Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence

Social Abstract Argumentation

João Leite
CENTRIA
Universidade Nova de Lisboa
jleite@di.fct.unl.pt

João Martins
Computer Science Department
Carnegie Mellon University
jmartins@cs.cmu.edu

Abstract

In this paper we take a step towards using Argumentation in Social Networks and introduce Social Abstract Argumentation Frameworks, an extension of Dung's Abstract Argumentation Frameworks that incorporates social voting. We propose a class of semantics for these new Social Abstract Argumentation Frameworks and prove some important non-trivial properties which are crucial for their applicability in Social Networks.

1 Introduction and Literature Review

Despite the overall growing number of users of the Web 2.0, there are also many who abandon it, often unsatisfied with the quality of their experience. One of the reasons pointed out is the unstructured, often chaotic, kind of interactions that characterise most of the available Social Networks. This prevents a fulfilling experience for those seeking deeper interactions and not just increasing their number of *friends*, *likes*, or similar measures. Among these we can find experts or enthusiasts who wish to debate some particular topic with people of similar interests, or people who simply want to follow the debate, but who nevertheless have an opinion on what is happening and wish to express it and influence the outcome through some simple interaction mechanisms such as voting.

There are already some online communities that actively engage in structured debates. In websites such as `debate.org`, two users can take opposing sides in debating some issue, following the usual debate rules, while all other users can vote on the winning party according to a set of predefined criteria (e.g. who made more convincing arguments, who had a better conduct, etc). Despite their merits, these websites have several characteristics that limit their adoption in a wide Social Web scale, namely: 1) two (and only two) antagonistic users can engage in a debate, while all remaining users are limited to participate by voting on the winner; 2) the debate structure is very rigid, proceeding in a number of pre-fixed rounds with very strict debate rules that are not known, nor easy to follow, by most; and 3) there are no facilities to reuse arguments and debates.

In order to promote wide scale richer interactions in the form of debates, a Social Network must facilitate:

- More *open participation* where users with different levels of expertise are able to easily express their arguments, even without knowing the formal rules of debate.
- More *flexible participation* where debates are not restricted to a pair of users arguing for antagonistic sides, but where there may be more than just two sides, more users can propose arguments for each side, and each user is allowed to contribute with arguments for more than one side of the debate.
- More *detailed participation* where users are allowed to express their opinions by voting on individual arguments, instead of just on the overall debate's outcome.
- Appropriate *feedback* to users so that they can easily assess the strength of each argument, taking into account not only the logical consequences of the debate, but also the popular opinion and all its subjectiveness.

We envision a self-managing online debating system capable of accommodating two archetypal levels of participation. On the one hand, experts, or enthusiasts, will be provided with simple mechanisms to specify their arguments and also a way to specify which arguments attack which other arguments. To promote participation, arguments can be anything such as a textual description of the argument, a link to some source, a picture, or any other piece of information these users deem fit. On the other hand, less expert users who prefer to take a more observational role will be provided with simple mechanisms to vote on individual arguments, and even on the specified attacks. The system will be able to autonomously maintain a formal outcome to debates by assigning a strength to each argument, taking into account both the opinion expressed through the votes, but also the structure of the argumentation graph composed of the arguments and attacks. Users would be able to interact with the debate system through a GUI that easily allows them to observe the current state of the debate (e.g. depicting arguments with shades or sizes proportional to their strengths).

A key component of this system is the underlying knowledge representation framework that accommodates all the information provided by the users i.e. the arguments, attacks, and votes, together with a semantics that precisely characterises the strength of each argument.

There are websites such as `debategraph.org` and `compendium.open.ac.uk` which focus more on providing tools to facilitate the drawing of an argument map, than on reasoning with the information gathered as we propose. Others, like `livingvote.org`, also allow users to both create arguments and vote on them, but limit their output to showing the users the total number of votes for each argument, stopping short of reasoning with the information towards determining the outcome of a debate where the attacks between arguments as well as the votes are taken into account.

