Measuring bi-polarization with argument graphs

Carlo Proietti[1], Davide Chiarella[1]

[1]National Research Council of Italy, Institute, for Computational Linguistics, via De Marini 6, Genova

Abstract

Multi-agent models play a significant role in testing hypotheses about the unfolding of opinion dynamics in complex social networks. The model of the Argument Communication Theory of Bi-polarization (ACTB), developed by Maes and Flache (2013), shows that simple circulation of arguments among individuals in a group can determine strong differentiation of opinions (bi-polarization effects) even with a small degree of homophily. The ACTB model and similar ones have nevertheless one limitation: given a topic of discussion, only direct pro and con arguments for it are considered. This does not allow to account for the topology of a more complex debate, where arguments may also interact indirectly with the topic at stake. This gap can be filled by using Quantitative Bipolar Argument Frameworks (QBAF). More specifically, by applying measures of argument strength for QBAFs in order to calculate the agents' opinion. In the present paper we generalize the ACTB measure of opinion strength to acyclic bipolar graphs and compare it with other measures from the literature. We then present a revised version of the ACTB model, where the agents' knowledge bases are structured as subgraphs of an underlying global knowledge base (described as a QBAF). We first test that the predictions of the ACTB model are confirmed when the underlying QBAF contains only direct pro and con arguments for a topic. We then explore more complex topologies of debate with two additional batches of simulations. Our first results show that changing the topology, while keeping the same number of pro and con arguments, has no significant impact on bi-polarization dynamics.

1. Introduction

In social psychology, group polarization is commonly understood as a situation where the opinions of individuals in a group tend to become more radical after discussing with peers [1]. Closely related to this are so-called bi-polarization effects, where the opinion of two subgroups split in opposite directions (both getting more radical).[1] *Social* and *informational influence* have been identified by psychologists as essential explanatory causes of both phenomena. In more recent years, multi-agent models have been developed to test these hypotheses via computer simulations on artificial societies [2, 3, 4, 5]. In the general setup of these models, agents interact with their direct links and revise their opinion about a given topic after every exchange. In a first family of models, of the kind inspired by [6], exchanges consist of individuals disclosing their opinion and revising it as a function of their previous opinion and of their neighbors' displayed opinion (mostly by a mechanism of averaging). These models aim to test the polarizing effect of standard social influence (we may call it peer pressure) as theorized by [7]. But, in order to show

5th Workshop on Advances in Argumentation in Artificial Intelligence
carlo.proietti@ilc.cnr.it (C. Proietti); davide.chiarella@ilc.cnr.it (D. Chiarella)

[1]Often, the term 'polarization' is used to denote the latter phenomenon. In what follows we stick to our distinction in order to avoid confusion.

bi-polarization effects, they need to assume both positive and negative influence (distancing) among individuals.

A second type of models, most prominently the model of *Argument Communication Theory of Bi-polarization* (ACTB) developed by [5], assumes that agents interact not by displaying their opinion to others, but by communicating arguments they have in their knowledge base and that are either in favor (*pro*) or against (*con*) the topic of discussion. Furthermore, the opinion of each agent is only a function of the newly acquired arguments and the ones he already has, and not of the opinion displayed by others. In a nutshell, the more pro arguments an agent owns, the more favorable will be its opinion, in a scale ranging from -1 (totally against) to +1 (totally in favor). The ACTB model was devised to test the explanatory hypothesis of the *persuasive arguments theory* by [8]. The latter assumes that the main driver of polarization is the circulation of novel and persuasive arguments in favor or against the given topic (rather than peer pressure). In virtue of these features, the ACTB model and similar ones can be classified as models of informational influence. In [5], the authors show that, by assuming a relatively small degree of homophily, i.e. the tendency of individuals to communicate with those who share a similar opinion, this mechanism of informational influence suffices to generate strong bi-polarization effects.

In the ACTB model, both the set of all potentially available arguments (we call it the *global knowledge base*) and the agents' *individual knowledge base* are constituted by pro and con arguments for a given topic. It is natural to frame such knowledge bases as directed graphs with two different types of arrows for supports and attacks, i.e. *bipolar argumentation frameworks* [9], and one terminal node representing the topic of discussion. This, in turn, suggests the possibility of using measures of *argument strength* to calculate the degree of opinion, essentially as the strength of the terminal node. Many such measures have been developed in the literature on *gradual argumentation* [10, 11, 12, 13, 14, 15, 16] and, as we shall show, the ACTB measure can be generalized into a new one. From this angle, one (global or individual) knowledge base in the ACTB model can be regarded essentially as a *star tree* (see Figure 1a), more precisely as a *rooted in-tree* with all nodes at maximum distance one from the root, i.e. the topic node. Nodes other than the root therefore work as independent attackers or supporters with equal weight (see Section 2 for explanation). This however constitutes a strong simplifying assumption of the model, since it suppresses a relevant dimension of an argumentative knowledge base, namely that pro and con arguments may interact with each other at different levels. To make this clear with an example, suppose our topic of discussion t is vaccination for COVID-19. One possible individual knowledge base k_1 could be constituted by the following con arguments:

