Constructing and Evaluating Bipolar Weighted Argumentation Frameworks for Online Debating Systems

Andrea Pazienza, Stefano Ferilli, and Floriana Esposito

Dipartimento di Informatica – Università di Bari
name.surname@uniba.it

Abstract. Discussions on social Web platforms carry a lot of information which is more and more difficult to analyze. Given a virtual community of users that discuss a particular topic of interest, an important task is to extract a model of the whole debate in order to automatically evaluate what are the most reliable claims. This paper proposes to approach this task using abstract argumentation, and define a new argument system, called Bipolar Weighted Argumentation Framework. It is able to capture all the useful information from a discussion thread, including the strength of positive (i.e., supports) and negative (i.e., attacks) relations between arguments. It also provides a way to assess an acceptability degree for each argument by means of the strength propagation of indirect relations ending to it, and a strategy to build such a framework from an online debate with a hierarchical structure. A model obtained from a real life discussion (a Reddit thread) is discussed and qualitatively evaluated.

1 Introduction

People in all societies argue, discuss, and debate not only to convince others of their own opinions, but because they want to explore the differences between their own understanding and the conceptualizations of others, and learn from them. Being one of the primary intellectual activities of the human mind, debating therefore naturally involves a wide range of conceptual capabilities and activities, ones that have only in part been studied from a computational perspective. One of the most influential computational models of argument was presented by Dung's Argumentation Frameworks [5] (in short, AF), which is roughly a directed graph where the vertices are the abstract arguments and the directed edges correspond to attacks between them. As there are no restrictions on the attack relation, cycles, self–attackers, and so on, are all allowed. Arguments do not have any particular structure and the precise conditions for their acceptance are defined by the semantics. Semantics produce acceptable subsets of the arguments, called *extensions*, that correspond to various positions one may take based on the available arguments. However, within the argumentation process, the construction of proper AFs can cause much more concern than expected. Therefore,

basic AFs may not necessarily be the best target systems for the instantiation. In order to address this problem, a research direction is to extend AFs by equipping relations with more expressive concepts such as support relations, giving rise to Bipolar Argumentation Frameworks (BAFs) [3], and weighted attacks, giving rise to Weighted Argumentation Frameworks (WAFs) [6]. Nevertheless, a single framework that takes both advantages of BAFs and WAFs is still lacking.

Different levels of agreement or disagreement may exist in real discussions with respect to an argument, whereby AFs, BAFs and WAFs may be unable to model certain situations. Moreover, there may exist indirect attack/support relations that in the aforementioned frameworks are not quantified.

In the present work we propose a new AF extension, called Bipolar Weighted Argumentation Framework (BWAF), which combines all the properties of BAFs and WAFs. In order to refashion the interactions that may subsist between arguments, BWAFs are able to express weighted directed relations and to compute the propagated strength of indirect relations. We also provide a general strategy to build such a model from an Online Debating System (ODS). We use Text Similarity and Sentiment Analysis techniques to identify weighted attack/support relations. Finally, we show the use of this framework in a real discussion thread taken from Reddit.

This paper is organized as follows. Section 2 briefly recalls the background on abstract argumentation and subsequent generalizations of Dung's Framework. Section 3 introduces our proposal, i.e. BWAF with its useful properties. Section 4 describes a strategy to build a BWAF from an ODS and shows its application to a Reddit discussion. Finally, Section 5 concludes the paper.

2 Background and Related Work

Let us start by providing the basics of Abstract Argumentation.

Definition 1 *An **AF** is a pair $F = \langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a finite set of arguments and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. Given $a, b \in \mathcal{A}$, the relation $a\mathcal{R}b$ means that a attacks b.*

An argumentation semantics is the formal definition of a method ruling the argument evaluation process. The most basic concepts shared by all argumentation semantics in the literature are *conflict-freeness* and *defense*.

Definition 2 *Let $F = \langle \mathcal{A}, \mathcal{R} \rangle$ be an AF, and $S \subseteq \mathcal{A}$:*

- *S is* conflict-free *if $\nexists a, b \in S$ s.t. $a\mathcal{R}b$;*
- *$a \in \mathcal{A}$ is* defended by *S if $\forall b \in \mathcal{A}$: $b\mathcal{R}a \Rightarrow \exists c \in S$ s.t. $c\mathcal{R}b$;*
- *$f_F \colon 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ s.t. $f_F(S) = \{a \mid a$ is defended by $S\}$ is called the* characteristic function *of F;*
- *S is* admissible *if S is conflict-free and S is defended by itself, i.e. $\forall a \in S, \forall b \in \mathcal{A}$: $b\mathcal{R}a \Rightarrow \exists c \in S$ s.t. $c\mathcal{R}b$.*

Then, standard acceptability semantics, introduced by Dung [5], characterize admissible sets of arguments:

Definition 3 *Let $F = \langle \mathcal{A}, \mathcal{R} \rangle$ be an AF and $S \subseteq \mathcal{A}$ be an admissible set. Then, S is a:*

- complete *extension iff $S = f_F(S)$;*
- grounded *extension iff S is the $\subseteq$-minimal complete extension;*
- preferred *extension iff S is a $\subseteq$-maximal complete extension;*
- stable *extension iff $\forall a \in \mathcal{A}, a \notin S, \exists b \in S$ s.t. $b \mathcal{R} a$.*

A Bipolar AF (*BAF*) [3] is an extension of Dung's AF in which two kinds of interactions between arguments are possible: the attack relation and the support relation. These two relations are independent (i.e., the support relation is not defined using the attack relation) and lead to a bipolar representation of the interaction between arguments.