Since the form of argumentation we propose is intrinsically subjective, ultimately, the real value of its semantics can only be assessed by means of testing it with human users. However, several properties seem to be a priori requirements if a semantics is to be accepted by the users of this system:

1) *It must provide a model (valuation of all arguments) for every possible debate*. From a purely logical standpoint, one may consider that some debates simply contain inconsistencies that make it impossible to assign them meaningful semantics. However, we are dealing with the Social Web, where inconsistency is the *norm*. We believe that most of its users would prefer a system that would, nevertheless, provide them some valuation of the arguments that is somehow justifiable, instead of telling them that the debate is inconsistent.

2) *It should provide only one model for every possible debate*. Even though one can argue that some debates may have more than one model – e.g. a debate in which two arguments attack each other should have two models, each of which with one argument defeating the other – in order to succeed in the Web 2.0, simplicity is key. We believe that most users would would be turned away by a system with semantics where they and their opponents are both right and wrong at the same time (there being multiple models). Arguably, in the example above, users would prefer a system that simply lets them know that both arguments have e.g. half their original strength because they attack each other.

3) *Strength should go beyond the classical True/Accepted or False/Defeated.* The inclusion of votes presupposes a sort of popular opinion which is seldom universal. One can hardly take any controversial issue and state it as simply true or false, but it is not shocking to state it to have a certain *degree of truth*. Realistic argumentation systems must incorporate this fuzzy or many-valued approach on argument strength.

4) *Argument strength should be limited by popular opinion, and every vote should count.* In a true social system, there should be no arguments of authority, stronger than their popular support. They can be weaker since arguments may be attacked by other arguments, but the direct opinion expressed by the votes should act as an upper bound on the strength of the argument. Also, every positive vote should increase the strength of the argument (how much can depend on many factors), and any negative vote should decrease the strength of the argument, unless it had no strength to start with.

5) *The semantics should be smooth.* Users would be turned away from a system where a single vote would result in substantial changes in the model. Bearing in mind that the system is to be used in a wide scale, with votes being casted regularly, we do not want the GUI to look like a *Christmas Tree* with arguments changing their strength too abruptly. Along with this property comes the "No majority rule" according to which an argument should not disappear (or have a zero strength) just because some majority cast a negative vote, or some stronger argument is attacking it. As long as some users voted positive for some argument, and some users voted negative on its attacker, then the argument should have a non-zero strength since there is a social support for it, even if very low.

In [Dung, 1995], Dung proposed the notion of *Abstract Argumentation Frameworks* (*AAF*). This model of argumentation takes a set of abstract arguments i.e. arguments without internal structure or specific interpretation, and an attack relation between them, and defines a semantics (actually more than one) indicating which sets of arguments are mutually compatible. The simplicity and generality of the model, due to the abstract nature of the arguments, and the relationship with non-monotonic reasoning formalisms such as Default Logic and Logic Programming have contributed to the huge body of research devoted to this model of argumentation.

Abstract arguments allow great flexibility in the process of specifying arguments. However, this can eventually lead to the proposal of statements that are not really arguments. We will rely on the intelligence of the users to vote these down, thus reducing their impact in the framework. The alternative would be to consider structured arguments with some underlying logic. However, the process of formalising arguments is quite difficult if not for anything else, because it requires technical understanding of the logic being used which would prevent its wide scale adoption. It is worth noting that abstract arguments can be seen as encapsulations of structured arguments which hide their structure. Considerable efforts have recently been made in making structured argumentation more accessible [Rahwan *et al.*, 2007; Reed and Walton, 2005] which could, in the future, bring added value to the system we envision.

Given the nature of the system described above, namely the desired flexibility of the notion of an argument, for now, we will adopt abstract arguments and AAFs. However, they need to be extended to deal with the votes. In the literature one can find several extensions to Dung's work that differentiate the strength of arguments.

In [Prakken and Sartor, 1997], the authors attach priorities to arguments defining a partial order between them. However, these priorities are used to resolve conflicts between arguments and do not seem appropriate to our purposes since we do not seek a two-valued semantics where the votes define what should be accepted or not, but rather a many-valued semantics where the votes partially define the argument's value.