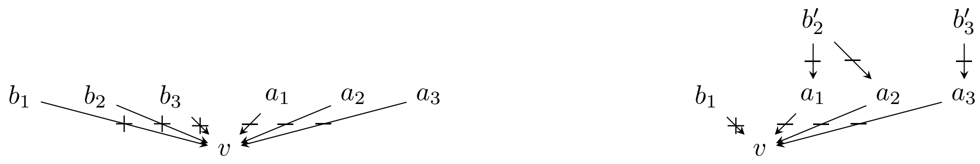
Con

a_1 Vaccination is useless because herd immunity will never be reached.
a_2 Vaccination is useless because it is not widespread in poor countries and therefore the virus would circulate anyway.
a_3 Vaccination is useless because vaccinated individuals can still infect others.

together with the following pro arguments:

Pro

b_1 Vaccination is a social duty for everybody.



(a) k_1 (b) k_2

Figure 1: Two different knowledge bases k_1 and k_2. A directed edge labelled with $+$ indicates support. One labelled with - indicates attack.

b_2 My doctor says I should get vaccinated.
b_3 I need the EU covid certificate.

A different knowledge base, say k_2, may instead consist of the same con arguments a_1, a_2, a_3, but a different set of pro arguments, namely:

Pro'

b_1 Vaccination is a social duty for everybody
b'_2 Herd immunity has been reached in many cases for viruses that have now disappeared (e.g. smallpox). And even when this is still not the case, viruses have often disappeared in parts of the world where vaccination was significant (e.g. polio).
b'_3 Vaccinated individuals are not totally immune, but the probability of getting infected, and therefore to infect others, is significantly lower.

k_1 and k_2 correspond, respectively, to the graph in Figure 1a and Figure 1b. Both knowledge bases have three pro and three con arguments w.r.t. the topic v, and therefore the ACTB measure cannot distinguish among them, so that the resulting opinion will be neutral (assuming that all arguments have equal weight). However, at an intuitive level, the opinion determined by k_2 is more likely to support a favorable attitude towards v.[2] The *generalized ACTB measures* we introduce follow this intuition and, *ceteris paribus*, predict a higher opinion strength in the case of k_2.

Given a larger variety of possible graph configurations and the new measures, it is then an interesting question to test whether, given an equal number of pro and con arguments, the topology of the underlying global knowledge base has an impact on the bi-polarization dynamics predicted by the ACTB model. In particular, it is desirable to ascertain whether augmenting the likelihood of a favorable attitude towards v forces more positive consensus among agents. Or alternatively, in cases where two subgroups end at the extreme poles of the opinion spectrum, if this increases the cardinality of the group of agents with an absolute pro opinion. To provide (partial) answers to these questions, we devised a revised version of the original multi-agent ACTB model, written in `Python`, where global and individual knowledge bases are encoded as bipolar graphs, and the opinion of the agents is obtained, at each step, by measuring the strength of the topic v in their individual knowledge base according to the revised measures. We first test that results agree with those by [5] in the case where the global

[2] Deciding whether and how much this is the case is a task for empirical research. On the other hand, measures of opinion strength need to account for such a distinction.

knowledge base has a star-tree structure like the one in Figure 1a. Then, we start exploring what happens by introducing a structural imbalance between pro and con arguments (while keeping their cardinality the same) as in the case of Figure 1b. As we shall see in Section 3.2, such modifications have no significant impact on bi-polarization dynamics in terms of rate of bi-polarizations, time for convergence, and cardinality of splitting subgroups. Therefore, the answer to the questions above is essentially negative. This means that, based solely on argument strength, the ACTB model cannot account for situations where opinion clustering generates minorities. As a consequence, it seems that further assumptions are needed to produce and explain such scenarios.

The paper proceeds as follows. In Section 2 we introduce the basic notions concerning *Quantitative Bipolar Argumentation Frameworks* (QBAF) and the standard ACTB measure. We then show how to generalize it as a measure of argument strength for acyclic QBAFs and discuss some of the properties of the new measure. In Section 3 we describe our revised version of the ACTB model. In Section 3.2 we check that our model predicts bi-polarization effects that are consistent with those predicted by the original model in [5] in cases where the first can be reduced to the latter (only direct pro and con arguments). We then account for our initial observations on two different structures where the cardinality of pro and con arguments is kept equal. Finally, in Section 4 we discuss possible expansions and more systematic simulation setups as well as future avenues for research.