Definition 4 *A **BAF** is a triplet $B = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$, where $\mathcal{A}$ is a set of arguments, $\mathcal{R}_{att}$ is a binary relation on $\mathcal{A}$ called attack relation and $\mathcal{R}_{sup}$ is another binary relation on $\mathcal{A}$ called support relation. Given $a, b \in \mathcal{A}$, $a \mathcal{R}_{att} b$ (resp., $a \mathcal{R}_{sup} b$) means that a attacks b (resp., a supports b).*

In BAFs, new kinds of attack emerge from the interaction between the direct attacks and the supports: there is a *supported attack* iff there is a sequence of supports followed by one attack, while, there is an *indirect attack* iff there is an attack followed by a sequence of supports. Taking into account sequences of supports and attacks leads to the following definitions applying to sets of arguments.

Definition 5 *Let $B = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ be a BAF. A set $S \subseteq \mathcal{A}$* set-attacks *$b \in \mathcal{A}$, iff there exists a supported attack or an indirect attack for b from an element of S. A set $S \subseteq \mathcal{A}$* set-supports *$b \in \mathcal{A}$, iff there exists a sequence of supports for b from an element of S. A set $S \subseteq \mathcal{A}$* defends *$a \in \mathcal{A}$, iff for each argument $b \in \mathcal{A}$, if $\{b\}$ set-attacks a, then b is set-attacked by S.*

In the following, we define the semantics for acceptability in BAFs [3].

Definition 6 *Let $B = \langle \mathcal{A}, \mathcal{R}_{att}, \mathcal{R}_{sup} \rangle$ be a BAF and $S \subseteq \mathcal{A}$. Then, S is:*

- conflict-free *iff $\nexists a, b \in S$ s.t. $\{a\}$ set-attacks b;*
- safe *iff $\nexists b \in \mathcal{A}$ s.t. S set-attacks b and either S set-supports b or $a \in S$;*
- d-admissible *iff S is conflict-free and $\forall a \in S, a$ is defended by S;*
- s-admissible *iff S is safe and $\forall a \in S$, a is defended by S;*
- c-admissible *iff S is d-admissible and closed for $\mathcal{R}_{sup}$.*
- *a* d-preferred *(resp.* s-preferred, c-preferred*) extension is a $\subseteq$-maximal d-admissible (resp. s-admissible, c-admissible) subset of $\mathcal{A}$;*
- *a* stable extension *iff S is conflict-free and $\forall a \notin S$, S set-attacks a.*

A Weighted AF (*WAF*) [6] is another extension of Dung's AF in which attacks between arguments are associated with a weight, indicating the relative strength of the attack.

Definition 7 *A **WAF** is a triplet $W = \langle \mathcal{A}, \mathcal{R}, w \rangle$, where $\langle \mathcal{A}, \mathcal{R} \rangle$ is the standard AF and $w \colon \mathcal{R} \mapsto \mathbb{R}^+$ is a function assigning real valued weights to attacks.*

Note that allowing 0-weight attacks is counter-intuitive since it can be interpreted as absence of attack relation. In this framework, some inconsistencies are tolerated in subsets S of arguments, provided that the sum of the weights of attacks between arguments of S does not exceed a given inconsistency budget $\beta \in \mathbb{R}^+_*$. Hence, given an inconsistency budget β, the meaning is that attacks up to a total weight of β are neglected. Dung's argument systems assume an inconsistency budget of 0, while, by relaxing this constraint, WAFs can achieve more solutions.

Definition 8 *Let $W = \langle \mathcal{A}, \mathcal{R}, w \rangle$ be a WAF. Given an inconsistency budget $\beta \in \mathbb{R}^+_*$, function sub returns the set of subsets T of $\mathcal{R}$ whose total weight does not exceed β, i.e.,*

$$sub(\mathcal{R}, w, \beta) = \{T \subseteq \mathcal{R} \mid \sum_{\langle \alpha_1, \alpha_2 \rangle \in T} w(\langle \alpha_1, \alpha_2 \rangle) \leq \beta\}.$$

Thus, intuitively, any set of arguments is consistent at some cost, and the cost required to make a set of arguments consistent immediately gives us a preference ordering over sets of arguments. Admissibility is defined in the standard way, and standard semantics are considered leading to various notions of β-extensions which echo Dung's ones (i.e., grounded, preferred, stable extensions).

Definition 9 *Given a WAF $W = \langle \mathcal{A}, \mathcal{R}, w \rangle$, let $\mathcal{E}_\mathcal{S}$ be an extension-based semantics. For $\beta \in \mathbb{R}^+$, the subset $\mathcal{E}^w_\mathcal{S}(\langle \mathcal{A}, \mathcal{R}, w \rangle, \beta)$ of $2^\mathcal{A}$ is given as*

$$\mathcal{E}^w_\mathcal{S}(\langle \mathcal{A}, \mathcal{R}, w \rangle, \beta) = \{S \subseteq \mathcal{A} \mid \exists T \in sub(\mathcal{R}, w, \beta) \wedge S \in \mathcal{E}_\mathcal{S}(\langle \mathcal{A}, \mathcal{R} \setminus T \rangle\}$$

where $\mathcal{E}_\mathcal{S}(\langle \mathcal{A}, \mathcal{R} \rangle) = \{S \subseteq \mathcal{A} \mid \mathcal{S}(S)\}$ returns the set of subsets of $\mathcal{A}$ for acceptability semantics $\mathcal{S}$.

A set $S \in \mathcal{E}^w_\mathcal{S}(\langle \mathcal{A}, \mathcal{R}, w \rangle, \beta)$ will be denoted as β-$\mathcal{S}$ set (extension), so that we refer to β-admissible sets, β-grounded extensions, β-preferred extensions, etc. So, for example, S is β-admissible if $\exists T \in sub(\mathcal{R}, w, \beta)$ such that S is admissible in the AF $\langle \mathcal{A}, \mathcal{R} \setminus T \rangle$.

