graph is balance-conservative, but it also would be a less complex procedure than searching for specific induced subgraphs.

6 A Tool for Social Network Simulations in the Alvim-Knight-Valencia Model

Given the use and continuous research on the AKV model (Alvim, Knight, and Valencia 2019), we develop a Python tool that implements the model and provides a catalog of initial configurations, influence graphs, and belief update functions already developed in the literature. This tool facilitates future research using the AKV model without the need to reimplement it and also allowing its reproducibility. The tool is publicly available as `akvmodel` on GitHub and on the Python Package Index (PyPI) under the MIT Licence allowing researchers to freely use it and contribute.

This chapter is structured as follows:

- Section 6.1 – The tool: Explains the tool, showing the specification of the `AKV` class and the catalogs.
- Section 6.2 – Examples: We show running examples of models simulated using our tool and the correspondence of our simulations to the work of Alvim, Amorim, et al. (2023).

6.1 The tool

We developed a Python tool that implements the AKV model and a useful tool of update functions and initial configurations. Our tools can use the update function to generate the belief states over time so the user can analyze the outcomes. In our tool, the user can define their own initial state, influence graph, and update function, or use the readymade ones we provide as useful test cases.

In this section, we describe the tool, the class `AKV` for instantiating the model, and the catalogs `UpdateFunctions`, `InitialConfigurations`, and `InfluenceGraphs`. The source code is available at our repository on GitHub[1] with its documentation[2].

6.1.1 `AKV` class

Aiming to work on the modeling belief about multiple issues (Alvim, Knight, and Valencia 2019), we expand this model to having a multi-valued belief for each agent.

Definition 6.1. *(Domain)* Let p be a proposition. Define

$$\mathcal{X} = \{0, 1, \cdots, n-1\} \tag{6.1}$$

as the set of possible outcomes (or truth values) for the proposition p.

With this domain, we redefine a belief configuration as

$$\mathcal{B} : \mathcal{A} \times \mathcal{X} \to [0, 1], \tag{6.2}$$

such that for each agent $i \in \mathcal{A}$ and outcome $x \in \mathcal{X}$, the value $B_i, x = B(i, x)$ repre-

[1] Repository: https://github.com/josecoliveira/akvmodel/tree/v1.2.3

[2] Documentation: https://github.com/josecoliveira/akvmodel/blob/v1.2.3/DOCUMENTATION.md

sents the confidence of agent $i \in \mathcal{A}$ that the outcome $x \in \mathcal{X}$ is true in the proposition p. We also allow agents to be *uncertain*, i.e. $\sum_{x \in \mathcal{X}} B_i, x < 1$, or *inconsistent*, i.e. $\sum_{x \in \mathcal{X}} B_i, x > 1$.

```python
class AKV:
    belief_state: list[list[float]]
    influence_graph: list[list[float]]
    update_function: Callable[
        [list[list[float]], list[list[float]]], list[list[float]]
    ]
    number_of_agents: int
    domain_size: int
    states: list[list[list[float]]]

    def __init__(
        self,
        belief_state: list[list[float]],
        influence_graph: list[list[float]],
        update_function: Callable[
            [list[list[float]], list[list[float]]], list[list[float
                ]]
        ],
    )
    def update(self) -> list[list[float]]
    def get_polarization(
        self, k: int = 201, K: float = 1000, alpha: float: = 1.0
    ) -> list[list[float]]
```

Listing 6.1: `AKV` class' attributes, constructor, and methods.

The class `AKV`, as its constructor is shown in Listing 6.1, takes the following parameters:

- `belief_state`: the initial belief state defined as a matrix $|\mathcal{A}| \times |\mathcal{X}|$ where each line represents a belief configuration given an outcome of $\mathcal{X}$.
- `influence_graph`: an adjacency matrix $|\mathcal{A}| \times |\mathcal{A}|$.
- `update_function` a function that takes a belief state and an influence graph.

The class has the following attributes:

- `belief_state`: the current belief state defined as an $|\mathcal{A}| \times |\mathcal{X}|$ matrix where each line represents a belief configuration given an outcome of $\mathcal{X}$.
- `influence_graph`: an $|\mathcal{A}| \times |\mathcal{A}|$ adjacency matrix.
- `number_of_agents`: $|\mathcal{A}|$.
- `domain_size`: $|\mathcal{X}|$.
- `states`: The list of states generated by the update function. It is a three-dimensional $(t+1) \times |\mathcal{A}| \times |\mathcal{X}|$ where t is the number of times the state was updated.

Finally, the class has the following methods:

- `update`: use `update_function` to compute the next state, add it to `states` and return it.
- `get_polarization`: return a list with the polarization of computed every state and for every domain. Therefore, it is a $t \times |\mathcal{X}|$ matrix. It takes the parameters `k` for the numbers of bins, and `K` and `alpha` for the parameters of Esteban-Ray polarization measure K and α.

6.1.2 `UpdateFunction` class

The `UpdateFunctions` class has only static functions to be used as parameters of `AKV` instances. We provide two built-in update functions which may be of interest to the user, or they can define their own. Both the following functions take a belief state and an influence graph. Listing 6.2 shows the functions' signatures.

```
1 class UpdateFunctions:
2     def classic(
3         belief_state: list[list[float]], influence_graph: list[list[
              float]]
4     ) -> list[list[float]]:
5     def confirmation_bias(
6         belief_state: list[list[float]], influence_graph: list[list[
              float]]
7     ) -> list[list[float]]
```

Listing 6.2: `UpdateFunctions` class for the catalog of update functions.

The followin updated functions implemented in the class come from the reviewed literature in this paper:

- `classic` is the implementation of the classic update function (authority non-confirmatory bias).
- `confirmation_bias` is the implementation of the confirmation-bias update function.