In [Bench-Capon, 2003] the author defines Value-based Abstract Argumentation by attaching to each argument the social values that it promotes, and making the semantics dependent on a particular preference order over values, representing a particular audience. Since the semantics is also based on a binary notion of argument acceptability, it is not adequate for our purposes.

In [Cayrol and Lagasquie-Schiex, 2005], the authors introduce Bipolar Argumentation Frameworks which extend Dung's AAF with a notion of support, dual to the notion of attack, and also provide an extension consisting of a valuation function that assigns to each argument a measure of its value, no longer restricted to the common two values. Even though this is in line with what we seek, the value of each argument is only a function of the values of the arguments that are related to it, not taking into account any strength assigned a priori to the arguments, like the votes are assigned in our system.

In [Matt and Toni, 2008] the authors presented a similar account to the degree to which an argument is acceptable, but based on game-theoretic notions. As with [Cayrol and Lagasquie-Schiex, 2005], no additional information other than the original argumentation graph is provided, so we cannot reuse such extension.

In [Dunne *et al.*, 2011], the authors introduce Weighted Argument Systems by extending Dung's AAF with weights on the attacks. These weights are taken into consideration when standard semantics have no models, and one is prepared to accept some contradiction, measured by the weights of the attacks we ignore. This proposal does not consider weights of arguments, so it does not directly apply to our case.

Even though there are many extensions of Dung's AAF that consider weights, none of them is appropriate to our envisaged system. To address this issue, in this paper we introduce the notion of *Social Abstract Argumentation Frameworks*, an extension of Dung's AAF with the possibility to associate votes to arguments, together with a semantics which assigns each argument a value, drawn from a pre-determined set of possible values, which represents the argument's strength taking into account both the structure of the graph and the social opinion expressed through the votes. This is done in Section 2, where we also illustrate the semantics with a simple example. In Sect. 3, we investigate some of the properties of the semantics, namely those encoding the desirable characteristics of our system. In Sect 4 we conclude.

We believe this paper to be a sound step towards enriching Web 2.0 with more structured forms of interactions, through debates, grounded on formal Argumentation Theory, but extended to deal with the subjective nature of the Social Web.

2 Social Abstract Argumentation Framework

The main formal concept introduced in this paper is that of a Social Abstract Argumentation Framework. It is an extension of Dung's AAF, composed of arguments and an attack relation to which we add an assignment of votes to each argument. To keep things simple, we will assume users will be given the possibility to cast *Pro* and *Con* votes for each argument. As will become clear below, if other more elaborate voting mechanisms are required, this framework can easily be extended as needed.

Definition 1 (Social Abstract Argumentation Framework) *A Social Abstract Argumentation Framework (SAF) is a triple $\langle \mathcal{A}, \mathcal{R}, V \rangle$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation between arguments and $V : \mathcal{A} \to \mathbb{N} \times \mathbb{N}$ is a total function mapping each argument to it's number of positive (*Pro*) and negative (*Con*) votes.*

Before proceeding we set the notation used throughout.

Notation 2 *Let $F = \langle \mathcal{A}, \mathcal{R}, V \rangle$ be a social abstract argumentation framework. We use $\mathcal{A}_F$, $\mathcal{R}_F$, and V_F to denote, respectively, the set of arguments ($\mathcal{A}$), the attack relation ($\mathcal{R}$), and the votes (V) of F. Let $a, b \in \mathcal{A}_F$ be arguments and $V_F(a) = (p, n)$. Then, $V_F^+(a) \triangleq p$ (resp. $V_F^-(a) \triangleq n$) denotes the number of positive (resp. negative) votes for argument a; $a\mathcal{R}_F b$ denotes that $(a, b) \in \mathcal{R}_F$; and $\mathcal{R}_F^-(a) \triangleq \{a_i : (a_i, a) \in \mathcal{R}_F\}$ denotes the set of (direct) attackers of argument a. Whenever unambiguous, we drop the subscript F from $\mathcal{A}_F$, $\mathcal{R}_F$, V_F, V_F^+, V_F^-, and $\mathcal{R}_F^-$.*

We can now tackle the semantics for SAFs. We will define, in an abstract way, the main components that will determine the semantics of SAF, which can then be instantiated to define custom tailored semantics.