Very recently, a *weighted bipolar* framework has been proposed in [2], in which is tackled the issue of arguments evaluation in a strong different perspective. In this setting, weights are associated to arguments which have a basic (*intrinsic*) strength and its evaluation method transforms it accordingly to attacks and supports received into an overall strength as acceptability degree. A big drawback in their semantics is the that it deals only with acyclic graphs.

While, the work proposed in [11] clearly distinguish between the intrinsic strength of an argument and the strength of relations. In fact, a preliminar version of the BWAF is already presented and a thorough discussion on how to handle both intrinsic strength of an argument (coming from the realiabilty

of its source) and the strength of relations with other arguments is accurately described. As well, [10] associates arguments with weights that express their source's *authority degree* and defines a strategy to combine them in order to determine which arguments withstand in a dispute concerning a given domain.

For applications involving a large number of arguments, it can be problematic to have only two levels of evaluations (arguments are either accepted or rejected). For instance, such a limitation can be questionable when using argumentation for debate platforms on the Web for a discussion. In order to fix these problems, a solution consists in using semantics that distinguish arguments not with the classical accepted/rejected evaluations, but with a larger number of levels of acceptability. Another way to select a set of acceptable arguments is to rank arguments from the most to the least acceptable ones. *Ranking-based semantics* [1] aim at determining such a ranking among arguments.

Definition 10 *A* Ranking-based semantics $\mathcal{S}$ *associates to any argumentation framework* $F = \langle \mathcal{A}, \mathcal{R} \rangle$ *a ranking* $\succeq_F^{\mathcal{S}}$ *on* $\mathcal{A}$*, where* $\succeq_F^{\mathcal{S}}$ *is a preorder (a reflexive and transitive relation) on* $\mathcal{A}$*. Given two arguments* $a, b \in \mathcal{A}$*,* $a \succeq_F^{\mathcal{S}} b$ *means that* a *is at least as acceptable as* b*.*

3 Bipolar Weighted Argumentation Framework

Among the various argumentation systems proposed in literature, the instantiation of the proper argumentation system is dependent from the experience of an expert. The need for a unique, general, extended argumentation system is required for two main reasons: (i) *theoretical*: in order to deal not only with argumentative reasoning without the human expertise for instantiation, but also with other areas of Artificial Intelligence, such as decision-making, planning, machine learning, dialogue, natural language processing, and multi-agent systems; (ii) *practical*: in order to facilitate a more direct and natural instantiation from a lot of applications, such as Online Debating Systems (ODS), Social Network analysis and moderation, virtual communities, legal cases, financial debates.

In these respects, two main generalizations of argumentation systems are detected to cope with all the above requirements: BAF and WAF. Therefore, by taking *the best of both worlds* we would be able to represent an argumentation system which combines BAFs and WAFs, extending them in a new framework while still resulting compatible with them. Then, the Bipolar Weighted Argumentation Framework (BWAF) is proposed as a further generalization of Dung-style AFs. The idea behind it is to allow not only weighted attack relations between abstract arguments, but also weighted support relations. This is achieved by assigning to each relation a weight which can be positive or negative. More formally:

Definition 11 *A* ***BWAF*** *is a triplet* $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$*, where* $\mathcal{A}$ *is a finite set of arguments,* $\hat{\mathcal{R}} \subseteq \mathcal{A} \times \mathcal{A}$ *and* $w_{\hat{\mathcal{R}}} : \hat{\mathcal{R}} \mapsto [-1, 0[\cup]0, 1]$ *is a function assigning a weight to each relation. Attack relations are defined as* $\hat{\mathcal{R}}_{att} = \{\langle a, b \rangle \in \hat{\mathcal{R}} \mid$

$w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \in [-1, 0[\ \}$ *and support relations as* $\hat{\mathcal{R}}_{sup} = \{\langle a, b\rangle \in \hat{\mathcal{R}} \mid w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \in]0, 1]\ \}$.

A BWAF can be represented as a directed graph whose nodes represent arguments, relations represent attacks and supports, and weights represent the relative strength of relations. We introduce a few new features to deal with bipolar weighted relations. First, we restrict the value of the relation's weight in a specific bounded interval, in order to define its maximum and minimum attack/support degree. Second, we assign negative real values as weights of attack relations: starting from the fact that having negative weight for a negative interaction relies on a more natural intuition, we would be able to better explore the graph with a new useful interaction paradigm dealing with the notion of defense. Last, but not least, we make the same assumptions also for support relations, assigning them positive real values as weights, following the meaning that positive weights would better represent positive interactions between arguments. As for WAFs, there can be several interpretations of these weights for support relations:

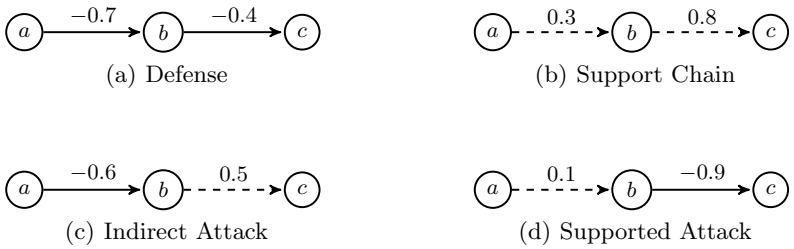
- *Votes on the supports*: represents a weight as the number of the votes in endorsement of the support, in the context of collective decision making.
- *Implicit strength of the supports*: equates weights to subjective beliefs, assigning value true to the supported argument when the supporting argument is true.
- *Explicit strength of the supports*: is the simplest interpretation; weights are used to rank the relative strength of supports between arguments, i.e., the higher the weight, the stronger the support. In this interpretation, one might consider subjective or objective criteria for ranking supports.

