Extending Social Abstract Argumentation with Votes on Attacks

Sinan Eğilmez[1], João Martins[1,2], and João Leite[1]

[1] CENTRIA and Departamento de Informática, FCT, Universidade Nova de Lisboa
`s.egilmez@campus.fct.unl.pt`, `jleite@fct.unl.pt`
[2] Computer Science Department, Carnegie Mellon University, Pittsburgh PA
`jmartins@cs.cmu.edu`

Abstract. Social abstract argumentation laid theoretical foundations for future online debating systems with formal backbones and semantics. The advantage of these envisioned new systems is their capability of formally justifying the social outcomes of their debates. Many recent extensions proposed for argumentation in general have addressed the issue that not all attacks between arguments are equal, especially in the eyes of the crowd. This work generalises social abstract argumentation to incorporate voting on attacks, inducing a social notion of attack strengths.

1 Introduction

The Web 2.0 proved extremely successful and its use has become second nature to most of the Internet population. With social networks now widely adopted and their users beating the one billion mark in 2013, the initial boom is over. As social networks become established, the patterns of these new social interactions slowly emerge. It is becoming apparent that many people are growing unsatisfied with the depth (or lack thereof) of interactions on social websites. A growing percentage of users are giving up on the Web 2.0 entirely for lack of intellectually stimulating discussions to which it is possible to attribute some sort of outcome.

This has given rise to websites that revolve around more meaningful interactions, and some of them purport to be a platform for serious debate.[1] Typically, these online debating systems (ODS) try to engage users with different degrees of desired involvement. On the one hand, experts and strongly opinionated people can propose their own debates, arguments, and go head to head against opponents. On the other hand, less involved users can simply share their opinion by means of simple voting mechanisms.

Despite their merits, these websites have several characteristics that limit their adoption in a wide Social Web scale, namely: 1) only two antagonistic users can engage in a debate, others can only vote for the winning side, but not on arguments themselves; 2) the debate structure is very rigid, with a pre-fixed number of rounds and very strict debate rules not known by most; 3) there are

---

[1] The websites `debategraph.org`, `idebate.org`, `debate.org` are a few examples.

E. Black, S. Modgil, and N. Oren (Eds.): TAFA 2013, LNAI 8306, pp. 16–31, 2014.

no facilities to reuse arguments and debates, although recent initiatives can help overcome this [6]; and 4) they stop short of reasoning with the debate data and votes/opinions, yielding very simplistic and naïve outcomes.

1.1 The Envisioned Online Debating System

Argumentation theory grounds debates in solid logical foundations and has in fact been shown to be applicable in a multitude of real-life situations [18]. *Social Abstract Argumentation* [16] in particular provided the theoretical foundations on which to build an ODS that gives deeper meaning to online debates, in a more robust, flexible, pervasive and interesting fashion than those currently available. In fact, it has already been used and extended in a prototype business directory allowing users to formally discuss and rank businesses [9], giving customers better control over who they hire.

Social Abstract Argumentation addresses many issues important to ODSs. It does away with the two-sides, one-winner approach typical of current systems. Instead, any user can propose any argument at any time. This yields a much more flexible debate structure, making it easier for users to get engaged and participate. The system also reasons and provides outcomes at the argument level at which users are now allowed to vote. The finer granularity makes outcomes more interesting, detailed and insightful.

When engaging in a debate, users always propose arguments for specific purposes, like making a claim central to the issue being discussed, or defeating arguments supporting an opposing claim. Thus, the envisioned ODS can allow users to formally describe an abstract argument, capable of attacking other arguments, simultaneously with its natural language (or image, video, link, etc.) representation. Therefore, the formal specification of arguments and attacks becomes a natural by-product of the users' intent when proposing new arguments. To make this process as painless and easy as possible, and enable more people to participate, no particularly deep knowledge (such as logics) can be required.

It is natural that a new argument might attack a previously proposed argument - indeed, that was likely the object of its creation. However, it is also possible that an older argument attacks the new argument as well. Therefore, the ODS should allow users to add this new attack relation formally to the system.

Those users who do not wish to engage in proposing arguments or attacks, for whatever reason, should also be accommodated in the system through a less complex participation scheme. Thus, in the ODS, users may simply read the arguments in natural language (or image, video, link, etc.) and formally state whether they agree with them. This induces a voting mechanism similar to what is found in current ODSs. There are alternatives, such as having argument's social trustworthiness be based on people's opinion's of who proposed it. Voting on arguments was chosen over these alternatives since it is the closest to current ODSs, and thus offers the path of least resistance. It is the role of the ODS to continuously provide an up to date view of the outcome of the debate e.g. by assigning value to each argument that somehow represents its social strength, taking the structure of the argumentation framework (arguments and attacks) and the votes into account. A nice GUI e.g. depicting arguments with a size and/or color proportional to these values would make the debate easier to follow, bringing forward relevant (socially) winning arguments, while downgrading unsound, unfounded (even troll) arguments. So that users may understand and follow a debate, small changes in the underlying argumentation framework and its social feedback (i.e. votes) should result in small changes to the formal outcome of the debate. If a single new vote entirely changes the outcome of a debate, users cannot gauge its evolution and trends, and are likely to lose interest.

Any debating system as the one envisioned must also ensure that a few crucial properties are satisfied. ODSs without the following properties are highly unlikely to be seriously adopted by online communities.