6.1.3 `InitialConfigurations` class

In the package, we include several representative initial configurations representing a range of interesting starting situations. The user can also define their own

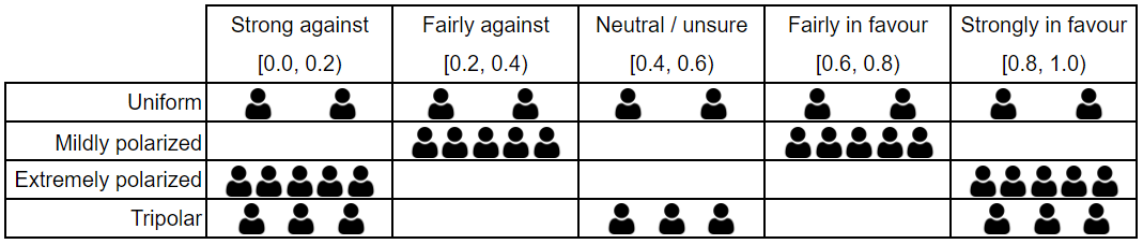


Figure 6.1: Initial belief configurations with 10 agents. (Alvim, Amorim, et al. 2023).

as desired. The `InitialConfigurations` class has only static functions that generate a belief state given the number of agents. It can be used to generate the initial belief configuration to instantiate an `AKV` object. Listing 6.3 shows the functions' signatures.

```
1 class InitialConfigurations:
2     def uniform(num_agents: int) -> list[list[float]]
3     def mildly(num_agents: int) -> list[list[float]]
4     def extreme(num_agents: int) -> list[list[float]]
5     def tripolar(num_agents: int) -> list[list[float]]
```

Listing 6.3: `InitialConfigurations` class for the catalog of initial configurations.

The following initial configurations implemented in the class come from the reviewed literature in this paper:

- `uniform` is the uniform belief configuration.
- `mildly` is the mildly polarized belief configuration.
- `extreme` is the extremely polarized belief configuration.
- `tripolar` is the tripolar belief configuration.

6.1.4 `InfluenceGraphs` class

As a starting point for the user, we define the built-in influence graphs that are a generalization of the influence graphs considered in Alvim, Knight, and Valencia (2019). The `InfluenceGraphs` class has only static functions that generate an influence graph. The parameters are different depending on the definition of the influence graph. The class can be used to generate the influence graph to instantiate an `AKV` object. Listing 6.4 shows the functions' signatures.

```
1  class InfluenceGraphs:
2      def clique(
3          num_agents: int, influence: float = 1
4      ) -> List[List[float]]
5      def circular(
6          num_agents: int, influence: float = 0.5
7      ) -> List[List[float]]
8      def faintly(
9          num_agents: int,
10         strong_influence: float = 0.5,
11         weak_influence: float = 0.1
12     ) -> List[List[float]]
13     def disconnected(
14         num_agents: int, influence: float = 0.5
15     ) -> List[List[float]]
16     def malleable_influencers(
17         num_agents: int,
18         influencer_1_influence: float = 0.8,
19         influencer_2_influence: float = 0.4,
20         influence_on_influencer_1: float = 0.1,
21         influence_on_influencer_2: float = 0.1,
22         other_agents_influence: float = 0.1,
```

```
    ) -> list[list[float]]
    def unrelenting_influencers(
        num_agents: int,
        influencer_1_influence: float = 0.6,
        influencer_2_influence: float = 0.6,
        other_agents_influence: float = 0.1,
    ) -> list[list[float]]
```

Listing 6.4: `InfluenceGraphs` class for the catalog of influence graphs.

The following influence graphs implemented in the class come from the reviewed literature in this paper:

- `clique` is the function for a clique influence graph. It takes the number of agents `num_agents` and the influence `influence` that every agent has on each other.
- `circular` is the function for a circular influence graph. It takes the number of agents `num_agents` and the influence `influence` that one agent has on the next one.
- `faintly` is the function for a faintly communicating influence graph. It takes the number of agents `num_agents`, the in-group influence `strong_influence`, and the out-group influence `weak_influence`.
- `disconnected` is the function for a disconnected influence graph. It takes the number of agents `num_agents` and the in-group influence `influence`.
- `malleable_influencers` is the function for a malleable influencers influence graph. It takes:
  - the number of agents `num_agents`,

  - the agent 0 influence towards the agents $1, \cdots, n-2$ `influencer_1_influence`,
  - the agent $n-1$ influence towards the agents $1, \cdots, n-2$ `influencer_2_influence`,
  - the agents $1, \cdots, n-2$ influence towards the agent 0 `influence_on_influencer_1`,
  - the agents $n-1, \cdots, n-2$ influence towards the agent 0 `influence_on_influencer_2`, and
  - the agents $n-1, \cdots, n-2$ influence on each other `other_agents_influence`.

- `unrelenting_influencers` is the function for an unrelenting influencers influence graph. It takes:
  - the number of agents `num_agents`,
  - the agent 0 influence towards the agents $1, \cdots, n-2$ `influencer_1_influence`,
  - the agent $n-1$ influence towards the agents $1, \cdots, n-2$ `influencer_2_influence`, and
  - the agents $n-1, \cdots, n-2$ influence on each other `other_agents_influence`.

6.2 Examples and validation

This section shows an working example using an the tool. The Jupyter Notebook file with this experiment can be found in the tool's repository[3].

Let an AKV model have 100 agents and a time frame of 100 steps. The initial configuration is mildly polarized, with a faintly connected influence graph, using the confirmation bias update function. To set these parameters, we call the constructor for `akvmodel` using the functions from the catalogs. Listing 6.5 shows an instantiation for the AKV model. We omit arguments for the initial configuration and use the default arguments set as in Alvim, Knight, and Valencia (2019).

```
1 akvmodel = AKV(
2     belief_state=InitialConfigurations.mildly(NUM_AGENTS),
3     influence_graph=InfluenceGraphs.faintly(NUM_AGENTS),
4     update_function=UpdateFunctions.confirmation_bias,
5 )
```

Listing 6.5: `akvmodel` instance with mildly polarized initial configuration, faintly communicating influence graph and confirmation bias update function.

Then, we use `AKV.update` to run the update function for 100 times, as in Listing 6.6.

```
1 for _ in range(NUM_STEPS):
2     akvmodel.update()
```

Listing 6.6: Updating the model for 100 steps.