In order to define acceptability in BWAFs, we propose a new interaction paradigm based on weighted relations. The notion of defense, since it lies at the heart of all argumentative evaluation strategies, still remains the central concept when evaluating the justification of sets of arguments. So, we first generalize the key concept of defense between two arguments. Then, we extend this notion to sets of arguments and subsequently define acceptability semantics for BWAFs. The argumentation graph evaluation methods we present are all based on the concept of transitivity which stipulates that relations between any two nodes in the graph can be described by paths between the two nodes. In the simplest case we can ask: If there is a path of signed edges between two nodes in the graph, what relation can we induce between the two nodes? We show that the solution to this question is a *multiplication rule* in which: (i) it is exemplified the basic Dung's notion in which even-length paths of attacks means a defense (i.e., *the attack of an attack is a defense*); (ii) BAF's notions of indirect attack and supported attack are generalized into a single definition.

Definition 12 *Let* $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ *be a BWAF. Given two arguments* $a, b \in \mathcal{A}$ *and a path* $\langle a, x_1, x_2, \ldots, x_n, b\rangle$ *from a towards b, then:*

- a bw-defends b *if the result of weights multiplications* $w_{\hat{\mathcal{R}}}(\langle a, x_1\rangle) \cdot w_{\hat{\mathcal{R}}}(\langle x_1, x_2\rangle) \cdot \ldots \cdot w_{\hat{\mathcal{R}}}(\langle x_n, b\rangle)$ *is positive.*

– a bw-attacks b *if the result of weights multiplications* $w_{\hat{\mathcal{R}}}(\langle a, x_1\rangle) \cdot w_{\hat{\mathcal{R}}}(\langle x_1, x_2\rangle) \cdot \ldots \cdot w_{\hat{\mathcal{R}}}(\langle x_n, b\rangle)$ *is negative.*



(a) Defense (b) Support Chain

(c) Indirect Attack (d) Supported Attack

Fig. 1. Multiplication rules: in (a): a **bw-defends** c with a total weight of $-0.7 \cdot -0.4 = 0.28$; in (b): a **bw-defends** c with a total weight of $0.3 \cdot 0.8 = 0.24$; in (c): a **bw-attacks** c with a total weight of $-0.6 \cdot 0.5 = -0.3$; in (d): a **bw-attacks** c with a total weight of $0.1 \cdot -0.9 = -0.009$.

As we can notice, the notions of "defense" and "support" can now merge in a unique definition. Indeed, in Figure 1 we show that the notions of defense (Fig. 1 (a)) and support chain (Fig. 1 (b)) convene the notion of bw-defense since the result of weights' multiplication is positive. On the other hand, the notions of indirect and supported attack (Fig. 1 (c) and Fig. 1 (d)) convene the notion of bw-attack since the result of weights' multiplication is negative.

Taking into account sequences of weighted supports and weighted attacks leads to the following definitions applying to sets of arguments.

Definition 13 *Let* $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ *be a BWAF. A set* $S \subseteq \mathcal{A}$ set-bw-attacks *an argument* $b \in \mathcal{A}$*, iff there exists a bw-attack for* b *from an element of* S*. A set* $S \subseteq \mathcal{A}$ set-bw-defends *an argument* $a \in \mathcal{A}$*, iff for each argument* $b \in \mathcal{A}$*, if* $\{b\}$ *set-bw-attacks* a*, then there exists a bw-defense for* a *from an element of* S*.*

Notice that the notion of collective defense is still required when evaluating the acceptability of subsets of arguments as in the standard extension-based semantics. Hence, admissibility can be still addressed in the classic way. So, BAF's extension-based semantics can be built upon our bipolar weighted version of defense.

Definition 14 *Let* $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ *be a BWAF and* $S \subseteq \mathcal{A}$*. Then,* S *is a:*

– bw-conflict-free *set iff* $\nexists a, b \in S$ *s.t.* $\{a\}$ *set-bw-attacks* b*;*
– bw-admissible *set iff* S *is bw-conflict-free and* $\forall a \in S$*,* a *is set-bw-defended by* S*;*
– bw-preferred *extension iff* S *is a* $\subseteq$*-maximal bw-admissible subset of* $\mathcal{A}$*;*
– bw-stable *extension iff* S *is bw-conflict-free and* $\forall a \notin S$*,* S *set-bw-attacks* a*.*

In WAFs, inconsistency budgets are used to solve the empty extension case. In the same way, we address the case when BWAF's extensions may be empty. This means that we would restrict the bipolar weighted graph under a positive inconsistency budget in order to neglect some support relations. The main question to account for is: why should we not consider supports? Consider to extract arguments from text (from a blog post, a social network discussion, a forum thread, etc.). There may exist some forms of action to stand up for an argument, according to the purpose and audience. Nevertheless, in some cases, the supporting arguments may be advantageous if their strength is high. On the contrary, they may result uneffective if their strength is low so that one may decide to avoid these supports. Some cases are: (i) "poor" *statistical evidence* which can translated into a supporting argument with low strength and it could not be useful; (ii) "low" *expert opinion* which draws different conclusions from the same information showing that opinions may not be as reliable as facts or personal experience; (iii) "bad" *visuals* which translates important information into a visual to aid in readability but, actually, providing unhelpful visual impact.

Therefore, in order to leave out some poor or uneffective consistencies for weighted support relations, we extend the notion of inconsistency budget and define the acceptability semantics for BWAF.

