

Towards a General Model for Abstract Argumentation Frameworks^{*}

Stefano Ferilli

Dipartimento di Informatica – Università di Bari Aldo Moro, Bari, Italy
stefano.ferilli@uniba.it

Abstract. In its original definition, the Abstract Argumentation framework considers atomic claims and a binary attack relationship among them, based on which different semantics would select subsets of claims consistently supporting the same position in a dispute or debate. While attack is obviously the core relationship in this setting, in more complex (and in many real-world) situations additional information may help, or might even be crucial, in determining such positions, and especially those that are going to win the debate. Examples are bipolarity (considering also the support relationship between pairs of claims) and weights (assigning different importance to different elements of the framework). These additional features have often been considered separately, yielding incompatible or anyhow disjoint models for argumentation frameworks. In this paper we propose a model that unifies all these perspectives, and further extends them by allowing to express contextual information associated to the arguments, in addition to their relationships.

Keywords: Abstract Argumentation · Argumentation Frameworks

1 Introduction

Argumentation is the inferential strategy for practical and uncertain reasoning aimed at coping with partial and inconsistent knowledge, in order to justify one of several contrasting positions in a discussion [14]. A typical case is a debate in which each participant tries to support one position with suitable claims (the *arguments*), also attacking the arguments put forward by others to support competing positions, and defending his position from the attacks of the others. Since different forms of disputes (or anyway situations with contrasting evidence) are ubiquitous in real life, the availability of automated techniques for carrying out argumentation would be extremely useful. Hence, the birth of a specific branch of Artificial Intelligence aimed at developing models, approaches, techniques and systems for dealing with different aspects of argumentative reasoning.

Abstract argumentation, in particular, focuses on the resolution of the dispute based only on ‘external’ information about the arguments (notably, the

^{*} Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

inter-relationships among them), neglecting their internal structure or interpretation. Traditional Abstract Argumentation Frameworks (AFs for short) can express only attacks among arguments. While already useful to tackle many cases (because the attack relationship is indeed the very core and driving feature in a debate), this is obviously a significant limitation in expressiveness. So, several lines of research tried to overcome such a limitation by introducing additional features to be considered in the argumentation frameworks. Most famous are the possibility of expressing supports between arguments (in addition to attacks), or the ‘strength’ of attacks (in the form of a number). These extensions were mainly developed independently of each other, so that they cannot be straightforwardly combined into a more powerful framework encompassing all of them.

This paper proposes a general framework that brings to a cooperation of the different features of the single frameworks, yielding a much more powerful model to carry out abstract argumentation. It can simulate any of those frameworks, and also provides for the additional possibility of assigning a degree of ‘strength’ also to the arguments, not just to their relationships. We call it *Generalized Argumentation Framework*, or GAF. With respect to our previous preliminary work in this direction, here we fix some notational issues, reorganize the model formalization and abstract away from details of specific computational approaches that can be applied on it. Indeed, we stress the fact that our aim is not proposing any evaluation strategy or computational procedure, but a model that can be specialized and tailored to different contexts and domains, and on which theoretical investigation can be carried out for defining semantics and evaluation strategies. We believe our proposal can be taken as a reference, both for porting solutions developed for previous partial extensions, and for developing new solutions that fully exploit its extended expressive power. Also, we show that our model can be easily expressed using matrix representations, which might bring significant improvements in efficiency in computing the argumentation outcomes thanks to the use of matrix operations.

The paper is organized as follows. After recalling basic concepts of abstract AFs and discussing related works in the next section, in Section 3 we will define the new generalized model and show how it maps onto existing AFs. Then, in Section 4 we propose a specialization of it that allows to consider user confidence in the arguments and trust in the other users. Section 5 discusses the advantages of expressing our model using matrix-based representations, while Section 6 concludes the paper.