After updating, we use `AKV.get_polarization` to get the polarization for every step and outcome in the domain, as in Listing 6.7. We omit arguments for the polarization and use the default arguments set as in Alvim, Knight, and Valencia (2019).

[3] https://github.com/josecoliveira/akvmodel/blob/b033807/example.ipynb

```
1 import matplotlib.pyplot as plt
2 p = akvmodel.get_polarization()
3 plt.plot(p[0])
```

Listing 6.7: Get the polarization from an updated model and plot it.

With `p`, we plotted the evolution of the polarization, shown in Figure 6.2.

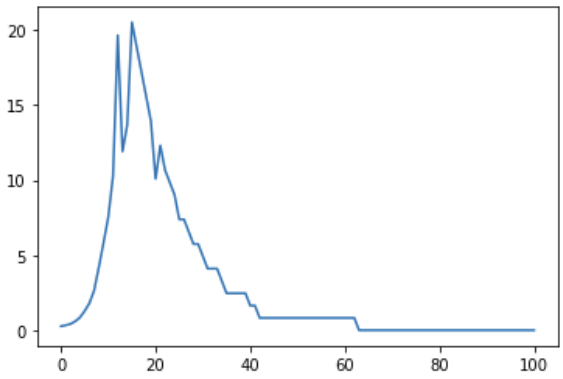


Figure 6.2: Evolution of the polarization over time

On Figures 6.3 and 6.4, we show the simulations for every combination of influence graph and initial configurations. In Alvim, Amorim, et al. (2021) and Alvim, Amorim, et al. (2023), many simulations were made with AKV models under confirmation bias. Those simulations correspond to the simulations made by our tool as shown on Figure 6.4.

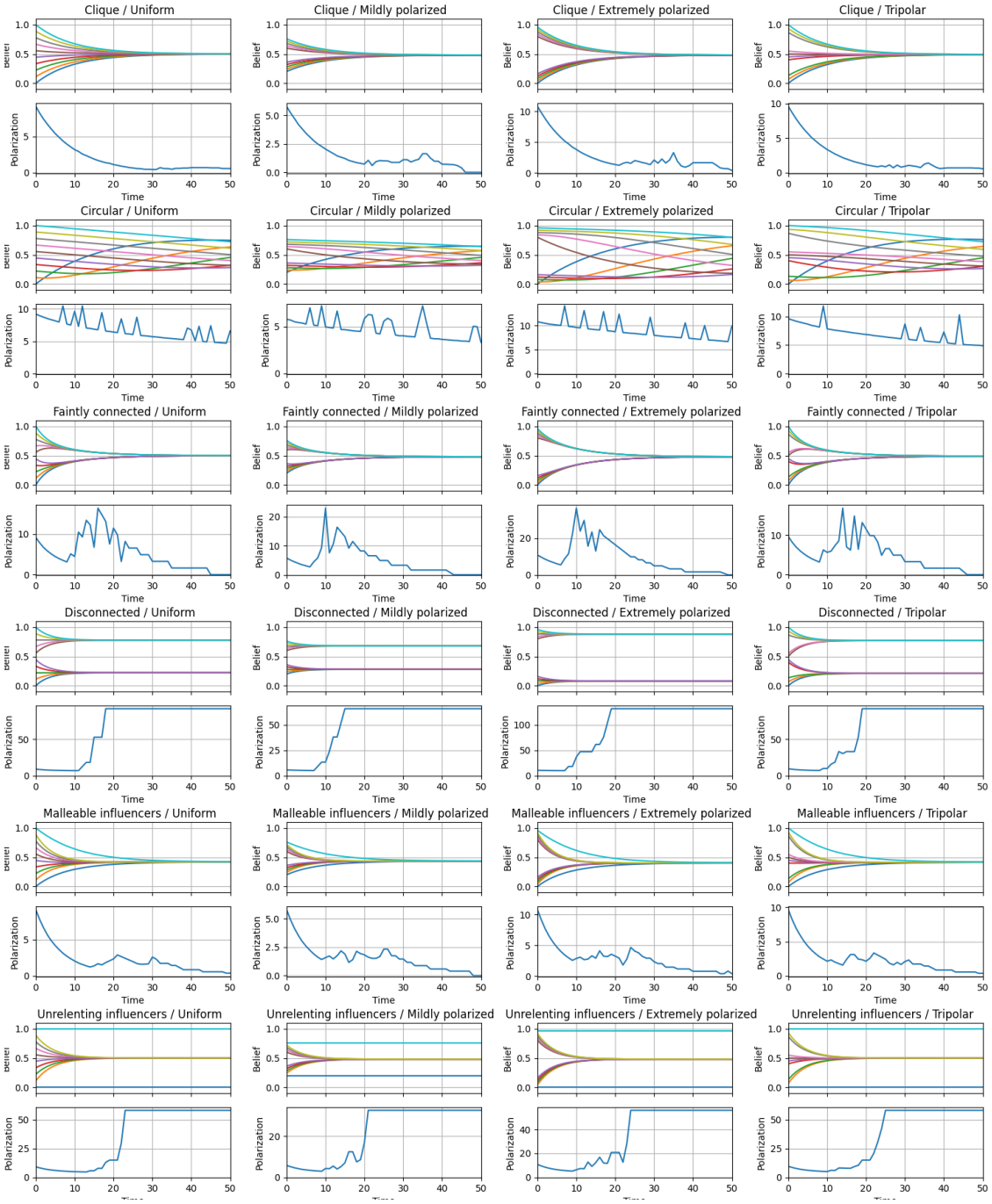


Figure 6.3: Experiments with all combinations of influence graphs and initial configurations under classic update function. Each experiment has 10 agents and ran for 50 timesteps.