2. Measures of opinion strength in Quantitative Bipolar Argumentation Frameworks

2.1. Preliminaries on abstract argumentation

The type of structures we deal with are instances of *Quantitative Bipolar Argumentation Frameworks* (QBAF) ([16], [15]), which are defined as follows:

Definition 2.1 (QBAF [16]). *A QBAF is a quadruple (A, R^-, R^+, w) consisting of a finite set A of arguments, a binary (attack) relation R^- on A, a binary (support) relation R^+ on A and a total function $w : A \longrightarrow I$ from A to a preordered set I.*

Here, for any $a \in A$, $w(a)$ is the *base score* of a, intended as the weight of an argument previous to any impact from other arguments. In what follows we adopt the interval $[0, 1]$, with the natural ordering relation on real numbers, as our preordered set for all measures. Some useful notation is the following. $R^-(a) = \{b \in A \mid (b, a) \in R^-\}$ denotes the set of direct attackers of a, whereas $R^+(a) = \{b \in A \mid (b, a) \in R^+\}$ is the set of its direct supporters. Following [16], it is useful to denote by $R^-_*(a)$ (resp. $R^+_*(a)$) the set of *effective* attackers (resp. supporters) of a.[3] Furthermore, let us denote $R = R^- \cup R^+$ the union of both relations. Then, let $Neg(a)$ be the set of all arguments b such that there is path $b = a_0R \ldots Ra_n = a$ (with $n \geqslant 1$) that contains an odd number of R^-. Let instead $Pos(a)$ be the set of all arguments b

[3]Depending on the modelling choice $R^-_*(a)$ may either be equal to $R^-(a)$ or to $R^-(a) \backslash \{b \in R^-(a) \mid s(b) = \bot\}$, where $\bot$ is the minimal element in the preorder I and $s(\cdot)$ is the strength function defined below. That is, with the second option we discount attackers with null strength. The same holds for supporters.

such that there is path $b = a_0 R \ldots R a_n = a$ (with $n \geqslant 1$) that contains an even number of R^-. Intuitively, $Neg(a)$ is the set of arguments with a negative influence on a, and $Pos(a)$ is the set of those with positive influence.[4] Here again, we set $Pos_*(a)$ (resp. $Neg_*(a)$) as the set of arguments with an effective positive (resp. negative) influence on a.[5]

The main idea behind gradual argumentation is to provide a semantics for the acceptability of arguments in terms of a *strength function* $s : A \longrightarrow I$. The function $s(a)$ is standardly provided as a function of the argument's base score $w(a)$ and of the strength of all other arguments affecting a. If we only consider elements of $R^-_*(a)$ and $R^+_*(a)$ as the ones affecting a, then we obtain a *local* measure of argument strength. If we instead consider all ancestors in $Pos_*(a)$ and $Neg_*(a)$, then we have a *global* measure [10]. To be neutral w.r.t. this choice we often write $Pro_*(a)$ to denote either $R^+_*(a)$ or $Pos_*(a)$, and $Con_*(a)$ to denote either $R^-_*(a)$ or $Neg_*(a)$.

2.2. The ACTB measure of opinion

In the formal model of ACTB [5] agents are equipped, at any step of the execution, with a set of n relevant arguments $a_1, \ldots, a_n$, chosen among a larger set of N possibly available arguments, that determine their opinion on a given topic v as a numerical value. Arguments are partitioned in two sets: $Pro(v)$ of pro arguments and $Con(v)$ of con arguments. Each argument a_l is assigned a weight $we(a_l)$ such that $we(a_l) = +1$ if $a_l \in Pro(v)$ and $we(a_l) = -1$ if $a_l \in Con(v)$. The opinion of agent i at time t is then provided by the following equation:

$$o_{i,t} = \frac{1}{|S_{i,t}|} \sum_{l=1}^{N} we(a_l) \cdot r_{i,t,l} \tag{1}$$

where $S_{i,t}$ is the set of relevant arguments for i at time t, and $r_{i,t,l}$ is the relevance, either 0 or 1, of a_l for i at time t. So the value of $o_{i,t}$ ranges in the interval $[-1, +1]$. For our purposes, it is easy to obtain an equivalent measure $o'_{i,t}$ ranging in the interval $[0, 1]$, by means of the following linear transformation:

$$o'_{i,t} = \frac{1 + o_{i,t}}{2} \tag{2}$$

Despite their polarity, all relevant arguments in this calculation have an equal strength of 1 and therefore an equal and independent impact on the opinion about v. Based on these features, it is natural to represent the knowledge base of agent i at time t as a star tree like the one of Figure 1a, where the node v is the topic, the upper nodes are the relevant arguments and the labelling of an edge from a_l to v identifies a_l either as a pro argument (if the label is +) or a con one (if the label is -). It then becomes natural to interpret $o_{i,t}$ as a measure of strength of the node v. Given a generic node a, this can be rewritten in the following form, using the terminology of QBAFs we introduced:

$$r(a) = \frac{\sum_{b \in Pro_*(a)} s(b) - \sum_{b \in Con_*(a)} s(b)}{|Pro_*(a)| + |Con_*(a)|} \tag{3}$$

[4]More fine-grained distinctions about positive and negative influence can be found in the literature (see e.g. [9]). This level of granularity is however enough for our present purpose.