- There should always be at least one solution to a debate. The users must get *some* outcome for their effort. If the system is incapable of providing solutions to every debate, then there is too much risk involved in using it.
- There should always be at most one solution to a debate. Logicians and mathematicians find it perfectly natural for there to be multiple, or even infinite, solutions to a given problem. However, in a social context as far-reaching as the Internet, it is disingenuous to assume that the general user-base, which likely covers a large portion of the educational spectrum, shares these views with the same ease. It is very hard for someone who has invested personal effort into a debate to accept that all arguments are in fact true (in a multitude of models)!
- Argument outcomes should thus be represented very flexibly. In particular, to accurately represent the opinions of thousands of voting users, arguments should be valuated using degrees of acceptability, or gradual acceptability. Two-valued or three-valued semantics risk grossly underrepresenting much of the userbase.
- Formal arguments and attacks must be easy to specify. For example, assuming knowledge of first-order logic for specifying structured arguments [13] would alienate many potential users when the present goal is to include as many as possible. Moreover, simpler frameworks turn implementing and deploying such a system in different contexts (web forums, blogs, social networks, etc) much easier.

The above properties have been studied in the context of Social Abstract Argumentation, which uses abstract arguments in the sense of Dung [7], but has argument outcomes take values in the $[0, 1] \subseteq \mathbb{R}$ interval. A "well-behaved" family of semantics is known to guarantee the existence of outcomes, whereas uniqueness of outcomes has been proven for specific semantics.

1.2 Contribution

Despite the interesting properties of social abstract argumentation, it is apparent that not all attacks bear the same weight. Some attacks might have an obvious logical foundation (e.g. undercuts or rebuts), thus gaining trust from the more perceptive users. Other attacks might be less obvious or downright senseless, especially in open online contexts, making users doubt or wish to discard them.

Thus, extending the ability to vote to attacks, already suggested in [16,18], becomes eminently desirable. Not only does voting on attacks more accurately represent a crowd's opinion in a variety of situations, but it also allows the ODS to self-regulate by letting troll-attacks be "downvoted" to irrelevance. Following this view, recent formalisms have incorporated the notion of attack weights [11,8].

In this work, social abstract argumentation is extended with votes on attacks, and the properties that hold in social argumentation investigated.

This paper is organised as follows. Sect. 2 develops the extension to social abstract argumentation and studies concrete semantics. Sect. 3 provides a concrete example highlighting the role of votes on attacks. Sect. 4 covers related work, and Sect. 5 concludes.

2 Extended Social Argumentation Frameworks

This section will present ESAFs as an extension of social argumentation frameworks [16] by adding votes to attacks, besides arguments. We refer to these votes as the *social support* of the respective argument or attack.

Extended social argumentation frameworks, which build on social argumentation frameworks from [16], have an added parameter for votes on attacks; votes on arguments were already a feature of SAFs.

Definition 1 (Extended social argumentation frameworks). *An* extended social argumentation framework *is a 4-tuple* $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$*, where*

- $\mathcal{A}$ *is the set of arguments,*
- $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary attack relation between arguments,*
- $V_{\mathcal{A}} : \mathcal{A} \to \mathbb{N} \times \mathbb{N}$ *stores the crowd's pro and con votes for each argument.*
- $V_{\mathcal{R}} : \mathcal{R} \to \mathbb{N} \times \mathbb{N}$ *stores the crowd's pro and con votes for each attack.*

Notation 1 *Let* $\mathcal{R}^{-}(a) \triangleq \{a_i \in \mathcal{A} : (a_i, a) \in \mathcal{R}\}$ *be the set of direct attackers of an argument* $a \in \mathcal{A}$*. Let also* $V_{\mathcal{A}}^{+}(a) \triangleq x$ *and* $V_{\mathcal{A}}^{-}(a) \triangleq y$ *whenever* $V_{\mathcal{A}}(a) = (x, y)$*. Votes on attacks are handled similarly with* $V_{\mathcal{R}}$*.*

Following the approach of [16], semantic frameworks are used to aggregate operators representing the several parametrisable components of a semantics:

- An operation to obtain the combined strength of an argument's attackers. This value should be computed by aggregating together their individual strengths into a single, stronger value.
- An operation to restrict an argument's attack strength by the respective attack's social support. In an attack, the attacker can never be stronger than its social support, nor stronger than the attack's own social support.
- An operation to restrict an argument's social support by the value of its aggregated attackers. Notice that it would be socially unacceptable for an argument's final value to be above what was originally its social support.

- An operation that computes a limiting factor from a given attack strength. This limiting factor can then be used to restrict an argument's original strength with the above operator.
- Computing social support values from pro/con votes cast by the community.

All the parametrisable components of a semantics, matching the operations mentioned above, are captured in the following definition.