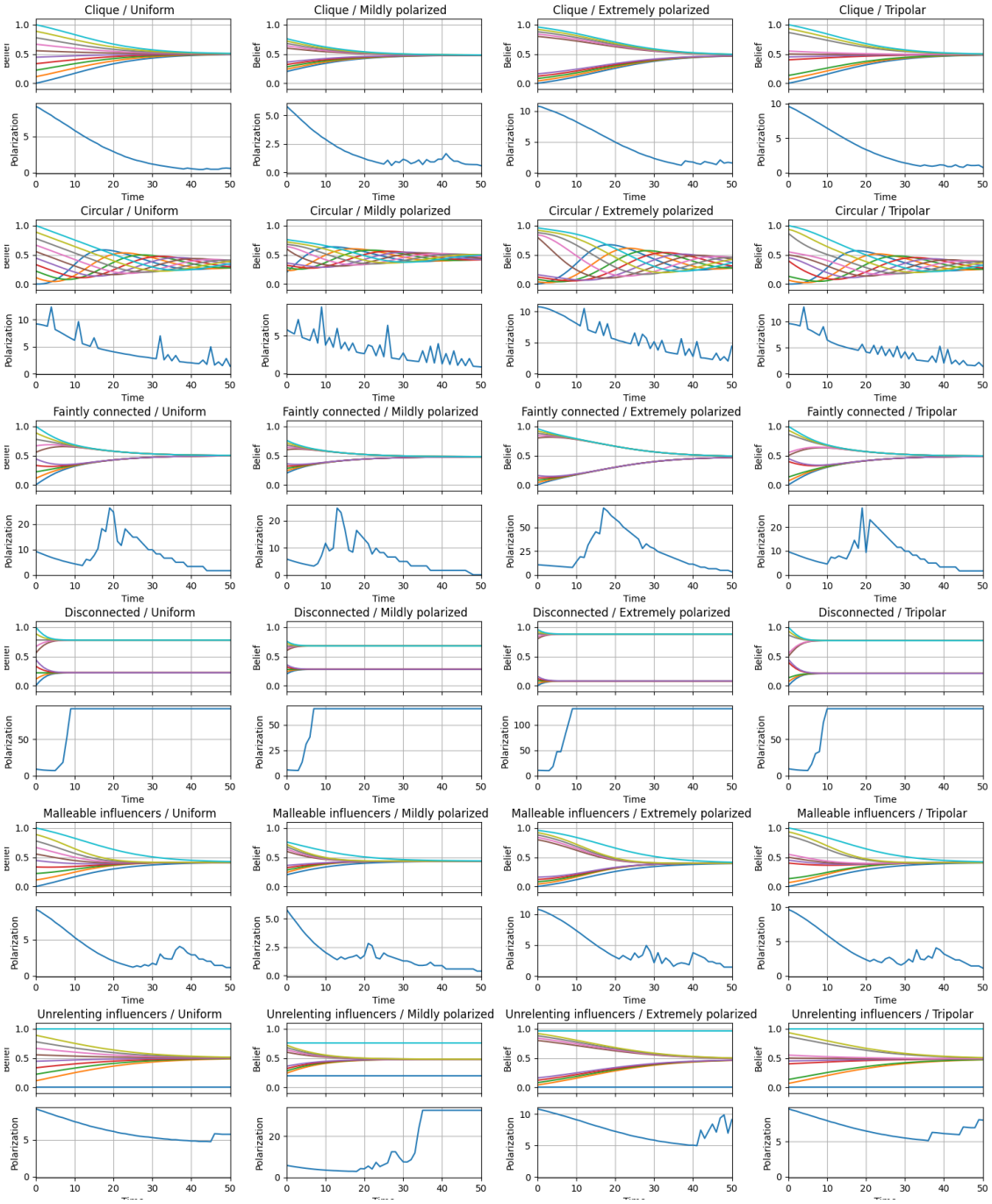


Figure 6.4: Experiments with all combinations of influence graphs and initial configurations under classic update function. Each experiment has 10 agents and ran for 50 timesteps.

7 A Tool for Paraconsistent Gödel Modal Logic

In Section 2.3, we introduced the Paraconsistent Gödel Modal Logic (Bílková, Frittella, and Kozhemiachenko 2022), or $\mathbf{KG}^2$, a logic to reason about agents, trust, inconsistencies, and uncertainty. $\mathbf{KG}^2$ is applicable to social networks. The models for this logic include a *trust function* between agents, representing how much each agent trusts each other agent. In the fuzzy logic, the modalities *infima* and *suprema* makes use of the trust function. These modalities allow an agent to optimistically or pessimistically, respectively, aggregate all the information they receive from the other agents, weighted by their trust for each other agent, and taking all the evidence both for and against the proposition into account. This aggregation gives an agent evidence for and against a proposition, based on the information from their contacts. Thus, these modalities can be thought of as a quantification of an agent's belief or knowledge about a proposition, based on all the information or evidence for and against the proposition that they receive from the other agents in the network, allowing inconsistent information. We believe that this logic can be fruitfully applied to social networks, where information is not consistent, uncertainty is important, and evidence for and against an issue must be taken into account.

The project in this chapter is a tool for $\mathbf{KG}^2$ that takes a social network situation, the model, and a $\mathbf{KG}^2$ formula as input, and returns an evaluation in $[0, 1] \times [0, 1]$,

representing the evidence for and against the formula in the social network. The tool is publicly available as `kg2` on GitHub and on the Python Package Index (PyPI) under the MIT Licence allowing researchers to freely use it and contribute.

This chapter is structured as follows:

- Section 7.1 – The too: Explains the tool, showing the specification of the `Model` class for instantiation of models and `Formula` for declaration and evaluation of formulas in $\mathbf{KG}^2$.
- Section 7.2 – Examples: We show running examples by instantiating a model and evaluating formulas with modalities.

7.1 The tool

We developed a Python tool that implements the $\mathbf{KG}^2$ model with creation of formulas in $\mathsf{bi}\mathcal{L}^{\neg}_{\Box,\Diamond}$ and its evalutions given a model. In this section, we describe the tool, the class `Model` for instantiating the model, and class `Formula` for creations and valuations of formulas. The source code is available at our repository on GitHub[1] with its documentation[2].