2 Basics & Related Work

The original (and now classical) Abstract Argumentation setting was proposed by Dung [7]. It can express only the attack relationship between pairs of arguments, as the core feature indicating inconsistency in the available information:

Definition 1 *An argumentation framework (AF for short) is a pair $F = \langle \mathcal{A}, \mathcal{R} \rangle$, where \mathcal{A} is a finite set of arguments and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relationship (meaning that, given $\alpha, \beta \in \mathcal{A}$, if $\alpha \mathcal{R} \beta$ then α attacks β).*

In this setting, no direct agreement between arguments can be expressed. Agreement can only indirectly be derived based on the attack relationship, yielding the notion of *defense*:

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

- $\alpha \in \mathcal{A}$ is defended by S if $\forall \beta \in \mathcal{A}: \beta \mathcal{R} \alpha \Rightarrow \exists \gamma \in S$ s.t. $\gamma \mathcal{R} \beta$;
- $f_F: 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}}$ s.t. $f_F(S) = \{\alpha \mid \alpha \text{ is defended by } S\}$ is the characteristic function of F .

So, an argument may ‘defend’ other arguments by attacking their attacker (or, in other words, attacking an attacker amounts to a defense).

An argumentation *semantics* is the formal definition of a method ruling the argument evaluation process. In particular, extension-based semantics determine which subset(s) of arguments in an AF, called *extensions*, can stand together and possibly be considered as the ‘winners’ of the dispute expressed by the AF. On the other hand, ranking-based semantics [1] individually evaluate single arguments rather than sets of arguments, and, given an AF, determine a ranking of the available arguments in the form of a pre-order (reflexive and transitive relation). We will not delve further into semantics in the following, since the aim of this paper is providing a unified framework in which the existing semantics can be transposed, and new ones can be developed, leveraging its additional features.

Several works tried to overcome the limitations of the classical AFs by generalizing them in different ways. The most investigated limitations were the possibility of expressing only attacks between pairs of arguments, and the inability of distinguishing different degrees of ‘strength’ for the single attacks. Research on the former led to the so-called *Bipolar* AFs (or *BAFs*) [6], allowing two kinds of interactions between arguments, expressed respectively by the *attack* relation and the *support* relation. Research on the latter led to the so-called *Weighted* AFs (or *WAFs*) [8], allowing to specify a numeric weight for each attack between arguments, indicating its relative strength. BAFs and WAFs cannot be immediately combined, because the computational procedures for WAFs are specified only for attacks, and are not simply applicable to supports if no strategy for combining overall attack and support assessment is provided.

This was the reason behind some attempts to define extensions encompassing both possibilities. Specifically, [11] proposed a formal extension of the framework (named *Bipolar Weighted Argumentation Framework*, or *BWAF*) and a gradual evaluation strategy, while [4] extended their previous work on graph-based computational strategies for unipolar AFs. BWAFs embed the notions of attack and support into the weights, by considering negative weights for attacks and positive weights for supports.

Definition 3 A BWAF is a triplet $F = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$, where \mathcal{A} is a finite set of arguments, $\hat{\mathcal{R}} \subseteq \mathcal{A} \times \mathcal{A}$ and $w_{\hat{\mathcal{R}}}: \hat{\mathcal{R}} \mapsto [-1, 0[\cup]0, 1]$ assigns a weight to each relation instance. Within $\hat{\mathcal{R}}$, the attack sub-relation is defined as $\hat{\mathcal{R}}_{att} = \{(\alpha, \beta) \in \hat{\mathcal{R}} \mid w_{\hat{\mathcal{R}}}((\alpha, \beta)) \in [-1, 0[\}$, while the support sub-relation is defined as $\hat{\mathcal{R}}_{sup} = \{(\alpha, \beta) \in \hat{\mathcal{R}} \mid w_{\hat{\mathcal{R}}}((\alpha, \beta)) \in]0, 1] \}$.

Weight 0 is not considered, since it would mean the absence of an attack or support relation. Note that this weighting scheme neatly distinguishes attacks from supports: a support is not considered as just the complement of an attack, but they are two distinct concepts, and only after determining the concept to be used (as the sign of the weight) the weight makes sense. This allows BWAFs to be consistent with previous bipolar approaches to Abstract Argumentation.