Definition 3 (Semantic Framework) *A social abstract argumentation semantic framework is a 5-tuple $\langle L, \tau, \curlywedge, \curlyvee, \neg \rangle$ where:*

- *L is a totally ordered set with top element $\top$ and bottom element $\bot$, containing all possible valuations of an argument. This can simply be a real number between 0 and 1, but it can also be some other set of values such as colours, textures, or any other set that is appropriate for the users and the GUI used.*
- *$\tau : \mathbb{N} \times \mathbb{N} \to L$ is a vote aggregation function which produces a valuation of an argument based on its votes, which we dub the argument's social support.*
- *$\curlywedge : L \times L \to L$ is a binary algebraic operation on argument valuations used to determine the valuation of an argument based on its valuation given by the votes and how weak its attackers are;*
- *$\curlyvee : L \times L \to L$ is a binary algebraic operation on argument valuations used to determine the valuation of a combined attack;*
- *$\neg : L \to L$ is a unary algebraic operation on argument valuations used to determine how weak an attack is.*

Notation 4 *Let F be a SAF, $\mathcal{S} = \langle L, \tau, \curlywedge, \curlyvee, \neg \rangle$ a semantic framework, $a \in \mathcal{A}$ an argument, and $R = \{x_1, x_2, ..., x_n\}$ a multiset of elements of L. Then, with small abuse of notation:*

- *$\tau(a) \triangleq \tau(V(a)) = \tau(V^+(a), V^-(a))$;*
- *$\curlyvee R \triangleq ((x_1 \curlyvee x_2) \curlyvee ... \curlyvee x_n)$.*

To make things more concrete, we now define a simple semantical framework which will not only help us illustrate some concepts, but will turn out to have many desirable properties that will be discussed later on.

Definition 5 (Simple Vote Aggregation) *A simple vote aggregation function is any function $\tau_\varepsilon : \mathbb{N} \times \mathbb{N} \to [0, 1]$ such that $\varepsilon \geq 0$*

$$\tau_\varepsilon(v^+, v^-) = \begin{cases} 0 & v^+ = v^- = 0 \\ \frac{v^+}{v^+ + v^- + \varepsilon} & \textit{otherwise} \end{cases}$$

The full relevance of the parameter ε will be clearer below.

Definition 6 (Simple Product Semantics) *A simple product semantic framework is any $\mathcal{S}_\varepsilon^{\cdot} = \langle [0, 1], \tau_\varepsilon, \curlywedge, \curlyvee, \neg \rangle$ where 1) $x_1 \curlywedge x_2 = x_1 \cdot x_2$, 2) $x_1 \curlyvee x_2 = x_1 + x_2 - x_1 \cdot x_2$, 3) $\neg x_1 = 1 - x_1$ and 4) $\varepsilon \geq 0$.*

The *Simple Product Semantics* simply uses the *Product T-Norm* and its dual, the *Probabilistic Sum T-CoNorm*.

We are now ready to define the semantics of SAFs.

Definition 7 (Social Model) *Let F be a social abstract argumentation framework and $\mathcal{S} = \langle L, \tau, \curlywedge, \curlyvee, \neg \rangle$ a semantic framework. A total mapping $M : \mathcal{A} \rightarrow L$ is a* social model of F under semantics $\mathcal{S}$*, or* $\mathcal{S}$-model of F*, if*

$$M(a) = \tau(a) \curlywedge \neg \curlyvee \left\{ M(a_i) : a_i \in \mathcal{R}^-(a) \right\} \qquad \forall a \in \mathcal{A}$$

*We use $\mathcal{M}_{\mathcal{S}}^{F}$ to denote the set of all $\mathcal{S}$-*models *of F. Whenever F or $\mathcal{S}$ are unambiguous, they may be omitted from $\mathcal{M}_{\mathcal{S}}^{F}$. We refer to $M(a)$ as the valuation, or value, of a in M, dropping the reference to M whenever unambiguous.*

The semantics is essentially given by fix-points of a set of equations that assign, for each argument, a value that is based on its social support and on how weak the attack it is being subjected to is.