Definition 15 *Let* $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ *be a BWAF,* $\hat{\mathcal{R}} = \hat{\mathcal{R}}_{att} \cup \hat{\mathcal{R}}_{sup}$ *and* $\alpha \in [-1, 0[$, $\beta \in]0, 1]$ *two inconsistency budgets. We define:*

$$sub_{att}(\hat{\mathcal{R}}_{att}, w_{\hat{\mathcal{R}}}, \alpha) = \{X \mid X \subseteq \hat{\mathcal{R}}_{att} \wedge \sum_{\langle a,b\rangle \in X} w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \geq \alpha\}$$

$$sub_{sup}(\hat{\mathcal{R}}_{sup}, w_{\hat{\mathcal{R}}}, \beta) = \{Y \mid Y \subseteq \hat{\mathcal{R}}_{sup} \wedge \sum_{\langle a,b\rangle \in Y} w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \leq \beta\}$$

$$\mathcal{E}^{\text{bw}}_{\mathcal{S}}(\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle, \alpha, \beta) =$$

$$\{S \subseteq \mathcal{A} \mid \exists R \in \{sub_{att}(\hat{\mathcal{R}}_{att}, w_{\hat{\mathcal{R}}}, \alpha) \vee sub_{sup}(\hat{\mathcal{R}}_{sup}, w_{\hat{\mathcal{R}}}, \beta)\} \wedge S \in \mathcal{E}_{\mathcal{S}}(\langle \mathcal{A}, \hat{\mathcal{R}} \setminus R\rangle)\}$$

where $\mathcal{E}_{\mathcal{S}}(\langle \mathcal{A}, \hat{\mathcal{R}} \rangle) = \{S \subseteq \mathcal{A} \mid \mathcal{S}(S)\}$ *returns the set of subsets of* $\mathcal{A}$ *for acceptability semantics* $\mathcal{S}$.

The function $sub_{att}(\cdot)$ returns the set of subsets R of $\hat{\mathcal{R}}_{att}$ whose total weight does not exceed α. While, the function $sub_{sup}(\cdot)$ returns the set of subsets R of $\hat{\mathcal{R}}_{sup}$ whose total weight does not exceed β. Therefore, $\mathcal{E}^{\text{bw}}_{\mathcal{S}}(\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle, \alpha, \beta)$ yields a subset of the power set of $\mathcal{A}$ whose elements contain only those arguments that are in relation with a weight less than α and greater than β. A bw-admissible set $S \in \mathcal{E}^{\text{bw}}_{\mathcal{S}}(\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle, \alpha, \beta)$ will be denoted as α-β-$\mathcal{S}$ extension so that we refer to α-β-admissible sets, α-β-preferred extensions, α-β-stable extensions, etc. So, for example, S is α-β-admissible if $\exists R \in sub_{att}(\hat{\mathcal{R}}_{att}, w_{\hat{\mathcal{R}}}, \alpha)$ or $R \in sub_{sup}(\hat{\mathcal{R}}_{sup}, w_{\hat{\mathcal{R}}}, \beta)$ such that S is admissible in the BAF $\langle \mathcal{A}, \hat{\mathcal{R}} \setminus R \rangle$.

3.1 BWAF Ranking Semantics based on Strength Progapation

One concern about extension-based semantics is that they do not fully exploit the weight of relations. Indeed, bw-admissibility for subsets of arguments is defined without considering the effective strength of relations and how these strengths affect the overall evaluation. The existing semantics declare all justified arguments of an argumentation framework as equally accepted. Ranking-based semantics take into account the graduality in this particular case.

In particular, one may wonder how much the weight of attack and support relations affect the overall strength of a path between two arguments. Moreover, more than one different path involving both attack and support relations may exist between two arguments. Additionally, an argument can be involved in many cycles, each of which may contain, in turn, arguments involved in other cycles, and so on. For this reason, we define an operator to assess the strength propagation of weighted relations that is able to deal with cycles.

Definition 16 *Let $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ be a BWAF and $a, b \in \mathcal{A}$ be two arguments such that there exists a simple path $\langle a, \ldots, b \rangle$. The* strength propagation *(in short, sp) from a towards b is defined as:*

$$sp(a,b) = \sum_{\langle a,\ldots,b\rangle} \Big(path\_weight(\langle a, \ldots, b\rangle) \cdot \prod_{c \in \langle a,\ldots,b\rangle} influence(c) \Big)$$

where

$$path\_weight(\langle v_1, \ldots, v_n \rangle) = \prod_{i=2}^{n} w_{\hat{\mathcal{R}}}(\langle v_{i-1}, v_i \rangle)$$

$$influence(v) = \begin{cases} \prod_{\langle v_1,\ldots,v_k\rangle} path\_weight(\langle v_1, \ldots, v_k \rangle) & \text{if } v_1 = v_k = v \\ 1 & \text{otherwise} \end{cases}$$

Function $path\_weight(\cdot)$ computes the strength of a simple path by multiplying every weight relation in it, while function $influence(\cdot)$ computes the influence of an node within the simple path on the basis of cycles to which it belongs. If it belongs to a cycle, it is computed as product of the path cycle starting and ending to it, as 1 otherwise. In the case of sub-cycles eventually involved in a path cycle, we only consider its "*maximal*" circuit. In this way, cycles and possibly involved sub-cycles are traversed exactly once. Since the evaluation of a cycle in an argument graph is always problematic, its presence within a path may influence significantly the evaluation of its overall strength. So, the influence of a node in a path, which also belongs to a cycle, may reduce drastically the path strength propagation. Hence, function $sp(a, b)$ detects all the simple paths starting from a towards b along possibly cycles in them and returns a positive value if there exists a (collective) bw-defense. Otherwise, it returns a negative value if there exists a (collective) bw-attack between them. Such a function gives a measure of *overall strength* of the relations between two arguments as part of a whole discussion.

We show its behavior in the following example.

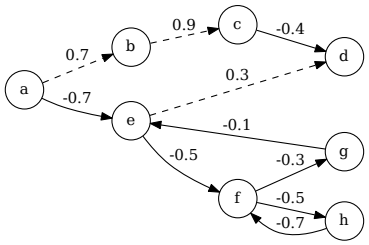


Fig. 2. Representation of BWAF for Example 1.