Definition 2 (Semantic Framework). *A semantic framework is a 6-tuple* $\langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *where:*

- L *is a totally ordered set with top and bottom elements* $\top$, $\bot$, *containing all possible valuations of an argument.*
- $\curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}} : L \times L \rightarrow L$, *are two binary algebraic operations used to restrict strengths to given values.*
- $\curlyvee : L \times L \rightarrow L$, *is a binary algebraic operation on argument valuations used to combine or aggregate valuations and strengths.*
- $\neg : L \rightarrow L$ *is a unary algebraic operation for computing a restricting value corresponding to a given valuation or strength.*
- $\tau : \mathbb{N} \times \mathbb{N} \rightarrow L$ *is a function that aggregates positive and negative votes into a* social support *value.*

Notation 2 *As a useful shortcut, let* $\tau(a) \triangleq \tau(V_{\mathcal{A}}^{+}(a), V_{\mathcal{A}}^{-}(a))$ *and* $\tau((a_1, a_2)) \triangleq \tau(V_{\mathcal{R}}^{+}((a_1, a_2)), V_{\mathcal{R}}^{-}((a_1, a_2)))$. *Let* $R = \{x_1, x_2, ..., x_n\}$ *be a multiset of elements of* L. *Then, with a small abuse of notation:*

$$\bigcurlyvee_{x \in R} x \triangleq (((x_1 \curlyvee x_2) \curlyvee ...) \curlyvee x_n)$$

Notice also that the valuation set L of arguments is parametrisable. L could be $[0, 1] \subseteq \mathbb{R}$, but it could also be any finite, countable or uncountable set of values such as booleans, colours, textures, or any other set that is deemed appropriate for users of the final application, so long as it is totally ordered.

The heart of the semantics is in the definition of a model, which combines the operators of a semantic framework $\mathcal{S}$ into a system of equations, one for each argument, that must be satisfied.

Definition 3 (Model). *Let* $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be a social argumentation framework,* $\mathcal{S} = \langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *be a semantic framework. A* $\mathcal{S}$*-model of* F *is a total mapping* $M : \mathcal{A} \rightarrow L$ *such that for all* $a \in \mathcal{A}$,

$$M(a) = \tau(a) \curlywedge_{\mathcal{A}} \neg \bigcurlyvee_{a_i \in \mathcal{R}^{-}(a)} (\tau((a_i, a)) \curlywedge_{\mathcal{R}} M(a_i))$$

The value assigned to an argument a by model M, or $M(a)$ is called the *valuation* of a under model M.

A model M is a solution to the equation system with one equation of the form in Definition 3 for each argument. An alternative interpretation is that models are fixpoints of the assignments induced by the equations.

We now analyse the equation to facilitate its understanding.

If an argument a_1 attacks another argument a_2, then the strength of the attack is the valuation of the attacking argument a_1 reduced by the social support of the attack: no argument's attack is stronger than either its own valuation or the social support of the attack itself. We use $\curlywedge_{\mathcal{R}}$ to restrict these values.

$$\tau\left(\left(a_1, a_2\right)\right) \curlywedge_{\mathcal{R}} M\left(a_1\right)$$

Since an argument may have multiple attackers, all of their attack strengths must be aggregated to form a stronger combined attack value, using operator $\curlyvee$.

$$\bigcurlyvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left(\left(a_i, a\right)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

The above equation results in a combined attack strength that must be turned into a restricting value using the $\neg$ operator.

$$\neg \bigcurlyvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left(\left(a_i, a\right)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

In a social context where the crowd has given its opinion of an argument a, it is clear that a's valuation should never turn out higher than a's social support $\tau(a)$. Thus, an argument's valuation is given by restricting $\tau(a)$ with the value of the aggregated attack using the final operator $\curlywedge_{\mathcal{A}}$.

$$\tau(a) \curlywedge_{\mathcal{A}} \neg \bigcurlyvee_{a_i \in \mathcal{R}^-(a)} \left(\tau\left(\left(a_i, a\right)\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right)$$

2.1 Specific Semantics

Fully specifying semantics means that any derivable properties are likely to apply only to that particular case. For that reason, this section starts by restricting a semantic framework in limited, sensible, even intuitively desirable ways, and investigating what derives therefrom. The following definition formalises well-behavedness in a social context.

Definition 4 (Well-behaved semantic frameworks). *A semantic framework* $\mathcal{S} = \langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *is well-behaved if*

- $\neg$ *is antimonotonic, continuous,* $\neg\bot = \top$, $\neg\top = \bot$ *and* $\neg\neg a = a$*;*
- $\curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}$ *are continuous, commutative, associative, monotonic w.r.t. both arguments and* $\top$ *is their identity element;*
- $\curlyvee$ *is continuous, commutative, associative, monotonic w.r.t. both arguments and* $\bot$ *is its identity element;*

Some important notions guided Definition 4. Continuity of operators guarantees small changes in the social inputs result in small changes in the models. Were this not the case, outcomes of debates would be very unstable, hard to follow and more easily exploited by trolls. The remaining algebraic properties simply state that the order in which arguments are attacked makes no difference; that an argument's valuation is proportional to its crowd support; that aggregated attacks are proportional to the attacking arguments; and so forth.

From these simple restrictions it is already possible to derive some important results regarding existence of models under well-behaved semantics.

Theorem 3 (Existence of Models). *Let* $F = \langle \mathcal{A}, \mathcal{R}, V_{\mathcal{A}}, V_{\mathcal{R}} \rangle$ *be an extended social argumentation framework and* $\mathcal{S} = \langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *a well behaved semantics. Then* F *has at least one* $\mathcal{S}$*-model.*

Consider now the following concrete semantics, suitable for studying specific behaviours and for implementation. It is a generalisation of its social abstract argumentation counterpart [16].