The previous definition implies that we consider the social support for an argument as an upper bound for the value assigned by the semantics, not only for the simple product semantics but also for a larger class of semantics studied below. This may feel strange to many logicians: things commonly accepted as perfectly logical propositions or axioms may not be in accord with the beliefs of the crowd. However, keep in mind that we are seeking a means of *subjective* reasoning. Our semantics is not expected to evaluate the *logical correctness* of arguments, but instead their *social acceptance*. It must garner that information from the crowd itself, and only then will social opinion truly influence reasoning.

Example 8 *One very common type of online debate revolves around technical products and gadgets, oftentimes playing a significant role in the purchasing decision. A typical forum discussion about which new generation phone to buy could be of the following form:*

a) *"The Wonder-Phone is the best new generation phone."*

b) *"No, the Magic-Phone is the best new generation phone."*

c) *links to a review of the M-Phone giving poor scores due to bad battery performance.*

d) *"c) is ignorant, since subsequent reviews noted only one of the first editions had such problems: [links]."*

e) *"d) is wrong. I found c)* knows *about that but withheld the information. Here's a link to another thread proving it!"*

It is worth noting that there are several types of arguments in this exchange. The first two arguments are unsupported claims, the third is merely a link making a point against the M-Phone, while the last two arguments are structured, with a claim supported by links. At our level of abstraction, meaningful arguments can be construed out of most participations.

*After a certain time, the above arguments accumulate votes, and it becomes apparent that the M-Phone and the W-Phone are subjectively at a stand-off, having 20 pro and con votes each. Then, perhaps many visitors followed **c)**'s link and fell in agreement with it, giving it a very substantial 60 positive votes. Similarly, users who followed **e)**'s link were forced to agree with it and disagree with **d)**, giving them symmetric votes of 40/10 for **e)** and 10/40 for **d)**. In the meantime, **c)**'s not quite outright lie garnered some 10 negative votes. The responsibility of discrediting **c)** thus falls to its attackers and to objective reasoning.*

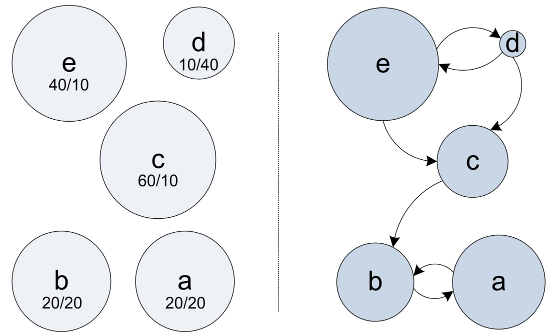

Figure 1: Framework with arguments scaled according to social support (left), and $\mathcal{S}_0^{\cdot}$-model (right)

*The votes amount to a social support of 0.5 for **a)** and **b)**, 0.86 for **c)**, 0.2 for **d)** and 0.8 for **e)**.*

*Figure 1 presents a graphical representation of the debate, where argument diameter is proportional to its crowd support on the left and its $\mathcal{S}_0^{\cdot}$-model on the right. More specifically, the $\mathcal{S}_0^{\cdot}$-model assigns the following values: 0.37 to **a)**, 0.25 to **b)**, 0.19 to **c)**, 0.05 to **d)** and 0.76 to **e)**.*

*Between the mutually exclusive **d)** and **e)** arguments, the difference already apparent in the social support is amplified further: in such an isolated mutually attacking situation, the stronger argument not only defends itself better against its counterpart's attack, but also mounts a heavier attack.*

*In a situation where the attack isn't mutual, however, **c)** finds its credibility heavily reduced despite the fact that it had the strongest social support. It still maintains a non-trivial fraction of its original strength, reflecting the weight given by the original agreement among the crowd.*