Example 1. Consider the BWAF depicted in Figure 2. Suppose that we want to compute the strength propagation for the path from a to d. First, we have to compute the existing path weigths from a to d. There exist two path weights:

$$path\_weight(\langle a, b, c, d\rangle) = 0.7 \cdot 0.9 \cdot (-0.4) = -0.252$$
$$path\_weight(\langle a, e, d\rangle) = (-0.7) \cdot 0.3 = -0.21$$

Since e belongs to a cycle, we need to compute its influence:

$$\begin{aligned} influence(e) &= path\_weight(\langle e, f, h, f, g, e\rangle) = \\ &= (-0.5) \cdot (-0.5) \cdot (-0.7) \cdot (-0.3) \cdot (-0.1) = -0.00525 \end{aligned}$$

Hence, we can compute the strength propagation from a to d:

$$\begin{aligned} sp(a, d) &= path\_weight(\langle a, b, c, d\rangle) \cdot 1 + \\ &+ path\_weight(\langle a, e, d\rangle) \cdot influence(e) = -0.2508975. \end{aligned}$$

With a big number of arguments, and a lot of users participating, it can be problematic to have a boolean evaluation (accepted/rejected). Then, a more informative evaluation is required. In the following, we propose a new semantics, i.e. ***sp*-ranking semantics**, which ranks arguments by comparing them with a numerical acceptability degree assigned to each argument. Such a ranking relies on bw-attacks and bw-defenses of each argument in order to evaluate its acceptability rank. We recall that an attack amounts to undermining one of the components of an argument, and has thus a negative impact on its target. Vice versa, a support amounts to stand up for an argument, and has thus a positive impact on its target. An evaluation of the overall acceptability of an argument becomes mandatory, namely for judging how much its conclusion is reliable. Therefore, *sp*-ranking semantics exploits the strength propagation of couples of arguments linked by path branches. A *path branch* from a to b is a sequence of nodes $\langle a_0, \ldots, a_k\rangle$, with $a_0 = a, a_k = b$ and $\forall i < k, \langle a_i, a_{i+1}\rangle \in \hat{\mathcal{R}}$ and a is not attacked nor supported. So, we first define a ranking opeator to assign an acceptability degree to each argument in the BWAF which deals with the

strength propagation of path branches ending to them. Then, we rank arguments according to their acceptability degree.

Definition 17 *Let $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ be a BWAF, $a \in \mathcal{A}$ an argument, $sp(\cdot, a)$ the strength propagation of path branch ending to a, $SP = \{sp(x_1, a), \ldots, sp(x_m, a)\}$ the set of all the strength propagations ending to a and $P = \{p_1, \ldots, p_n\}$ the set of all directed paths towards a in G, with $p_i = \langle x, \ldots, a \rangle \in P$, $\forall i \leq n$. The sp-*rank function $\mathrm{spr}\colon \mathcal{A} \mapsto [0, 2]$ *is defined as:*

$$\mathrm{spr}(a) = \begin{cases} 1 & \text{if } \forall x \in \mathcal{A}\colon \langle x, a \rangle \notin \hat{\mathcal{R}} \\ \frac{1}{n} \sum_{sp(x_i,a) \in SP}^{m} 1 + sp(x_i, a) & \text{otherwise} \end{cases}$$

Our approach is based two principles: the impact of unrelated argument, which play a key role in the (extension-based) acceptability of an argument, and the strength propagation of its attackers and of its supporters. It is worth to clarify why we differently consider the set P of directed paths and the set SP of strength propagation paths. The strength propagation function $sp(\cdot)$ computes the strength of all the existing multipaths between two nodes in the argumentation graph at once. If there exists at most one path for each couple of nodes, then $n = m$, otherwise we have that $m < n$. In this respect, we ensure that the ranking of arguments is always in the interval $[0, 2]$. Generally, arguments that not receive any attack or support play a key role in the (classical) acceptability. So, we set for them a ranking of 1. In this way, bw-defended arguments will achieve a ranking > 1, otherwise bw-attacked ones will achieve a ranking < 1. In this sense, the ranking of 1 will *tip the scale*, meaning that we would be able to consider not only an acceptability ranking for arguments, but also a mapping to the classical accepted/rejected evaluation. Finally, we can rank arguments according to their $spr(\cdot)$ function.

Definition 18 *The sp-*ranking semantics *spr associates to any BWAF $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ a ranking $\succeq_G^{\mathrm{spr}}$ on $\mathcal{A}$ such that $\forall a, b \in \mathcal{A}, a \succeq_G^{\mathrm{spr}} b$ iff $\mathrm{spr}(a) \geq \mathrm{spr}(b)$.*

4 Construction of a BWAF from an ODS

The task of building an argument system from a debate platform has been already addressed in [4, 7, 8]. Here, we consider classical thread discussions, that involve a tree (i.e., hierarchical) structure. A typical online debate thread would consist of a discussion topic, i.e. a content shared by a user, followed by comments from other users. In response to a comment there may be an answer, so comments are organized in a tree discussion where the root node, i.e. the major claim, is the shared content and the other nodes represent comments. Each of these comments has as children the comments in response to it. Hence, considering the similarity between the comments, the sentiment associated with them and their hierarchical structure, it is possible to extract a BWAF that models the discussion by identifying weighted attacks and supports depending on their strength.