Definition 5 (Product semantics). *Let* $\mathcal{S}^{\cdot}_{\epsilon} = \langle [0,1], \curlywedge^{\cdot}, \curlywedge^{\cdot}, \curlyvee^{\cdot}, \neg, \tau_{\epsilon} \rangle$ *be a semantic framework,* $x, y \in [0,1]$ *and*

- $x \curlywedge^{\cdot} y = x \cdot y$, *i.e. the product T-norm.*
- $x \curlyvee^{\cdot} y = 1 - (1-x) \cdot (1-y)$, *i.e. the T-conorm dual to the product T-norm.*
- $\neg x = 1 - x$
- $\tau_{\epsilon}(a) = \frac{V^{+}(a)}{V^{+}(a)+V^{-}(a)+\epsilon}$, *with* $\epsilon > 0$, *and similarly for attacks*

The τ_{ϵ} function used to compute the social support deserves special mention. It is a minor variation of a simple percentage. The reason for this modification will become apparent after the following example.

Example 1. Imagine a symmetric situation where two mutually attacking arguments, a and b have only received positive votes, as have their attacks. Figure 1 represents such a scenario, with the annotations indicating the social support of arguments and attacks, using τ_0.

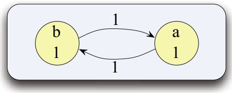

Fig. 1. Symmetric situation with mutually attacking arguments

Considering semantics $\mathcal{S}^{\cdot}_{0}$, it is easy to derive that the valuations of a and b are given by $M(a) = 1 - M(b)$ and $M(b) = 1 - M(a)$. Therefore, the system of equations that results from using $\tau_0(a) = \frac{V^{+}(a)}{V^{+}(a)+V^{-}(a)}$ is linearly dependent and results in infinitely many models with $M(a) + M(b) = 1$.

It turns out multiple models only arise in the extremely symmetrical situations depicted in Figure 1, with only positive votes. Similar odd-length cycles have a unique model. Furthermore, a single negative vote on any argument or attack is enough to break linear dependence and making uniqueness hold again.

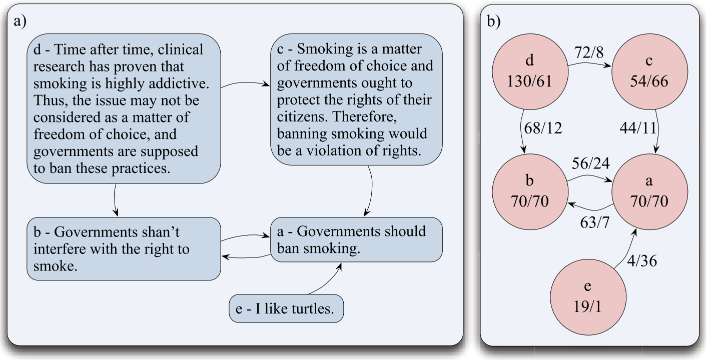

Fig. 2. Social Argumentation Framework: a) arguments and attacks; b) votes

This led to the introduction of the arbitrarily small $\epsilon > 0$ in Definition 5, making scenarios such as those of Figure 1 have unique models. The resulting single model has the property that $M(a) = M(b)$, which preserves the argumentation framework's symmetry in the outcome as well.

The situation when $\epsilon = 0$ can still be made sense of by taking the limit of ϵ as it goes to 0. In fact, $\lim_{\epsilon \to 0} M(a) = \lim_{\epsilon \to 0} M(b) = 0.5$, which is the model that best preserves the symmetry of the framework, as do the models when $\epsilon > 0$.

The product semantics is a well-behaved semantics, which means that the existence of a model is guaranteed as per Theorem 3. The following result provides some clarification in regards to the uniqueness of models.

Theorem 4 (Uniqueness of Models). *Let F be an ESAF such that $|\mathcal{R}^{-}(a)| \cdot V_{\mathcal{A}}(a) < 1$, for every $a \in \mathcal{A}$. Then, F has one and only one model under $\mathcal{S}_{\epsilon}^{\cdot}$.*

Additionally, we expect the result will hold for $\epsilon > 0$, without the $|\mathcal{R}^{-}(a)| \cdot V_{\mathcal{A}}(a) < 1$ condition, though this has not been proven yet.

3 Example

Consider a social interaction inspired by [21] where several participants, while arguing about the role of the government in what banning smoking is concerned, set forth the arguments and attack relations depicted in Fig. 2 a).

Despite the fact that these arguments are structurally different: a and b are unsupported claims, c and d contain multiple premises and a conclusion, while e, despite being rather consensual (who doesn't like turtles?), seems to be totally out of context and can hardly be seen as an attack on a (here, the attack by e on a is meant to represent a troll attack). Our goal is to show that ESAFs' level of abstraction allows meaningful arguments to be construed out of most participations – in fact, with suitable GUIs, arguments could even be built from videos, pictures, links, etc. – while the participation through voting will help deal