*Finally, without external influence, both **a)** and **b)** would have been weakened equally, maintaining a balance. With **c)**'s negative influence, the scales are tipped in **a)**'s favour. The difference between the models of **a)** and **b)** represents the fact that there was a minor issue regarding the technical reliability of M-Phone's manufacturing. On the other hand, since by and large the crowd feels the issue has been resolved, that difference is not overly accentuated. The result is a visible preference for Wonder-Phone over Magic-Phone, but not a definite one.*

3 Properties

We now turn our attention to the properties of SAFs and their semantics. The notion of semantic framework defined above was intentionally made general, to be able to accommodate semantics with many distinct features. In this section we are interested in investigating classes of semantics whose properties are in line with the desired features mentioned above. The first class we will define is that of Well-behaved Semantics which essentially enforces the operators used in the semantics framework to behave in the way we arguably already expected them to behave.

Definition 9 (Well-behaved Semantics) *A semantic framework $\mathcal{S} = \langle L, \tau, \curlywedge, \curlyvee, \neg \rangle$ is well-behaved if*

- *$\neg$ is antimonotonic, continuous, $\neg\bot = \top$, $\neg\top = \bot$ and $\neg\neg a = a$;*
- *$\curlywedge$ is continuous, commutative, associative, monotonic w.r.t. both arguments and $\top$ is its identity element;*
- *$\curlyvee$ is continuous, commutative, associative, monotonic w.r.t. both arguments and $\bot$ is its identity element;*
- *τ is monotonic w.r.t. the first argument and antimonotonic w.r.t the second argument.*

Proposition 10 *Any simple product semantic framework $\mathcal{S}_\varepsilon^{\cdot}$ is well behaved.*

An immediate consequence of this is that an unattacked argument retains its social value in well-behaved semantics.

Proposition 11 (Value of unattacked arguments) *Let F be a social abstract argumentation framework and $\mathcal{S} = \langle L, \tau, \curlywedge, \curlyvee, \neg\rangle$ a well-behaved semantic framework and $a \in \mathcal{A}$ an unattacked argument i.e. $\mathcal{R}^-(a) = \emptyset$. Then, $M(a) = \tau(a), \forall M \in \mathcal{M}_\mathcal{S}^F$.*
Proof. *Let $M \in \mathcal{M}_\mathcal{S}^F$. Then, $M(a) = \tau(a) \curlywedge \neg \curlyvee \{M(a_i) : a_i \in \mathcal{R}^-(a)\} = \tau(a) \curlywedge \neg \curlyvee \emptyset = \tau(a) \curlywedge \neg\bot = \tau(a)$.* ■

But more important, we are able to state one of the main results of this paper for the class of well behaved semantics.

Theorem 12 (Existence of Social Models) *Let F be a social abstract argumentation framework and $\mathcal{S}$ a well behaved semantics. Then F has at least one $\mathcal{S}$-model.*
Proof. *(Sketch) The proof uses Brouwer's fixed point theorem and the observation that M is continuous since the operators $\curlywedge$, $\curlyvee$ and $\neg$ of well behaved semantics are continuous.* ■

This previous result is of utmost importance if we are to widely use SAFs as users would be turned away from a system that was not capable of assigning a semantics to any situation it encounters. One can already imagine the trolls manipulating the system into a situation without a model.

Next we focus on the existence of a single model. As we discussed before, assigning multiple models to one SAF may be interesting from a theoretical point of view, but we expect the kind of users of this system to expect (if not demand) a single model. We are able to prove that, under certain conditions, there is a single model under the *Simple Product Semantics*:

Theorem 13 (Uniqueness of Social Models) *Let F be a social abstract argumentation framework such that $|\mathcal{R}^-(a)| \cdot \tau(a) < 1$, for every $a \in \mathcal{A}$. Then, F has one and only one $\mathcal{S}_\varepsilon^{\cdot}$-model.*
Proof. *(Sketch) The proof uses Banach's fixed point theorem together with the observation that we can define a function $T : [0,1]^n \to [0,1]^n$ (where n is the number of arguments in $\mathcal{A}$) based on M, which maps vectors of argument valuations into vectors of argument valuations, and is a contraction mapping in a specific compact metric space $A \subset [0,1]^n$, hence having a fix point, which is then shown to be in A, hence being the only fix point. This follows a similar proof in [Madrid and Ojeda-Aciego, 2011].* ■

We expect the following stronger result to also hold:

Conjecture 14 *Let F be a social abstract argumentation framework and $\varepsilon > 0$. Then, F has one and only one $\mathcal{S}_\varepsilon^{\cdot}$-model.*

Note that $\varepsilon > 0$ is an important requirement to ensure uniqueness of social models as illustrated by the following example.