[5]As before, we can either set $Pos_*(a)$ as equal to $Pos(a)$ or as $Pos(a) \setminus \{b \in Pos(a) \mid s(b) = \bot\}$. Same for $Neg_*(a)$.

which we normalize as

$$s(a) = \frac{1 + r(a)}{2} \tag{4}$$

Again, $Pro_*(a)$ is either $R_*^+(a)$ or $Pos_*(a)$, and $Con_*(a)$ either $R_*^-(a)$ or $Neg_*(a)$, depending on whether we adopt a local or a global measure. In the case of a star-tree graph, choice among the two options is clearly indifferent to calculate $s(v)$, since the set of ancestor nodes coincides with that of direct attackers and supporters. Clearly, Equation 4 cannot be employed to calculate the strength of the initial nodes nor of those with ineffective ancestors, since we would get a division by 0. Typically, such cases are covered by postulating $s(a) = w(a)$, so that we get the following definition by cases:

$$s(a) = \begin{cases} w(a) & \text{if } Pro_*(a) \cup Con_*(a) = \varnothing \\ \frac{1+r(a)}{2} & \text{otherwise} \end{cases} \tag{5}$$

Now, this measure is well-defined for finite acyclic graphs, since it allows to calculate the strength of every node starting from the initial ones. Moreover, it is fully consistent with the original ACTB measure when we assume $w(b) = 1$ for the initial nodes (i.e. maximal weight).

It should be noticed that here, as in the ACTB measure, the strength of non-initial nodes is fully determined by the strengths of the affecting nodes. This way of measuring $s(a)$ fully discounts the base score of a, since it makes it depend only on the impact of its ancestors. To take this into account we may want to generalize our measure further to the following:

$$s'(a) = p_1 \cdot w(a) + p_2 \cdot s(a) \tag{6}$$

where $p_1 + p_2 = 1$. Here p_1 and p_2 are parameters that determine a weighted average for updating the argument strength. Intuitively, the weight of p_1 tells us how much to count the base score, while the parameter p_2 says how much to weight the shift determined by the affecting nodes. We retrieve the ACTB measure by setting $p_1 = 0$.[6]

2.3. Differences among measures: global and local

The choice between a local and a global measure of argument strength can make a substantial difference in terms of opinion dynamics. This can be seen by considering a simple case of *reinstatement* as that of Figure 2. Here we assume that $w(b) = w(c) = 1$, and v is always the topic at stake. For both the local and the global approach we obtain $s(c) = w(c) = 1$ and $s(b) = 0$ (since b has only one ancestor and it is an attacker with maximal strength). But then, the value of $s(v)$ crucially depends on the choice. With a local measure, v has only one node affecting it (b) with null strength. If we consider this as a limit case where $Pro_*(a) \cup Con_*(a) = \varnothing$ (see fn. 3 and Equation 5), then we get $s(v) = w(v)$. Otherwise, $s(v) = 0.5$. Differently, in the global approach we cannot be in a limit case, since $c \in Pos_*(v)$. The value of $s(v)$ then depends on whether or not b counts as effective. If not, then $s(v) = 1$, otherwise we have $s(v) = 0.75$, since b although being of null strength, mitigates the reinstatement effect of c by having an impact on the denominator of $r(v)$.

[6] Note that setting p_1 to 0 allows so-called big-jumps ([15]), i.e. the strength of any node a can be easily brought to 0 by its attackers, even though a has a high base score. Vice versa, a node with low base score can be brought to 1 by its supporters. Augmenting the weight of p_1 mitigates this phenomenon.

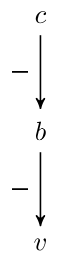


Figure 2: Reinstatement

2.4. Properties of the measures

Recent work in gradual argumentation ([16], [15]) explores general properties that can serve as desiderata for measures of argument strength. The most comprehensive list is provided by [16]. Given that these properties are provided for local measures, we can only test them on the local interpretation, i.e by reading $Pro_*(a)$ as $R^+_*(a)$ and $Con_*(a)$ as $R^-_*(a)$ in Equation 3. It is however interesting to check that our measure $s(a)$ differs from any other measure provided in the literature w.r.t. satisfaction of at least some of these properties (see [16] Sect. 5).