7.1.1 `Model` class

The class `Model` allows the instantiation and manipulation of a model.

```
1 class Model:
2     worlds_size: int
3     relation: list[list[float]]
```

[1]Repository: https://github.com/josecoliveira/kg2/tree/v1.1.0
[2]Documentation: https://github.com/josecoliveira/kg2/blob/v1.1.0/DOCUMENTATION.md

```
    valuation1: dict[str, list[float]]
    valuation2: dict[str, list[float]]

    def __init__(
        self,
        worlds_size: int,
        relation: list[list[float]] = None,
        valuation1: dict[str, list[float]] = None,
        valuation2: dict[str, list[float]] = None,
        self_relation: bool = False,
    )
    def set_relation(
        self, world1: int, world2: int, value: float
    ) -> None
    def set_variable_valuation1(
        self, variable: str, value: list[float]
    ) -> None
    def set_variable_valuation2(
        self, variable: str, value: list[float]
    ) -> None
    def set_variable_valuation1_for_world(
        self, variable: str, world: int, value: float
    )
    def set_variable_valuation2_for_world(
        self, variable: str, world: int, value: float
    )
```

Listing 7.1: `Model` class' attributes, constructor, and methods.

The class `Model`, as its constructor is shown in Listing 7.1, takes the following parameters:

- `worlds_size`: the number of worlds (or states, agents) in the model.
- `relation`: adjacency matrix for the accessibility relation between worlds.
- `valuation1`: a dictionary of variables mapped to a list containing their valuation 1 (v_1) for each world. E.g.: `{"p": [1, 1, 0.4, 0.4]}`
- `valuation2`: a dictionary of variables mapped to a list containing their valuation 2 (v_2) for each world.
- `self_relation`: if set `self_relation` and the user does not define a relation, the relation will be a identity matrix. Otherwise, the relation will be a null matrix.

The class `Model` takes the parameters `worlds_size`, `relation`, `valuation1`, `valuation2`. Finally, the class has the following methods for model manipulation:

- `set_relation`: sets the trust between two agents.
- `set_variable_valuation1`: takes a string as a variable and a list to add or replace the valuation 1 (v_1) of a value for all agents.
- `set_variable_valuation2`: takes a string as a variable and a list to add or replace the valuation 2 (v_2) of a value for all agents.
- `set_variable_valuation1_for_world`: replaces the valuation 1 (v_1) of a variable for an agent.
- `set_variable_valuation2_for_world`: replaces the valuation 2 (v_2) of a variable for an agent.

7.1.2 `Formula` class and subclasses

The class `Formula` allows the instantiation of formulas and run valuations given a model. This class is an interface, meaning that it a superclass to define other classes. These subclasses instatiate atomic formulas, i.e. variables, and non-attomic formulas using the logical connectives in the $\mathsf{bi}\mathcal{L}^{\neg}_{\Box,\Diamond}$ language. Listing 7.2 shows the class `Formula`.

```
1 class Formula:
2     def valuation1(self, model: Model, world: int) -> float:
3     def valuation2(self, model: Model, world: int) -> float:
```

Listing 7.2: `Formula` class' methods.

The class `Formula` has two methods, `valuation1` and `valuation2`. They take a model and a world and return an valuation. As an interface class, these methods are not defined in `Formula`, but it its subclasses. Listing 7.3 shows all `Formula`'s subclasses and their contructors.

```
1  class Variable(Formula):
2      def __init__(self, variable: str)
3  class Negation(Formula)
4      def __init__(self, operand: Formula)
5  class Conjunction(Formula)
6      def __init__(self, left: Formula, right: Formula)
7  class Disjunction(Formula)
8      def __init__(self, left: Formula, right: Formula)
9  class Implication(Formula)
10     def __init__(self, left: Formula, right: Formula)
11 class Coimplication(Formula)
12     def __init__(self, left: Formula, right: Formula)
13 class Box(Formula)
```

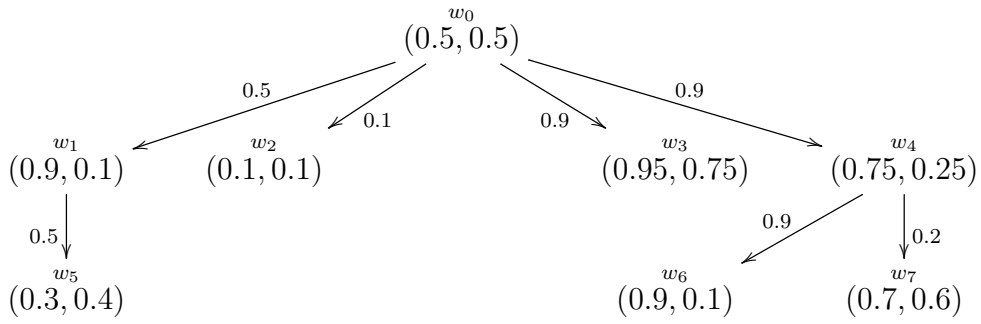
```
14      def __init__(self, operand: Formula)
15  class Diamond(Formula)
16      def __init__(self, operand: Formula)
```

Listing 7.3: `Formula` class' methods.

Naturally, `Variable` does have parameters, the unary operators take one formulas as parameters, and the binary operators take two formulas as parameters. Therefore, the class `Formula` implements a parsing tree where the class `Variable` represents the leaves of the tree.

7.2 Examples

In this section we present a example with a **K**biG model for fuzzy frames with evaluations of some formulas. Let $\mathfrak{M}$ be the model described by the diagram bellow:



(7.1)

We instantiate this model in Listing 7.4.

```
1  model = Model(8)
2  model.set_relation(0, 1, 0.5)
3  model.set_relation(0, 2, 0.1)
4  model.set_relation(0, 3, 0.9)
5  model.set_relation(0, 4, 0.9)
```

```
model.set_relation(1, 5, 0.5)
model.set_relation(4, 6, 0.9)
model.set_relation(4, 7, 0.2)
model.set_variable_valuation1(
    "p", [0, 0.9, 0.1, 0.95, 0.75, 0.3, 0.9, 0.7]
)
model.set_variable_valuation2(
    "p", [0.1, 0.1, 0.75, 0.25, 0.9, 0.1, 0.1, 0.6]
)
```

Listing 7.4: Instantiation of $\mathfrak{M}$.