Example 15 *Consider a social abstract argumentation framework F with just two arguments a and b attacking each other and $V^+(a) = V^+(b) > 0$ and $V^-(a) = V^-(b) = 0$. Then, F has an infinite number of $\mathcal{S}_0^{\cdot}$-models. For example, $M_1(a) = 1, M_1(b) = 0$ is a $\mathcal{S}_0^{\cdot}$-model, $M_2(a) = 0.3, M_2(b) = 0.7$ is also a $\mathcal{S}_0^{\cdot}$-model, just as any M_i such that $M_i(a) = 1 - M_i(b)$.*

It is now clear the reason for the introduction of ε in the function τ. It is there to deal with these situations where arguments with only positive votes (or arguments a such that $\tau(a) = \top$ to be more precise) cause an infinite number of models to occur when they are involved in circular attacks with other similar arguments. With a non-zero value for ε, such arguments with $\tau(a) = \top$ no longer occur. In the previous example, $lim_{\varepsilon\to 0} M(a) = lim_{\varepsilon\to 0} M(b) = 0.5$ which better reflects the fact that both arguments should be indistinguishable since they both only have positive votes, and they are both attacked by a similar argument.

Since multiple models only occur when arguments with $\tau(a) = \top$ are involved in circular attacks, a stronger theorem could actually be proved. In a practical use of this semantics, since every such argument is attacked by some other argument in the cycle, we could easily adopt a convention whereby each attack would count an additional negative vote to the attacked argument. This would allow the elimination of ε from the pre-conditions of the previous theorem, and even from the vote aggregation function altogether, as long as we also ensure that $\tau(a) = \bot$ for arguments with no votes (to prevent a division by 0).

For the cases covered by Theorem 13, it is possible to compute the unique $\mathcal{S}_\varepsilon^{\cdot}$-model, M_F, by determining the limit of a converging sequence. With $\mathcal{A} = \{a_1, ..., a_n\}$, let $M \in [0,1]^n$ be the vector such that its i^{th} component is $M(i) = M_F(a_i)$ and $T_F : [0,1]^n \to [0,1]^n$ be defined as follows:

$$T_F(I)(i) = \tau_\varepsilon(a_i) \cdot \prod_{a_j \in \mathcal{R}^-(a_i)} (1 - I(j))$$

Then, for every $I_0 \in [0,1]^n$, the sequence $I_{k+1} = T_F(I_k)$ converges to M. Consequently, for the cases covered by Theorem 13, we can approximate the unique $\mathcal{S}_\varepsilon^{\cdot}$-model by iterating T_F, given some initial assignment of arbitrary values to arguments.

In order to promote the trust of the users in social systems as the one we envision, it is important to guarantee other desirable properties, two of which encoded in the following proposition:

Proposition 16 (Influence of Votes) *Let F and F' be two social abstract argumentation frameworks such that F' results from F by adding a single positive (resp. negative) vote to some argument a and $\varepsilon > 0$. Let M be a $\mathcal{S}_\varepsilon^{\cdot}$-model of F. Then, there exists an $\mathcal{S}_\varepsilon^{\cdot}$-model M' of F' such that*

$M'(a) > M(a)$ (resp. $M'(a) \leq M(a)$). Furthermore, if Conjecture 14 holds, then M (resp. M') are the unique $\mathcal{S}^{\cdot}_{\varepsilon}$-models of F (resp. F').