The set of arguments $\mathcal{A}$ is made up of tree nodes representing the targeted discussion. The set of relations $\hat{\mathcal{R}}$ is built starting from each comment, which can be in favor or against the argument to which one is replying. In particular, the comments in the first level of the tree, related to the major claim, can be analyzed to extract the sentiment polarity with respect to that argument. If the polarity is positive, such comments are arguments supporting the major claim, otherwise, if the polarity is negative, they are attacking arguments. Specifically, we assume that the arguments in the first level of the tree address the major claim without going "off-topic". Regarding the answer comments to a comment, there can be an attack or a support with respect to it, depending on the extent they address the same argument. In fact, if a comment addresses the same topic of the comment it answers then there is a support, otherwise there is an attack. Formally, the task of weighing the relation between two arguments $a, b \in \mathcal{A}$ can be quantified as follows:

$$w_{\hat{\mathcal{R}}}(\langle a, b\rangle) = \textit{similarity}(\langle a, b\rangle) \cdot \textit{sentiment}(b) \tag{1}$$

where *similarity*: $\mathcal{A} \times \mathcal{A} \mapsto [0, 1]$ is a function evaluating the similarity between two arguments, and *sentiment*: $\mathcal{A} \mapsto [-1, 1]$ is a function that evaluates the polarity of sentiment associated with an argument. Given $a, b \in \mathcal{A}$, if $w_{\hat{\mathcal{R}}}(\langle a, b\rangle) < 0$ then there exists an attack from a towards b, otherwise if $w_{\hat{\mathcal{R}}}(\langle a, b\rangle) > 0$ then there exists a support from a towards b.

In order to determine if two arguments are treating the same theme, we adopt a measure of similarity based on *word embeddings*. These techniques are particularly effective in a large variety of applications related to the similarity of the contents. Generally speaking, in the context of semantic spaces models, a word embedding is a dense vector representation of a word. These vectors, learned from a corpus of text, capture the similarity between words by using them in the context in which they appear more frequently. In this way, it is satisfied the distributional hypothesis for two words in which the more semantically similar, the higher they tend to recur in similar contexts [9]. For this task, we exploited the GloVe model [12], which is a well-known model for learning word embeddings. In particular, the similarity between two sentences is computed as the cosine similarity between the average of the word embedding associated with the token present in sentences.

Concerning the task to determine the sentiment polarity of a sentence, we exploited the model presented in [13], which is a well-known model in the field of sentiment analysis. This model learns complex relationships between the terms of a sentence to determine the sentiment polarity. In particular, polarity classes provided by the model are "very positive", "positive", "neutral", "negative" and "very negative", for which it has been established the correspondence, respectively, with the numerical values -1, -0.5, 0, 0.5 and 1.

The procedure of (weighted) argument graph construction just described may embed some noise. Nevertheless, the procedure is simple and computationally fast, so that the graph instantiation is still quite reliable.

4.1 Application to a Reddit Thread

In order to identify arguments and their interaction in real cases, we considered the content on the platform *Reddit*[1], a big community that allows users to share links, opinions, real-time content and information and to discuss and comment on what is published on the platform. In this case, we consider a Reddit discussion of an episode of popular TV series[2] divided into several arguments with attacks and supports.

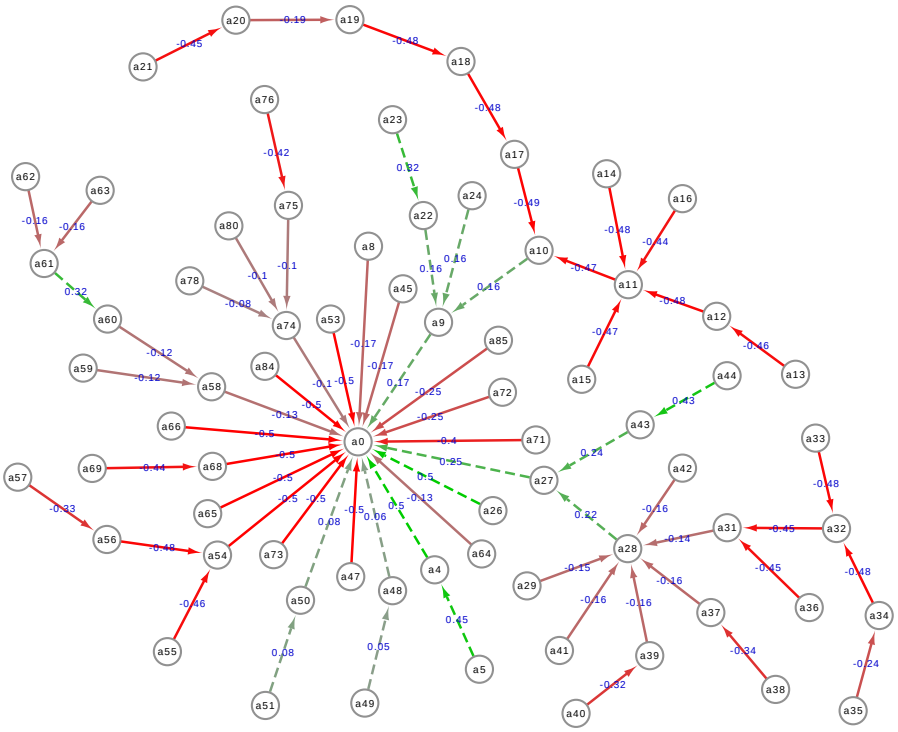


Fig. 3. BWAF representation for the considered Reddit Thread.

The produced BWAF is made up of 70 arguments and 69 relations, of which 52 attacks and 17 supports. The corresponding graph is reported in Figure 3. By analyzing the thread, it is possible make some considerations useful to judge the quality of the generated BWAF and to assess it with both extension- and ranking-based semantics. Node $a0$ in Figure 3 acts as the major claim for the whole discussion. While, other nodes in the BWAF represent replies to the major claim and subsequently comments on them. We note that the Equation (1) is

[1] www.reddit.com
[2] http://bit.ly/2fQxq4I

able to capture the meaning of nodes relating to other ones and that the weight of relation helps us in identifying some useful semantics. The *bw-stable semantics* has a unique extension, which is the following:
{$a4$, $a5$, $a8$, $a13$, $a14$, $a15$, $a16$, $a17$, $a19$, $a21$, $a22$, $a23$, $a24$, $a26$, $a29$, $a33$, $a35$, $a36$, $a38$, $a40$, $a41$, $a42$, $a43$, $a44$, $a45$, $a47$, $a48$, $a49$, $a50$, $a51$, $a53$, $a55$, $a57$, $a59$, $a62$, $a63$, $a64$, $a65$, $a66$, $a69$, $a71$, $a72$, $a73$, $a76$, $a78$, $a80$, $a84$, $a85$}.