Now, we test some formulas with the modalities from $\mathbf{KG}^2$:

Example 7.1. $v(\Box p, w_0) = (0.75, 0.9)$.

```
formula = Box(Variable("p"))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0.75, 0.9)
```

Listing 7.5: Evaluation of $v(\Box p, w_0)$.

Example 7.2. $v(\Diamond p, w_0) = (0.9, 0.1)$.

```
formula = Diamond(Variable("p"))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0.9, 0.1)
```

Listing 7.6: Evaluation of $v(\Diamond p, w_0)$.

Example 7.3. $v(\Box\Box p, w_0) = (0.3, 0.2)$.

```
formula = Box(Box(Variable("p")))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0.3, 0.2)
```

Listing 7.7: Evaluation of $v(\Box\Box p, w_0)$.

Example 7.4. $v(\Box\Diamond p, w_0) = (0, 0.9)$.

```
formula = Box(Diamond(Variable("p")))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0, 0.9)
```

Listing 7.8: Evaluation of $v(\Box\Diamond p, w_0)$.

Example 7.5. $v(\Diamond\Box p, w_0) = (0.9, 0)$.

```
formula = Diamond(Box(Variable("p")))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0.9, 0)
```

Listing 7.9: Evaluation of $v(\Diamond\Box p, w_0)$.

Example 7.6. $v(\Diamond\Diamond p, w_0) = (0.9, 0.1)$.

```
formula = Diamond(Diamond(Variable("p")))
formula.valuation1(model, 0), formula.valuation2(model, 0)
# >>> (0.9, 0.1)
```

Listing 7.10: Evaluation of $v(\Diamond\Diamond p, w_0)$.

8 Conclusions and Future Work

In this work, we presented models for social networks, experiments and results from these models, and tools for simulations and evaluations using quantitative formal methods and multi-agent logics.

Models in subjective logic We have used subjective logic to consistently extend the dynamic models from Alvim, Knight, and Valencia (2019) with *uncertainty*, allowing a more nuanced and realistic model of complex opinions in social networks. We identified four desiderata of properties that describe our interpretation of rationality when updating opinions. From subjective logic, we selected the trust discount and belief fusion operators to define our update function. We showed through examples that our function does not satisfy the desiderata.

On the other hand, we showed that our update function with cumulative belief fusion has the potential to represent a kind of interaction not described in Alvim, Knight, and Valencia (2019), even though it does not meet our initial desired properties for an update function. We also showed that if the agents disagree about a proposition, i.e. they have opposite opinions, there is a close enough distance between them where the update is stable, and they converge to a non-radical point. But if they are far enough, they will still converge, but also radicalize, converging to completely agree or completely disagree about the proposition. The next step for the update function with cumulative fusion is to have a clear definition of this behavior by calculating the convergence of the update function for each case.

Finally, we defined a new update function using subjective logic that is similar to a definition of classical belief update function from Alvim, Knight, and Valencia (2019) that already satisfies our desiderata. We showed that the new update function satisfies our desiderata, showing that it is possible to define at least one belief update function in subjective logic. We also showed that our subjective logic belief update function corresponds to the belief update function from Alvim, Knight, and Valencia (2019). This shows that subjective logic is promising for modeling social networks, since it can already capture the information used in Alvim, Knight, and Valencia (2019), and also is equipped with an additional method for reasoning about agents' uncertainty.

AKV model with dynamic influence We developed an extended AKV model that, unlike the original model, allows the update of the influence graph over time. We ran experiments on a model that updates both the belief configuration and the influence graph with the classic update function. In this model, the influence between every pair of agents becomes the proposition in discussion.

Using results from Alvim, Amorim, et al. (2023), we define balance-conservative influence graphs, graphs that, for every timestep, every agent influences and is influenced by the same degree. We show that clique graphs and bipartite graphs with equal-sized groups are balanced-conservative, and we show examples of star, wheel, friendship and bipartite graphs with different-sized groups that are non-balance-conservative. Finally, we propose two conjectures: 1) if a balanced influence graph contains an induced non-balanced-conservative influence graph, then this graph is non-balance-conservative; and 2) in a balanced influence graph, if the total influence of two agents is different, then the influence graph is non-balance-conservative. If they are proven to be true, then we have procedures to check if an influence graph

can lead to polarization.

AKV Python tool Given the use and continuous research on the AKV model, we developed a Python tool for the AKV model that implements the model and provides a catalog of initial configurations, influence graphs, and belief update functions already defined in the literature. The tool is available publicly and under the MIT license, allowing researchers to freely use it and contribute. With our tool we reproduced the results from Alvim, Amorim, et al. (2023) and our simulations correspond to theirs.

$\mathbf{KG}^2$ tool The Paraconsistent Gödel Modal Logic, or $\mathbf{KG}^2$, is a modal comparative logic equipped with Kripke models that can represent agents and belief with uncertainty and inconsistencies. These characterists allow a more interesting and powerful way to model social networks. With the modalities, we can reason about more degrees of relationship between agents. We implemented a Python tool for instantiation of the model and evaluation of formulas in $\mathbf{KG}^2$. The tool is available publicly and under the MIT license, allowing researchers to freely use it and contribute.

Future work With these contributions and tools for formal models in social networks, we aim them use it on real data to verify whether these models can explain the increasing polarization in social networks. Furthermore, these models can guide researchers to develop experiments with people and contribute with real life results and guidelines to help users in social networks to get into consensus instead of radicalize.