This proposition encodes the desirable property that every positive vote on an argument has the immediate effect of increasing its valuation, and a counterpart for negative votes. Naturally that the amount of change due to a single vote depends on many factors such as the total number of votes on that argument and the specific semantics used. Note that both properties are not completely symmetrical since it is possible that a negative vote has no immediate influence in the valuation of the argument e.g. when the valuation of the argument was already the lowest possible i.e. $M(a) = \perp$ (or 0 in the case of the simple product semantics).

4 Discussion and Conclusions

The (family of) semantics for SAFs based on Social Models introduced in this paper can be tailored in different ways meet the specific needs of the applications and their users. For example, we can change:

- *The Vote Aggregation Function*: we can employ a different way of aggregating the votes altogether, or we can simply adjust the value of ε in the *Simple Vote Aggregation*. The larger the value of ε, the smaller the effect the initial positive votes will have in the social value of an argument, which can be useful to prevent *trolls* from creating arguments with just one positive vote, but capable of causing great harm to those being attacked. Also, the vote aggregation function can easily be changed so that it distinguishes arguments with different number of negative votes and no positive votes, e.g. by adding some positive ξ to its numerator.
- *The $\curlywedge$, $\curlyvee$ and $\neg$ operators*: instead of the *Product T-Norm* and the *Probabilistic Sum T-CoNorm*, other Archimedean operators could be used. For example, using the *Hamacher Product T-Norm* and its corresponding T-coNorm we would obtain a semantics that would diminish the effects of attacks when compared with the *Product T-Norm*.

Given the subjective nature of wide scale social debate, ultimately, we will have to conduct experiments with human users to assess the adequacy of each semantics, and choose one over the set of existing ones.

The framework can also be extended in several ways.

First, and foremost, we should also let users vote on the attacking relation. As it is, we assume that attacks are always at *full strength*. We should allow users to express how strong they consider each attack to be. This would be a straightforward extension to SAFs, relatively easy to be dealt with. We left it out of this paper to simplify the presentation.

Another possible extension is to consider more general *Vote Aggregation Functions* that take into account other variables to produce the social value of the argument. These additional variables could be, for example, the total number of votes in the debate to include some measure of robustness of each argument (e.g. what should happen with two mutually attacking arguments with 10/10 and 1000/1000 votes?).

In closing, we believe that *Social Argumentation Frameworks* lay the theoretical foundations for a deeper, more serious social web. By providing debates with formal, justifiable and yet subjective outcomes, it counteracts the growing trend of superfluous discussion. A system built on top of *Social Argumentation Frameworks* would maintain a detailed, reusable knowledge-base, and provide the infrastructure for more open, flexible debates than current systems allow. Furthermore, we prove properties that guarantee consistent, understandable feedback, thus facilitating the adoption of the envisioned system by people with a serious interest and experts alike. Systems like `livingvote.org`, `debategraph.org` and `compendium.open.ac.uk` are prime candidates to immediately benefit from Social Argumentation Frameworks and their semantics.

Acknowledgments

We would like to thank Carlos Damásio, Paolo Torroni and the anonymous reviewers for their insightful comments and suggestions.

References

[Bench-Capon, 2003] T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. of Logic and Computation*, 13(3):429–448, 2003.

[Cayrol and Lagasquie-Schiex, 2005] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *J. of Artificial Intelligence Research*, 23:245–297, 2005.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Dunne *et al.*, 2011] P. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011.

[Madrid and Ojeda-Aciego, 2011] N. Madrid and M. Ojeda-Aciego. On the existence and unicity of stable models in normal residuated logic programs. *Intl. J. of Computer Mathematics*, 2011. in press.

[Matt and Toni, 2008] P.-A. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Procs. of JELIA'08*, volume 5293 of *LNCS*, pages 285–297. Springer, 2008.

[Prakken and Sartor, 1997] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *J. of Applied Non-Classical Logics*, 7(1), 1997.

[Rahwan *et al.*, 2007] I. Rahwan, F. Zablith, and C. Reed. Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171(10-15):897–921, 2007.

[Reed and Walton, 2005] C. Reed and D. Walton. Towards a formal and implemented model of argumentation schemes in agent communication. *Autonomous Agents and Multi-Agent Systems*, 11(2):173–188, 2005.