First of all, $s(\cdot)$ is not *balanced*, and a fortiori not *strictly balanced* ([16] Sect. 4). Indeed, it is not even the case that, when the set of attackers and supporters of a have equal strength[7] then $s(a) = w(a)$. This is due to the fact that the strength of an argument is determined solely by its ancestors, and not by its base score, in non-limit cases. It is easy to check that among the properties implied by (strict) balance and listed by [16], only **GP1** (aka *stability*) is satisfied.[8] For the abovementioned reason neither **GP2** (*weakening*), **GP3** (*strengthening*), **GP4** (*weakening soundness*), nor **GP5** (*strengthening soundness*) are guaranteed to hold.

On the other hand, $s(\cdot)$ is *strictly monotonic* ([16] Sect. 4). Monotonicity means, that if a and b are such that $w(a) \leqslant w(b)$, the set of supporters of b is at least as strong as the set of supporters of a, and the set of attackers of a is at least as strong as the set of attackers of b, then $s(a) \leqslant s(b)$.[9] Strict monotonicity means that whenever we replace 'at least as strong' with 'strictly more strong' in one of the preconditions, then $s(a) < s(b)$. By consequence, the properties **GP6**-**GP11**[16], implied by strict monotonicity, hold for this measure.

3. The multi-agent model

Our goal is to test whether the underlying topology of a debate has an impact on the bi-polarization dynamics predicted by the multi-agent ACTB model of [5]. To do this, we implemented a variation of this model. As a main modification, the set of all potentially available arguments (the global knowledge base) is now structured as an acyclic QBAF with one terminal

[7]Here 'equal strength' means that there is a bijection f between the elements of the two sets such that $s(b) = s(f(b))$ for any b.

[8]That is, If $R^-_*(a) = R^+_*(a) = \varnothing$ then $s(a) = w(a)$.

[9]Here, set B is at least as strong as A means that there is an injective map f from A to B such that for any $a \in A$, $s(a) \leqslant s(f(a))$.

node (the topic v). In accordance with this, the individual knowledge base of any agent, at any point of the execution, is a subgraph of it, and it always contains v. Then, in order to calculate the agent's opinion at any point we adapt our measures of Section 2.2.

3.1. General description of the model

The multi-agent model consists of a society of n interdependent agents, which simultaneously participate in an artificial influence process. Each agent i is attributed an opinion $o_{i,t}$ about the given issue v at each time point t. This is expressed by a numerical value that, in our case, ranges between 0 and +1. As in the original model, there is a limited number N of potential arguments about the issue at stake, and they are divided into two sets of, respectively, pro and con arguments. As mentioned, this global knowledge base is structured not as a vector but as a connected and acyclic QBAF F_g with v as its terminal node. The structure of the graph also determines the polarity of the arguments: each argument in $Pos(v)$ (see Section 2) is counted as a pro argument and each one in $Neg(v)$ as a con argument.[10] Each argument a is attributed an initial base score $w(a)$ such that $0 \leqslant w(a) \leqslant 1$. As in the standard model, we assume that, at each time t, the opinion of agent i is based only on a subset $S_{i,t}$ of relevant arguments, where $|S_{i,t}| \leqslant N$. This is summarized, for each agent i, by a relevance vector of N elements. The relevance $r_{i,t,l}$ of argument l for agent i at time t is either 1 (relevant) or 0 (not relevant).

In the standard model $o_{i,t}$ is determined by Equation 1. Here, $o_{i,t}$ is calculated by means of the measure described in Equation 5. More in detail, we first determine the subgraph $F_{i,t}$ of the global knowledge base, constituted by all arguments relevant for i at time t and the edges among them (as determined by the global knowledge base). Then, we calculate $s(a)$ for all nodes starting from the initial ones down to the terminal v.[11] We set $Pro_*(a) = Pro(a)$ and $Con_*(a) = Con(a)$ (see Section 2.2). The main issue, though, is that the graph $F_{i,t}$ can easily be disconnected (see below), and therefore we need to decide whether $Pro(a)$ and $Con(a)$ are evaluated w.r.t. F_g or $F_{i,t}$. Choice between the two options is a parameter of the model.[12] The value of $s(v)$ thus calculated is our $o_{i,t}$. The acyclicity assumption ensures that this process terminates.