References

Alvim, Mário S., Bernardo Amorim, et al. (2021). "A Multi-agent Model for Polarization Under Confirmation Bias in Social Networks". In: *Formal Techniques for Distributed Objects, Components, and Systems - 41st IFIP WG 6.1 International Conference, FORTE 2021, Held as Part of the 16th International Federated Conference on Distributed Computing Techniques, DisCoTec 2021, Valletta, Malta, June 14-18, 2021, Proceedings*. Ed. by Kirstin Peters and Tim A. C. Willemse. Vol. 12719. Lecture Notes in Computer Science. Springer, pp. 22–41. DOI: 10.1007/978-3-030-78089-0_2. URL: https://doi.org/10.1007/978-3-030-78089-0_2 (cit. on pp. 31, 83).

— (2023). "A Formal Model for Polarization under Confirmation Bias in Social Networks". In: *Log. Methods Comput. Sci.* 19.1. DOI: 10.46298/LMCS-19(1:18)2023. URL: https://doi.org/10.46298/lmcs-19(1:18)2023 (cit. on pp. 3, 17, 19, 21, 31, 52, 55, 73, 78, 83, 95, 96).

Alvim, Mário S., Sophia Knight, and Frank Valencia (2019). "Toward a Formal Model for Group Polarization in Social Networks". In: *The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy - Essays Dedicated to Catuscia Palamidessi on the Occasion of Her 60th Birthday*. Ed. by Mário S. Alvim et al. Vol. 11760. Lecture Notes in Computer Science. Springer, pp. 419–441. DOI: 10.1007/978-3-030-31175-9_24. URL: https:

//doi.org/10.1007/978-3-030-31175-9_24 (cit. on pp. 3, 5, 7, 16, 31, 33, 44, 46, 48, 50–53, 55, 61, 73, 74, 79, 82, 94, 95).

Baltag, Alexandru, Zoé Christoff, et al. (2019a). "Dynamic Epistemic Logics of Diffusion and Prediction in Social Networks". In: *Stud Logica* 107.3, pp. 489–531. DOI: 10.1007/S11225-018-9804-X. URL: https://doi.org/10.1007/s11225-018-9804-x (cit. on p. 29).

— (2019b). *Dynamic Epistemic Logics of Diffusion and Prediction in Social Networks*. Vol. 107. 3, pp. 489–531. DOI: 10.1007/S11225-018-9804-X. URL: https://doi.org/10.1007/s11225-018-9804-x (cit. on p. 30).

Baltag, Alexandru and Sonja Smets (2008a). "A Qualitative Theory of Dynamic Interactive Belief Revision". In: pp. 11–58. URL: https://www.jstor.org/stable/j.ctt46mz4h.4 (visited on 07/15/2024) (cit. on p. 29).

— (2008b). "Probabilistic dynamic belief revision". In: *Synth.* 165.2, pp. 179–202. DOI: 10.1007/S11229-008-9369-8. URL: https://doi.org/10.1007/s11229-008-9369-8 (cit. on p. 29).

Bílková, Marta, Sabine Frittella, and Daniil Kozhemiachenko (2022). "Paraconsistent Gödel Modal Logic". In: *Automated Reasoning - 11th International Joint Conference, IJCAR 2022, Haifa, Israel, August 8-10, 2022, Proceedings*. Ed. by Jasmin Blanchette, Laura Kovács, and Dirk Pattinson. Vol. 13385. Lecture Notes in Computer Science. Springer, pp. 429–448. DOI: 10.1007/978-3-031-10769-6_26. URL: https://doi.org/10.1007/978-3-031-10769-6%5C_26 (cit. on pp. 22, 23, 25, 26, 86).

— (2023). "Non-standard Modalities in Paraconsistent Gödel Logic". In: *Logics in Artificial Intelligence - 18th European Conference, JELIA 2023, Dresden, Germany, September 20-22, 2023, Proceedings*. Ed. by Sarah Alice Gaggl, Maria

Vanina Martinez, and Magdalena Ortiz. Vol. 14281. Lecture Notes in Computer Science. Springer, pp. 420–436. DOI: 10.1007/978-3-031-43619-2_29. URL: https://doi.org/10.1007/978-3-031-43619-2%5C_29 (cit. on p. 26).

Christoff, Zoé and Jens Ulrik Hansen (2015). "A logic for diffusion in social networks". In: *J. Appl. Log.* 13.1, pp. 48–77. DOI: 10.1016/J.JAL.2014.11.011. URL: https://doi.org/10.1016/j.jal.2014.11.011 (cit. on p. 29).

Christoff, Zoé, Jens Ulrik Hansen, and Carlo Proietti (2016). "Reflecting on Social Influence in Networks". In: *J. Log. Lang. Inf.* 25.3-4, pp. 299–333. DOI: 10.1007/S10849-016-9242-Y. URL: https://doi.org/10.1007/s10849-016-9242-y (cit. on p. 29).

Delgrande, James P., Bryan Renne, and Joshua Sack (2019). "The logic of qualitative probability". In: *Artif. Intell.* 275, pp. 457–486. DOI: 10.1016/J.ARTINT.2019.07.002. URL: https://doi.org/10.1016/j.artint.2019.07.002 (cit. on p. 30).

Ditmarsch, Hans van, Jérôme Lang, and Abdallah Saffidine (2013). "Strategic voting and the logic of knowledge". In: *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013), Chennai, India, January 7-9, 2013*. Ed. by Burkhard C. Schipper. URL: http://www.tark.org/proceedings/tark_jan7_13/p196-van-ditmarsch.pdf (cit. on p. 30).

Esteban, Joan-María and Debraj Ray (1994). "On the Measurement of Polarization". In: *Econometrica* 62.4, pp. 819–851. ISSN: 00129682, 14680262. URL: https://doi.org/10.2307/2951734 (visited on 07/15/2024) (cit. on pp. 17, 31).

Fagin, Ronald, Joseph Y. Halpern, and Nimrod Megiddo (1990). "A Logic for Reasoning about Probabilities". In: *Inf. Comput.* 87.1/2, pp. 78–128. DOI: 10.1016/

0890-5401(90)90060-U. URL: https://doi.org/10.1016/0890-5401(90)90060-U (cit. on p. 30).

G1 (2023). "Mensagens mostram como bolsonaristas articularam ato em Brasília que levou a invasão de STF, Congresso e Planalto". In: *G1*. URL: https://g1.globo.com/politica/noticia/2023/01/08/mensagens-bolsonaristas-terroristas-brasilia.ghtml (cit. on p. 1).