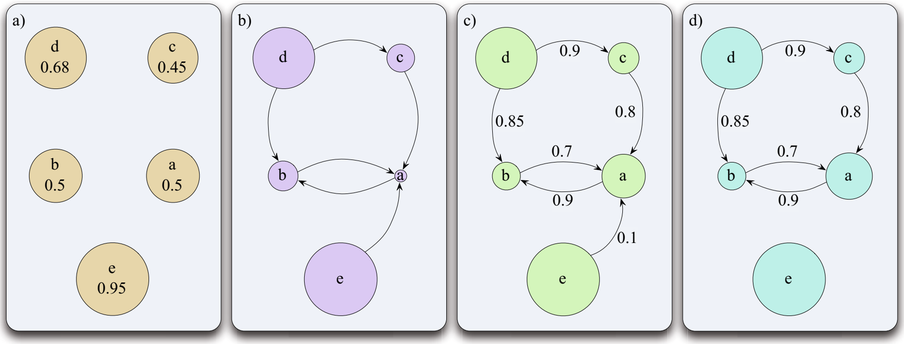

Fig. 3. Model of the Social Abstract Argumentation Framework considering: a) social support only; b) attacks but not their strength; c) attack strength; d) attack strength, without the attack from e to a

with mitigating the disturbing effect of unsound arguments and poorly specified (troll) attacks.

After a while, the arguments and attacks garner the pro/con votes depicted in Fig. 2 b). Arguments a and b obtain the same direct social support as expressed by the 70 *pro* and *con* votes. Meanwhile, a's attack on b is deemed stronger than its counterpart, judging from their votes. One might speculate that this is a consequence of a delivering a more direct message. Whereas argument c does not get much love from the crowd (a vote ratio of 54/66), its attack on a is still supported by the community (44/11). Perhaps initially there was a better sentiment towards c but the introduction of d, which amassed a decent amount of support itself (130/61), turned the odds against c. Both of d's attacks on b and c materialise to be strong enough, the former being slightly weaker (72/8 versus 68/12). Lastly, argument e received just a mere number of votes, most being positive (19/1). However, there seems to have been a significant effort from the users on discrediting the attack on a by e (4/36). Note that e is a perfectly legitimate argument. Indeed the crowd endorses the fondness for turtles – it's the attack, not the argument, that is not logically well-founded.

With the abstract argumentation framework and the votes on arguments and attacks in hand, we can turn our attention to the valuation of the arguments.

If we consider the social support of each argument, i.e. its value considering only the votes it obtained while ignoring attack relations, we obtain the following values:[2] $\tau_0(a) = 0.50$, $\tau_0(b) = 0.50$, $\tau_0(c) = 0.45$, $\tau_0(d) = 0.68$ and $\tau_0(e) = 0.95$, as depicted in Fig. 3 a) (where the size of each node is proportional to its value).

The original Social Abstract Argumentation semantics [16], which considers attacks between arguments but not the votes on attacks, assigns the following values to arguments: $M(a) = 0,02$, $M(b) = 0,16$, $M(c) = 0,14$, $M(d) = 0,68$ and $M(e) = 0,95$, as depicted in Fig. 3 b). As expected, d and e retain their initial social support values, since they are not attacked, while the remaining

---

[2] We will consider the Product Semantics as in Def.5, with $\epsilon = 0$.

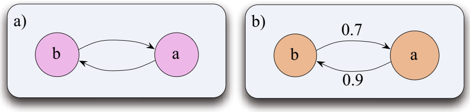

Fig. 4. a) symmetric attacks without attack strength; b) symmetric attacks with attack strength

arguments see a decrease in their social support value. Argument a decreases the most while b and c maintain a reasonable fraction of their initial strength. Since two of a's attackers – b and c – are attacked by d, which is a non-attacked argument with strong social support, their value is weakened, so their effect on a is lessened. Thus, we can conclude that the main cause for the downfall in a's value is e's attack.

We can now turn our attention to the model proposed in this paper, which also takes votes on attacks into consideration, and assigns the following values to arguments: $M(a) = 0,35$, $M(b) = 0,14$, $M(c) = 0,17$, $M(d) = 0,68$ and $M(e) = 0,95$, as depicted in Fig. 3 c). The value assigned to a by the model increases from 0.02 to the more plausible level of 0.34, mostly due to e's weakened capability to attack a. Indeed, the crowd's overwhelming con votes on the (troll) attack of e on a essentially neutralised it. To confirm, we compare it with the model obtained if the attack from e to a was simply removed, depicted in Fig. 3 d), whose valuations of $M(a) = 0,39$, $M(b) = 0,14$, $M(c) = 0,17$ and $M(d) = 0,68$ are very similar to those obtained in the presence of the very weakened attack from e to a, which allows us to conclude for the success of ESAF's in discounting attacks that are socially deemed unsound, such as troll attacks. Since the weights of the remaining attacks are relatively high and also close to each other at the same time, their impact is somewhat minimal.

For convenience, all values regarding the figures are listed in Table 1.

One last remark worth noting is that the inclusion of votes on attacks enables the model to break stand-stills when attacks are not equally strong. Figure 4 a) and b) show how two arguments enjoying the same direct social support, arguments a and b in our example, become distinguishable once the weights of attacks are taken into account.

4 Related Work

4.1 Gabbay's Equational Approach to Argumentation

Here we compare with [10,11], and how some of their proposed semantics can be captured by ESAFs. Gabbay's equational approach to argumentation [10] is a very general formalism that deserves its differences to ESAFs highlighted.