As in the ACTB model, each agent is attributed a *recency vector* of N elements, each one having a value ranging from 0 to $S_{i,t}$, where a higher value indicates that the corresponding argument has been taken into account more recently, and where the value of 0 indicates that the argument is not relevant (because it has never entered the database or because it is too old and therefore disregarded). For each argument a_l, we denote its recency for agent i at time t by the number $r_{l,i,t}$ The recency vector is then updated at each step following the same mechanism

[10] In the general case of an acyclic QBAF, it is possible for an argument to fall in both sets $Pos(v)$ and $Neg(v)$. This does not constitute a problem when implementing our measures. However, for our present purpose, we decided to initialize our graphs as rooted in-trees, so to ensure that $\{Neg(v), Pos(v)\}$ forms a partition of the set of all arguments (See Section 3.2). So, the graph contains $N + 1 = |Pos(v)| + |Neg(v)| + 1$ nodes.

[11] This can be done either with a local or a global interpretation of $s(\cdot)$ (see Section 2.2). Choice between interpretations is set as a parameter of the model.

[12] Both options are intuitively grounded according to different interpretations of one agent's background knowledge. Choice between them can make a substantial difference w.r.t. the resulting opinion dynamics. Indeed, when $Pro(a)$ and $Con(a)$ refer to $F_{i,t}$, disconnected arguments, no matter how strong, have no impact on the calculation of $o_{i,t}$.

described by [5]: each new argument is attributed a value of $S_{i,t}$ and all others are diminished by one.

Exactly as in the original model, the opinion of each agent evolves as the result of a sequence of events, each one corresponding to one interaction between two agents. Each interaction goes in two sequential phases. (i) A selection phase where one agent i is randomly picked and then a partner j is selected with a probability proportional to the similarity of its opinion with that of i (*opinion homophily*).[13] (ii) A social influence phase, where the opinion of agent i is updated as a result of the interaction with j. Here again, we implement the exact same mechanism of the original model: one of j's relevant arguments is picked, and is then adopted by i.[14] Then, i updates its recency vector as described. This mechanism ensures that one argument is added and another is discarded, and therefore the number of arguments in the knowledge base of i is kept constant. As in the original model, each run iterates events until equilibrium is reached, and there are two kinds of equilibria: *perfect consensus* and *maximal bi-polarization*. In perfect consensus all agents hold the same opinion based on the same set of arguments. In maximal bi-polarization there are two maximally distinct subgroups, where members agree with each other in the same way (i.e. same opinion based on the same set of arguments). Both equilibria, and only them, are stable situations, as explained by [5] p. 6.

3.2. Setup and preliminary results

We implemented three different configurations of a global knowledge base, as in Figure 3. All scenarios have the same number of pro and con arguments but different topologies, to the effect that the strength $s(v)$ of the topic node v increases from 0.5 (Scenario 1) to 0.75 (Scenario 3).[15] This enables to answer our initial question as to whether providing stronger reasons for a pro attitude towards v has an influence on bi-polarization dynamics, e.g. by inducing more general consensus for v or, in case of a group split, by determining larger clusters of agents with a favorable attitude.

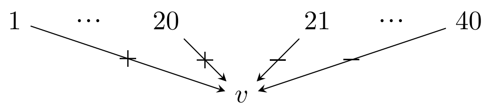
The first configuration consists of a star-tree with equal number of pro and con arguments. This configuration reduces to the vector configuration of the original ACTB model, and therefore should give similar results. To test this, we initialized a QBAF consisting of 41 arguments, i.e. the topic node v, 20 direct attackers (con arguments) and 20 direct supporters (pro) of v, all with maximal base score $w = 1$, as in Figure 3a, so that $s(v) = 0.5$ in the global knowledge base. We impose a strong level of homophily for the selection phase [h=9] (See [5], Equation 3). The total number of agents is $n = 20$ and all agents consider $S = 4$ arguments as relevant for opinion formation. Given S, there are S+1 possible distributions of relevant pro and con arguments for one agent, in our case S+1=5: 4 pro and 0 con arguments, 3 and 1, 2 and 2 etc. In our setup we randomly distributed the number of agents along such configurations, and then randomly attributed to each agent a number of pro and con arguments that fits its configuration.[16] We

[13]The measure of similarity $sim_{i,j,t}$ between i and j at time t and the corresponding probability of matching are described in Equation (2) and (3) of [5].

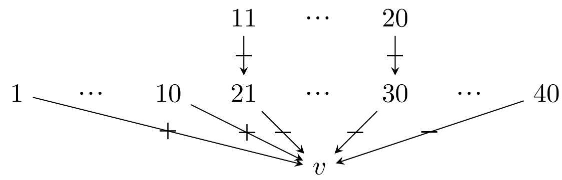
[14]By running the model, we observed that the directionality of the exchange has a strong impact on the resulting bi-polarization dynamic. Indeed, by setting i as the speaker and j as the receiver, the rate of bi-polarizations in simulation runs drops dramatically.

[15]Here, $s(v)$ can be regarded as measuring the opinion of an omniscient agent.