Table 1. *sp*-ranking semantics

arg	spr	arg	spr	arg	spr	arg	spr	arg	spr	arg	spr
$a4$	1.45	$a80$	1.0	$a59$	1.0	$a38$	1.0	$a13$	1.0	$a32$	0.8176
$a43$	1.43	$a78$	1.0	$a57$	1.0	$a36$	1.0	$a8$	1.0	$a34$	0.76
$a22$	1.32	$a76$	1.0	$a55$	1.0	$a35$	1.0	$a5$	1.0	$a11$	0.7077
a10	1.108	$a73$	1.0	$a53$	1.0	$a33$	1.0	$a58$	0.9641	$a39$	0.68
$a19$	1.0855	$a72$	1.0	$a51$	1.0	$a29$	1.0	$a28$	0.9595	$a56$	0.67
$a50$	1.08	$a71$	1.0	$a49$	1.0	$a26$	1.0	$a18$	0.9589	$a37$	0.66
$a48$	1.05	$a69$	1.0	$a47$	1.0	$a24$	1.0	$a74$	0.954	$a75$	0.58
a9	1.0425	$a66$	1.0	$a45$	1.0	$a23$	1.0	$a60$	0.9488	$a68$	0.56
$a17$	1.0197	$a65$	1.0	$a44$	1.0	$a21$	1.0	***a0***	0.9225	$a20$	0.55
a27	1.0035	$a64$	1.0	$a42$	1.0	$a16$	1.0	$a31$	0.9047	$a12$	0.54
$a85$	1.0	$a63$	1.0	$a41$	1.0	$a15$	1.0	$a54$	0.8492		
$a84$	1.0	$a62$	1.0	$a40$	1.0	$a14$	1.0	$a61$	0.84		

While, in Table 1 is reported the *sp*-ranking semantics. It becomes apparent to note that all the arguments belonging to the bw-stable extension hold an acceptability degree ≥ 1. A slight difference in the two semantics is that *sp*-ranking one yields as defended also arguments $a9$, $a10$, and $a27$. This is because the computation of the (collective) strength propagation of all path branches ending to them takes advantages of weight of relations and, in particular, of weighted support relations. In fact, one limit of extension-based semantics is that they do not fully profit by support relations.

5 Conclusions and Future Work

This work proposed a generalization of Dung's AF, the BWAF, which is able to express weighted attack and support relations. We presented a new characterization of the notion of defense, along with an extended version of inconsistency budget, two classes of extension-based semantics and a new ranking-based semantics. Since weighted relations can affect indirect attacks/supports, we handle this with the definition of strength propagation, in order to quantify the positive or negative strength of indirect relations between two arguments. Another contribution regards the strategy to construct BWAFs starting from real data. The detection of weighted relations is based on the similarity measure of textual contents and on their associated sentiment polarity. To prove its effectiveness, we built a BWAF from a Reddit discussion and we examined the resulting model by

comparing extension- and ranking-based semantics. It is shown that our framework is able to clearly represent the whole discussion and the *sp*-ranking semantics better exploits the effectiveness of (weighted) support relations.

Our work opens some issues for further research. In the phase of evaluation of accepted arguments, one may find that not all the arguments of a discussion are essential when drawing conclusions. Especially when the cardinality of the set of arguments is high, identifying the most relevant arguments is a tricky task. In huge argumentation graphs, the analysis of its synthesis may favor better interpretability and may allow you to extract semantics that include the strongest arguments.

Acknowledgments

This work was partially funded by the Italian PON 2007-2013 project PON02_00563_3489339 'Puglia@Service'.

References

[1] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In *SUM 2013*, pages 134–147. Springer, 2013.

[2] L. Amgoud and J. Ben-Naim. Evaluation of arguments in weighted bipolar graphs. In *ECSQARU 2017*, pages 25–35, 2017.

[3] C. Cayrol and M. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *ECSQARU*, volume 3571 of *Lecture Notes in Computer Science*, pages 378–389. Springer, 2005.

[4] F. Cerutti et al. A pilot study in using argumentation frameworks for online debates. In *SAFA*, pages 63–74, 2016.

[5] 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):321–357, 1995.

[6] P. E. Dunne et al. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457 – 486, 2011.

[7] V. Evripidou and F. Toni. Quaestio-it.com: a social intelligent debating platform. *Journal of Decision Systems*, 23(3):333–349, 2014.

[8] J. Leite and J. Martins. Social abstract argumentation. In *IJCAI*, volume 11, pages 2287–2292, 2011.

[9] G. A. Miller and W. G. Charles. Contextual correlates of semantic similarity. *Language and cognitive processes*, 6(1):1–28, 1991.

[10] A. Pazienza, F. Esposito, and S. Ferilli. An authority degree-based evaluation strategy for abstract argumentation frameworks. In *Proceedings of CILC*, pages 181–196, 2015.

[11] A. Pazienza, S. Ferilli, and F. Esposito. On the gradual acceptability of arguments in bipolar weighted argumentation frameworks with degrees of trust. In *ISMIS 2017*, pages 195–204, 2017.

[12] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–43, 2014.

[13] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, volume 1631, page 1642, 2013.