It uses a fixed domain of $[0, 1]$, unlike a generic L. For every argument a, whose attackers are $x_1, ..., x_n$, the semantics and models are given by $\mathbf{h}_a$:

$$\mathbf{f}(a) = \mathbf{h}_a(\mathbf{f}(x_1), ..., \mathbf{f}(x_n))$$

Table 1. Model of the Social Abstract Argumentation Framework considering: a) social support only; b) attacks but not their strength; c) attack strength; d) attack strength, without the attack from e to a

Argument	Fig. a	Fig. b	Fig. c	Fig. d
a	0.50	0.02	0.35	0.39
b	0,50	0,16	0,14	0.14
c	0,45	0,14	0,17	0.17
d	0,68	0,68	0,68	0.68
e	0,95	0,95	0,95	0.95

The intuition is that a's model depends on the models of its attackers $x_1 \dots x_n$ according to formula $\mathbf{h}_a$, entirely like ESAFs, implying models are fixpoints.

Because different $\mathbf{h}_a$ can be used for each argument, the equational approach can be seen as more general than ESAFs (cf. Definition 3). It is nonetheless interesting to note that all proposed semantics use the same formula for all arguments, except for the *suspect semantics* whose desired meaning is unclear.

The complete generality of the equational approach is appealing because any formula can be written, so presumably any other formalism is capturable by it. Ultimately, however, this generality results in very few derivable properties. In fact, [10] focuses mostly on studying specific semantics and possible extensions to those semantics rather than on properties of the general system.

The comparison that follows will focus on subsequent work [11] which is of a more social nature than [10]. It allows initial weights on attacks, which fill the same role as social support in ESAFs.

The semantics proposed therein make use of $V_0(a)$ as the initial value for arguments and of $\xi(a_1, a_2)$ for attacks. They are called inverse semantics and maximum semantics, defined below:

$$\mathbf{f}_{inv}(a) = V_0(a) \cdot \prod_{a_i \in \mathcal{R}^-(a)} \left(1 - \xi(a_i, a)\, \mathbf{f}_{inv}(a_i)\right)$$

$$\mathbf{f}_{max}(a) = V_0(a) - \max_{a_i \in \mathcal{R}^-(a)} \left\{\xi(a_i, a)\, \mathbf{f}_{max}(a_i)\right\}$$

These semantics can be captured in ESAFs, as the following results show.

Proposition 1. *Eq_{inv} can be represented using ESAFs when initial values are rational.*

Proof. *First, it will be necessary to show that the initial values $V_0(a), V_{\mathcal{R}}(a_1, a_2) \in \mathbb{Q}$ can be represented as votes. The following lemma shows that they can.*

Lemma 1. *Let $x \in \mathbb{Q}$ such that $0 \leq x \leq 1$. Then there is $y, z \in \mathbb{N}$ such that $x = \frac{y}{y+z}$ and $\frac{y}{y+z}$ is irreducible.*

Proof. *From $x \in \mathbb{Q}$ it follows that there is an irreducible fraction $x = \frac{a}{b}$. From $0 \leq x \leq 1$ it follows that $a \leq b$. Then, let $y = a$ and $z = b - a \Leftrightarrow z + a = b \Leftrightarrow y + z = b$.*

Back to the proposition. Consider the product semantics defined above as $\mathcal{S}^{\cdot}_{\epsilon} = \langle [0,1], \curlywedge^{\cdot}, \curlywedge^{\cdot}, \curlyvee^{\cdot}, \neg, \tau \rangle$. *Then, since* $V_0(a), V_{\mathcal{R}}(a_1, a_2) \in \mathbb{Q}$, *by the lemma, there are* $V_{\mathcal{A}}^{+}(a)$, $V_{\mathcal{A}}^{-}(a)$, $V_{\mathcal{R}}^{+}(a_1, a_2)$, $V_{\mathcal{R}}^{-}(a_1, a_2) \in \mathbb{N}$ *such that* $\tau(a) = V_0(a)$ *and* $\tau(a_1, a_2) = V_{\mathcal{R}}(a_1, a_2)$.

T-norms and their dual T-conorms are distributive, and the generalised de Morgan laws apply. This justifies the first step in the following derivation.

$$\begin{aligned} M(a) &= \tau(a) \curlywedge^{\cdot} \neg \mathop{\curlyvee^{\cdot}}_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(\tau\left(a_i, a\right) \curlywedge^{\cdot} M\left(a_i\right)\right) \\ &= \tau(a) \curlywedge^{\cdot} \mathop{\curlywedge^{\cdot}}_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(\neg\left(\tau\left(a_i, a\right) \curlywedge^{\cdot} M(a_i)\right)\right) \\ &= V_0(a) \cdot \prod_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(1 - \xi\left(a_i, a\right) \cdot M(a_i)\right) = \boldsymbol{f}_{inv}\left(a\right) \end{aligned}$$

While it was not necessary for Eq_{inv}, two different $\curlywedge$ operators are needed to subsume Eq_{max}, since it uses both subtraction and multiplication as restrictive operations.