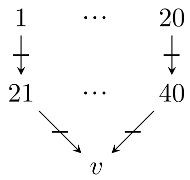
[16]More precisely, each agent is randomly assigned a pair (n, k) such that $n + k = S$. Then, n pro arguments



(a) Scenario 1. $s(v) = 0.5$



(b) Scenario 2. $s(v) = 0.625$



(c) Scenario 3. $s(v) = 0.75$

Figure 3: Global knowledge bases for scenarios 1-3.

initialized our model so that, when calculating the agents' opinion, the sets $Pro(a)$ and $Con(a)$ are evaluated w.r.t. F_g, so that arguments disconnected from v in the individual knowledge base are still counted as relevant for opinion formation.[17] We then ran the model as described in Section 3.1. More precisely, halt conditions are triggered (a) when all agents have opinion value either 0 or 1, and therefore maximal bi-polarization is bound to obtain[18]; (b) when the number of arguments considered as relevant by at least some of the agents is equal to S, which implies perfect consensus; and (c) after 6M events for space limits (which rarely happens, in this configuration, before (a) or (b) are triggered). Out of 500 simulation runs, we obtained bi-polarization 424 times with subgroups of equal cardinalities (an average of 10,02 con-oriented individuals and 9.98 pro-oriented), consistently with the results of [5].

As a second and third setup, we instead organized our graph as a rooted in-tree with nodes at a maximum distance of 2 from the root. As in the previous case, 20 pro and 20 con nodes are present. However, in the second setup we have 20 direct attackers, 10 direct supporters and 10 defenders (attackers of attackers), as in Figure 3b. In the third setup we instead have 20 direct attackers and 20 defenders (see Figure 3c). When we calculate the opinion strength by evaluating $Pro(v)$ and $Con(v)$ relative to F_g, these choices guarantee that $s(v)$ is higher in the global knowledge base of Scenario 2 than in Scenario 1 ($s(v) = 0.625$), and even stronger in Scenario 3 ($s(v) = 0.75$). In both the second and third setup all remaining parameters are kept the same. Again, we ran 500 simulations per setup and did not observe significant differences w.r.t. Scenario 1 in terms of numbers of bi-polarizations, nor concerning the cardinalities of subgroups of pro and con oriented agents. Indeed, we obtained 433 and 415 bi-polarizations respectively (Figure 4(a)), both with a slight average variation in time steps for obtaining group split: 4098 for Scenario 1 against 4140 and 3655 for Scenario 2 and 3 (see Figure 4(b)). The average of pro-oriented agents after group split is 10.33 in Scenario 2 and 10.40 in Scenario 3

---

and k con arguments are randomly selected and attributed different values from 1 to S in the agent's recency vector.

[17]For this star-tree configuration, this choice is indifferent, but it will be not for our second and third setup.

[18]This because the probability of communication between agents at the opposite poles is 0 as by [5], Equation 3.

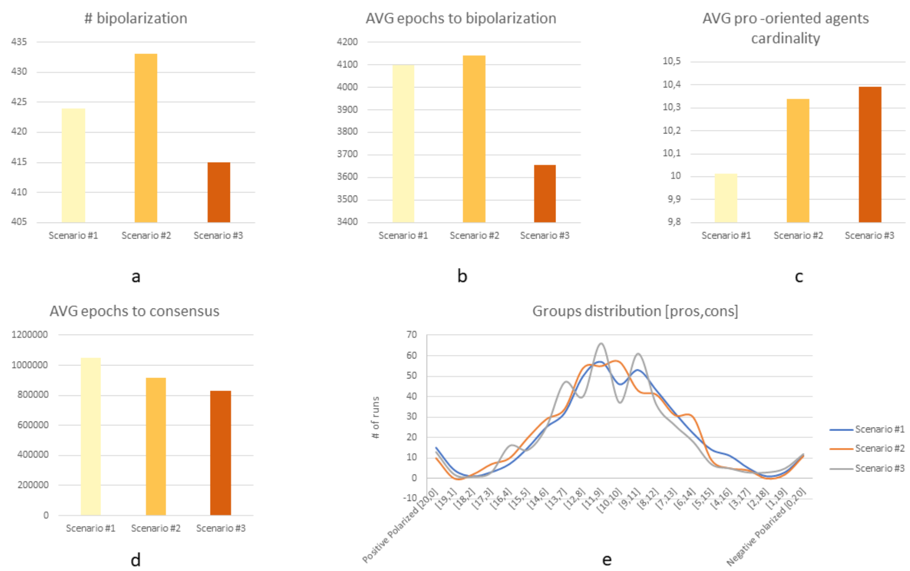


Figure 4: Results from simulation experiments on Scenario 1,2 and 3 (500 runs per condition, N = 20, pro = con = 20, S = 4).