Gargiulo, Floriana and Yérali Gandica (2017). "The Role of Homophily in the Emergence of Opinion Controversies". In: *J. Artif. Soc. Soc. Simul.* 20.3. DOI: 10.18564/JASSS.3448. URL: https://doi.org/10.18564/jasss.3448 (cit. on p. 31).

Golub, Benjamin and Matthew O. Jackson (Feb. 2010). "Naïve Learning in Social Networks and the Wisdom of Crowds". In: *American Economic Journal: Microeconomics* 2.1, pp. 112–49. DOI: 10.1257/mic.2.1.112. URL: https://www.aeaweb.org/articles?id=10.1257/mic.2.1.112 (cit. on p. 31).

Hunter, Aaron (2017). "Reasoning About Trust and Belief Change on a Social Network: A Formal Approach". In: *Information Security Practice and Experience - 13th International Conference, ISPEC 2017, Melbourne, VIC, Australia, December 13-15, 2017, Proceedings*. Ed. by Joseph K. Liu and Pierangela Samarati. Vol. 10701. Lecture Notes in Computer Science. Springer, pp. 783–801. DOI: 10.1007/978-3-319-72359-4_49. URL: https://doi.org/10.1007/978-3-319-72359-4_49 (cit. on p. 30).

Jackson, Jasper (2017). "Eli Pariser: activist whose filter bubble warnings presaged Trump and Brexit". In: *The Guardian*. URL: https://www.theguardian.com/media/2017/jan/08/eli-pariser-activist-whose-filter-bubble-warnings-presaged-trump-and-brexit (cit. on p. 29).

Jøsang, Audun (2016). *Subjective Logic - A Formalism for Reasoning Under Uncertainty*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer. ISBN: 978-3-319-42335-7. DOI: 10.1007/978-3-319-42337-1. URL: https://doi.org/10.1007/978-3-319-42337-1 (cit. on pp. 5, 8–10, 12).

Jøsang, Audun, Ross Hayward, and Simon Pope (2006). "Trust network analysis with subjective logic". In: *Computer Science 2006, Twenty-Nineth Australasian Computer Science Conference (ACSC2006), Hobart, Tasmania, Australia, January 16-19 2006*. Ed. by Vladimir Estivill-Castro and Gillian Dobbie. Vol. 48. CRPIT. Australian Computer Society, pp. 85–94. URL: https://dl.acm.org/citation.cfm?id=1151710 (cit. on p. 5).

Jøsang, Audun, Dongxia Wang, and Jie Zhang (2017). "Multi-source fusion in subjective logic". In: *20th International Conference on Information Fusion, FUSION 2017, Xi'an, China, July 10-13, 2017*. IEEE, pp. 1–8. DOI: 10.23919/ICIF.2017.8009820. URL: https://doi.org/10.23919/ICIF.2017.8009820 (cit. on pp. 5, 12–15).

Li, Lin et al. (2013). "Consensus, Polarization and Clustering of Opinions in Social Networks". In: *IEEE J. Sel. Areas Commun.* 31.6, pp. 1072–1083. DOI: 10.1109/JSAC.2013.130609. URL: https://doi.org/10.1109/JSAC.2013.130609 (cit. on p. 31).

Liu, Fenrong, Jeremy Seligman, and Patrick Girard (2014). "Logical dynamics of belief change in the community". In: *Synth.* 191.11, pp. 2403–2431. DOI: 10.1007/S11229-014-0432-3. URL: https://doi.org/10.1007/s11229-014-0432-3 (cit. on p. 30).

Oliveira, José C. (2022). "Subjective Logic as a Model for Social Networks". In: *Proceedings of the ESSLLI 2022 Student Session*, pp. 128–140. DOI: 10.21942/

uva.20370810.v1. URL: https://doi.org/10.21942/uva.20370810.v1 (cit. on p. 3).

Oliveira, José C. and Sophia Knight (2024a). *josecoliveira/akvmodel: v1.2.3*. Version v1.2.3. DOI: 10.5281/zenodo.12817825. URL: https://doi.org/10.5281/zenodo.12817825.

— (2024b). *josecoliveira/kg2: v1.1.0*. Version v1.1.0. DOI: 10.5281/zenodo.12982711. URL: https://doi.org/10.5281/zenodo.12982711.

Rodrigues, Marcus et al. (2023). "Bolsonaristas invadem Congresso, Planalto e STF em manifestação antidemocrática". In: *Metrópoles*. URL: https://www.metropoles.com/distrito-federal/bolsonaristas-extremistas-manifestacao-brasilia (cit. on p. 1).

Seligman, Jeremy, Fenrong Liu, and Patrick Girard (2013). "Facebook and the epistemic logic of friendship". In: *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013), Chennai, India, January 7-9, 2013*. Ed. by Burkhard C. Schipper. URL: http://www.tark.org/proceedings/tark_jan7_13/p229-seligman.pdf (cit. on p. 29).

Sîrbu, Alina et al. (Mar. 2019). "Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model". In: *PLOS ONE* 14.3. Ed. by Floriana Gargiulo, e0213246. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0213246. URL: http://doi.org/10.1371/journal.pone.0213246 (cit. on p. 31).

Venâncio, Otavio R et al. (2024). "Unraveling User Coordination on Telegram: A Comprehensive Analysis of Political Mobilization during the 2022 Brazilian Presidential Election". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 18, pp. 1545–1556. DOI: 10.1609/icwsm.v18i1.31408.

URL: https://ojs.aaai.org/index.php/ICWSM/article/view/31408 (cit. on p. 1).

Weterman, Daniel (2023). "Manifestantes fazem ato na Esplanada para pedir intervenção militar e até prisão de Lula". In: *Estadão*. URL: https://www.estadao.com.br/politica/manifestantes-fazem-ato-na-esplanada-para-pedir-intervencao-militar-e-ate-prisao-de-lula/ (cit. on p. 1).