Proposition 2. Eq_{max} *can be represented using ESAFs when initial values are rational.*

Proof. Let $\mathcal{S}^{max} = \langle L, \curlywedge_{\mathcal{A}}, \curlywedge_{\mathcal{R}}, \curlyvee, \neg, \tau \rangle$ *such that:* $L = [0,1]$; $V_{\mathcal{A}}(a)$ *and* $V_{\mathcal{R}}(a_1, a_2)$ *are handled as in Proposition 1;* $l_1 \curlywedge_{\mathcal{A}} l_2 = l_1 - l_2$, *arithmetic subtraction;* $l_1 \curlywedge_{\mathcal{R}} l_2 = l_1 \cdot l_2$, *arithmetic multiplication;* $l_1 \curlyvee l_2 = \max\{l_1, l_2\}$, *maximum aggregation operation; and* $\neg l_1 = l_1$.

Notice that $\mathcal{S}^{max}$ *is* not *a well-behaved framework since* $\curlywedge_{\mathcal{A}}$, *or subtraction, is not commutative. The de Morgan laws also do not apply.*

$$\begin{aligned} M(a) &= \tau(a) \curlywedge_{\mathcal{A}} \neg \mathop{\curlyvee}_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(\tau\left(a_i, a\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right) \\ &= V_0(a) \curlywedge_{\mathcal{A}} \neg \mathop{\curlyvee}_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(\xi\left(a_i, a\right) \curlywedge_{\mathcal{R}} M\left(a_i\right)\right) \\ &= V_0(a) - \mathop{\curlyvee}_{a_i \in \mathcal{R}^{\text{-}}(a)} \left(\xi\left(a_i, a\right) \cdot M\left(a_i\right)\right) \\ &= V_0(a) - \max_{a_i \in \mathcal{R}^{\text{-}}(a)} \left\{\xi\left(a_i, a\right) M\left(a_i\right)\right\} = \boldsymbol{f}_{max}\left(a\right) \end{aligned}$$

These propositions serve as evidence that several semantics can be captured using ESAFs. Some accommodate the notion of well-behaved semantics naturally, but others do not. This could stem from how closely related with multivaluedness the semantics are. For example, $\mathbf{f}_{max}$ is very two-valued at its core. It actually coincides with stable extensions [7] when $V_0(a) = 1$ for all $a \in \mathcal{A}$, $\xi\left(a_1, a_2\right) = 1$ for all $(a_1, a_2) \in \mathcal{R}$, and an argument is taken to be accepted when $\mathbf{f}_{max}\left(a\right) = 1$. In fact, it follows that $\mathbf{f}_{max}\left(a\right) \in \{0, 1\}$! The product semantics is more tightly connected with the valuations being $[0,1] \subseteq \mathbb{R}$, and that appears to be part of the reason for its well-behavedness.

4.2 Graduality in Argument Valuations

Other proposals have broached the subject of graduality, or multi-valuedness, in argumentation, which is so important in large-scale online debates. Graduality in argumentation was studied in [5], proposing $\mathbb{R}$ as the domain of argument valuations, and aggregation and reduction operators applied in an explicit equational way, not unlike the present work. However, no social context is assumed, and so all attacks and arguments are perfectly sound by default. Their initial values, or social support, are 1, not a function of a crowd's opinion.

As seen in Section 4.1, [10] uses a very flexible equational method with argument valuations over $[0, 1]$. In [11], that approach, extended with voting, is applied to the context of merging several agents' different perceptions of a single debate into a unified framework with a social outcome. Each agent votes positively on arguments and attacks it agrees with locally, and votes negatively on those that he doesn't agree with. This can be seen as a restricted use of the voting mechanism of ESAFs.

4.3 Social Contexts in Argumentation

Whereas ESAFs incorporate the social aspect of argumentation using voting, alternative proposals explicitly model properties of the social context in which debates are carried out.

In [20], arguments have values (e.g. free-market, human rights, family), and a specific audience will order values according to personal preferences. By incorporating a probabilistic model of the audience's preferences, [20] estimates the probability that each argument will convince the audience of a certain proposition. The arguments more likely to convince the audience can be chosen first.

Along a similar line, arguments can be related to topics on which certain people are experts [15]. This allows votes on attacks to have different weights depending on the expertise of the voter in the related topics. Furthermore, if a debate is controversial, i.e. with a fickle outcome balancing on a knife's edge, [15] draws some preliminary considerations on computing the best expert to call forward to propose the next argument, hopefully settling the debate.

4.4 Applications of Argumentation to the Web 2.0

Some recent applications have focused on applying argumentation theory to Twitter as a source of abundant social information in the form of concise comments. These systems are very close to the intended application of SAFs [16] and ESAFs. In fact, using the notion of arguments and attacks, it is possible to automatically mine Twitter for arguments and attacks between them [14]. These result in argument trees such as those found in classical dialogue-based argumentation.

A different approach, more reliant on users, is to allow them to annotate their tweets with agreement or disagreement towards a particular subject [12]. The new annotations are used to create arguments and attacks. Given a stream given by a hashtag, it is thus possible to induce an argumentation framework that represents the ongoing discussion, and obtain formal outcomes. This functionality has been implemented and tested.

The above implementations are showing how argumentation theory can be applied to the Web 2.0, although they currently do not feature crowd voting.

5 Conclusions

In closing, the work on *Extended Social Argumentation Frameworks* takes another step towards capturing the essence of social debates. It builds on the theoretical foundations laid by [16] that provided debates with formal and justifiable outcomes. In this work the social notion of attack strengths has been introduced via incorporating votes on attacks. A new family of semantics is introduced for the new framework and illustrated by the means of an example. Certain semantics suggested in [10,11] are proven to be special cases of ESAF.