(Figure 4(c)). Furthermore, in cases where perfect consensus is reached with halt condition (b), the average opinion is 0.5 in all three scenarios. The only difference consists in a decrease of the average time for convergence to perfect consensus, which ranges from 1.048.000 events in Scenario 1 against 918.000 in Scenario 2 and 831.000 in Scenario 3 (Figure 4(d)). But here again, this is balanced by the fact that the time-limit condition (c) is triggered only once in Scenario 1, while it occurs ten times for Scenario 2 and eleven times for Scenario 3. Finally, we also checked the standard deviations concerning these data and could not assess significant differences. For example, Figure 4(e) shows that the distribution of the cardinalities of subgroups that polarize in opposite directions is uniform over the three scenarios. As a consequence, we are not yet able to assess a significant impact of the topology of a debate on bi-polarization dynamics.

4. Discussion and future work

In this paper we generalized the ACTB measure of opinion into a measure of argument strength for QBAFs, and show how to integrate QBAFs into a multi-agent model of opinion dynamics. This opens up for the study of how topological features of a debate may influence consensus and bi-polarization effects. At the present stage, our results do not witness any significant impact. However, the generality of the model opens up for the exploration of a large parameter space and many questions can be framed via simulative experiments. As a first step, we need to test our preliminary observations on the three scenarios of Section 3.2 by varying the initialized parameters. Then, in order to look for robust results about the initial questions we asked, we need to analyze more scenarios and different ways of implementing the generalized ACTB measure of opinion, as provided in Section 2.3. It can be of further interest to also implement other measures from the literature in our model, to check whether their general properties, mentioned in Section 2.4, have an impact on the opinion dynamics of our model. There is then another interesting and probably more relevant question that our framework allows to ask, and it concerns the relevance update mechanism of the original ACTB model that we used. Indeed, once we are able to calculate the strength of the arguments in an individual knowledge base, it becomes natural to investigate how preferential communication of strong arguments or discarding of weaker ones may influence our opinion dynamics. These are only few of the possible venues for further research on a simulative basis.

Acknowledgements

The authors wish to thank the GARR Consortium which has provided the infrastructure for the simulative part of this work.

References

[1] D. J. Isenberg, Group polarization: A critical review and meta-analysis., Journal of personality and social psychology 50 (1986) 1141.

[2] M. W. Macy, J. A. Kitts, A. Flache, S. Benard, Polarization in dynamic networks: A hopfield model of emergent structure (2003).

[3] D. Baldassarri, P. Bearman, Dynamics of political polarization, American sociological review 72 (2007) 784–811.

[4] A. Flache, M. W. Macy, Small worlds and cultural polarization, The Journal of Mathematical Sociology 35 (2011) 146–176.

[5] M. Mäs, A. Flache, Differentiation without distancing. Explaining bi-polarization of opinions without negative influence, PloS one 8 (2013) e74516.

[6] R. P. Abelson, Mathematical models in social psychology, in: Advances in experimental social psychology, volume 3, Elsevier, 1967, pp. 1–54.

[7] L. Festinger, A theory of social comparison processes, Human relations 7 (1954) 117–140.

[8] A. Vinokur, E. Burstein, Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach., Journal of Personality and Social Psychology 29 (1974) 305.

[9] C. Cayrol, M.-C. Lagasquie-Schiex, On the acceptability of arguments in bipolar argumentation frameworks, in: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Springer, 2005, pp. 378–389.

[10] C. Cayrol, M.-C. Lagasquie-Schiex, Gradual valuation for bipolar argumentation frameworks, in: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Springer, 2005, pp. 366–377.

[11] P.-A. Matt, F. Toni, A game-theoretic measure of argument strength for abstract argumentation, in: European Workshop on Logics in Artificial Intelligence, Springer, 2008, pp. 285–297.

[12] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, G. Bertanza, Automatic evaluation of design alternatives with quantitative argumentation, Argument & Computation 6 (2015) 24–49.
[13] A. Rago, F. Toni, M. Aurisicchio, P. Baroni, Discontinuity-free decision support with quantitative argumentation debates (2016).
[14] P. Baroni, G. Comini, A. Rago, F. Toni, Abstract games of argumentation strategy and game-theoretical argument strength, in: International Conference on Principles and Practice of Multi-Agent Systems, Springer, 2017, pp. 403–419.
[15] L. Amgoud, J. Ben-Naim, Evaluation of arguments in weighted bipolar graphs, International Journal of Approximate Reasoning 99 (2018) 39–55.
[16] P. Baroni, A. Rago, F. Toni, From fine-grained properties to broad principles for gradual argumentation: A principled spectrum, International Journal of Approximate Reasoning 105 (2019) 252–286.