The originality of the proposed framework lies in its practicality and generality. The semantics for ESAFs can be tailored in different ways to meet the needs and expectations of varied applications and user groups. Whereas we focused more on gradual valuations, in a scenario where a clear decision is to be taken, it might make more sense to use a semantics with a family of operators that results in a classical in/out approach as in Dung-like argumentation frameworks.

Furthermore the framework can be extended in multiple ways.

Some authors have advocated the addition of a support relation between arguments (e.g. [1,3]). Whereas there has been a debate regarding the adequacy of such relation – some argue that since arguments are accepted by default, any support should take the form of an attack on its attackers – its incorporation into ESAF's might prove beneficial, and certainly worth future investigation.

Allowing votes on attacks resembles the abstract resolution semantics in [2]. Even though it's not exactly how we envision our system, it might be an interesting idea to explore a possible extension by admitting arguments with internal structure, taking into careful consideration the observation that properties of abstract resolution semantics are not always preserved by instantiations [19], which could result in the loss the list of desired properties of our framework.

The work in [4] allows the automatic detection and generation of the abstract arguments from natural language. Such a framework can be utilized hand in hand with ESAFs to capture the semantics of unstructured social debate platforms.

Another possible extension is to consider attacks on the attack relations themselves. Prior work on this topic such as [17] builds on [7] by following the generic accepted/defeated approach regarding the arguments. In this regard ESAFs can improve the novelty of the work by the flexible evaluation mechanism.

Finally, it should be noted that the work on ESAFs will proceed in near future via articulating the capabilities of the framework proposed in this paper in capturing the semantics of other existing approaches in the field.

Acknowledgments. We would like to thank the anonymous reviewers for their insightful comments and suggestions. J. Leite was partially supported by FCT project ERRO – PTDC/EIA-CCO/121823/2010.

References

1. Amgoud, L., et al.: On bipolarity in argumentation frameworks. Int. J. Intell. Syst. 23(10), 1062–1093 (2008)
2. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artif. Intell. 171(10-15), 675–700 (2007)
3. Boella, G., et al.: Support in abstract argumentation. In: Procs. of COMMA 2010. FAIA, vol. 216. IOS Press (2010)
4. Cabrio, E., Villata, S.: Natural language arguments: A combined approach. In: Procs. of ECAI 2012. FAIA, vol. 242. IOS Press (2012)
5. Cayrol, C., Lagasquie-Schiex, M.C.: Graduality in argumentation. J. Artif. Intell. Res. 23, 245–297 (2005)
6. Chesñevar, C., et al.: Towards an argument interchange format. Knowledge Eng. Review 21(4), 293–316 (2006)
7. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–358 (1995)
8. Dunne, P., et al.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artif. Intell. 175(2), 457–486 (2011)
9. Evripidou, V., Toni, F.: Argumentation and voting for an intelligent user empowering business directory on the web. In: Krötzsch, M., Straccia, U. (eds.) RR 2012. LNCS, vol. 7497, pp. 209–212. Springer, Heidelberg (2012)
10. Gabbay, D.M.: Equational approach to argumentation networks. Argument & Computation 3(2-3), 87–142 (2012)
11. Gabbay, D., Rodrigues, O.: A numerical approach to the merging of argumentation networks. In: Fisher, M., van der Torre, L., Dastani, M., Governatori, G. (eds.) CLIMA XIII 2012. LNCS (LNAI), vol. 7486, pp. 195–212. Springer, Heidelberg (2012)
12. Gabbriellini, S., Torroni, P.: Large scale agreements via microdebates. In: Procs. of AT 2012. CEUR Workshop Proceedings, vol. 918 (2012)
13. García, A., Dix, J., Simari, G.: Argument-based logic programming. In: Rahwan, I., Simari, G.R. (eds.) Argumentation in Artificial Intelligence, pp. 153–171. Springer (2009)
14. Grosse, K., Chesñevar, C.I., Maguitman, A.G.: An argument-based approach to mining opinions from Twitter. In: Procs. of AT 2012. CEUR Workshop Proceedings, vol. 918 (2012)
15. Kontarinis, D., et al.: Picking the right expert to make a debate uncontroversial. In: Procs. of COMMA 2012. FAIA, vol. 245. IOS Press (2012)
16. Leite, J., Martins, J.: Social abstract argumentation. In: Procs. of IJCAI 2011. IJCAI/AAAI (2011)
17. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. 173(9-10), 901–934 (2009)
18. Modgil, S., et al.: The added value of argumentation. In: Ossowski, S. (ed.) Agreement Technologies. Law, Governance and Tech. Series, vol. 8, pp. 357–403. Springer (2013)

19. Modgil, S., Prakken, H.: Resolutions in structured argumentation. In: Procs. of COMMA 2012. FAIA, vol. 245, IOS Press (2012)
20. Oren, N., Atkinson, K., Li, H.: Group persuasion through uncertain audience modelling. In: Procs. of COMMA 2012. FAIA, vol. 245. IOS Press (2012)
21. Walton, D.: Argumentation theory: A very short introduction. In: Rahwan, I., Simari, G.R. (eds.) Argumentation in Artificial Intelligence, pp. 1–22. Springer (2009)