Some researchers pointed out that not only the relationships among arguments, but also the arguments themselves may have different degrees of ‘strength’ or ‘reliability’. E.g., according to [2], the *intrinsic strength* of an argument may come from different sources: the certainty degree of its reason [3], the importance of the value it promotes if any [5], the reliability of its source [10]. In this line of thought, albeit there is no agreement in the literature about the possibility of using contextual information in an AF, [12] further extended the BWAF framework into the *Trust-affected Bipolar Weighted Argumentation Framework* (or T-BWAF), introducing the possibility of weighting also the arguments by determining their *intrinsic strength* as the result of several factors, internal to the argument (the authority of the source of the argument and its own confidence in the validity of the argument) or external to it (the trust of a community in the source of the argument¹).

Definition 4 A T-BWAF is a tuple $F = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\mathcal{A}}, w_{\hat{\mathcal{R}}}, \mathcal{K}, \text{conf} \rangle$, where \mathcal{A} , $\hat{\mathcal{R}}$ and $w_{\hat{\mathcal{R}}}$ are as in BWAFs, $w_{\mathcal{A}}: \mathcal{A} \mapsto [0, 1]$ assigns a weight to each argument, $\mathcal{K} = \{T_i\}_{i \in \mathcal{T}}$ is a set of Trust Users Graphs² for a set of topics \mathcal{T} and users \mathcal{U} , and $\text{conf}: \mathcal{U} \times \mathcal{A} \mapsto [0, 1]$ is the User Argument Confidence function.

where the additional components with respect to BWAFs are $w_{\mathcal{A}}$, \mathcal{K} , and conf . While the perspectives according to which T-BWAFs assess the arguments’ strength are sensible (authority, confidence and trust), unfortunately the resulting framework is totally integrated with the specific evaluation solutions proposed in the paper. This motivates the work in this paper, that generalizes [12] so as to abstract away from the specific computational approaches.

[13] proposed a matrix representation for BWAFs, showing how to use it for computing some traditional semantics and defining a new semantics specifically associated to such a representation. We propose the use of matrix representations also for our extended framework.

¹ We are aware of other works about trust in argumentation, but since they are later than [12] and do not refer it, we assume there is no sufficient relationship to this work, which builds on [12].

² A *Trust Users Graph* is a directed weighted graph $T = \langle \mathcal{U}, \mathcal{E}, w_{\mathcal{U}}, w_{\mathcal{E}} \rangle$ where:

- \mathcal{U} is a set of users,
- $\mathcal{E} = \mathcal{U} \times \mathcal{U}$ (a complete graph),
- $w_{\mathcal{U}}: \mathcal{U} \mapsto [0, 1]$ assigns a weight to each user, expressing his *subjective confidence* about the topic;
- $w_{\mathcal{E}}: \mathcal{E} \mapsto [0, 1]$, given a pair of users $(u_1, u_2) \in \mathcal{E}$, expresses the *trust* that u_1 has for u_2 (0 meaning full distrust, 0.5 full ignorance, and 1 full trust).

3 The Generalized Argumentation Framework

In this section we formalize our generalized model that extends traditional AFs with bipolarity, weights on both attacks and supports, and weights on the arguments. It comes with no embedded solutions for the use of such components. Rather, it provides a flexible way for representing different possible interpretations and perspectives on them, and a basis to implement different evaluation procedures, including those proposed by previous works. As said, we will present the new model by referring and comparing it to [12], which is the most comprehensive model proposed so far. In recalling the elements of [12], we also reorganize their definitions and presentation in order to make it more comfortable and fix some formal and notational issues of the original work.

Definition 5 A Generalized Argumentation Framework (**GAF**) is a tuple $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$, where:

- \mathcal{A} is a finite set of arguments,
- $\mathcal{S}(\mathcal{A})$ is a system providing external information on the arguments³ in \mathcal{A} ,
- $w_{\mathcal{A}}: \mathcal{A} \times \mathcal{S}(\mathcal{A}) \mapsto [0, 1]$ assigns a weight to each argument, to be considered as its intrinsic strength, also based on $\mathcal{S}(\mathcal{A})$, and
- $w_{\mathcal{R}}: \mathcal{A} \times \mathcal{A} \mapsto [-1, 1]$ assigns a weight to each pair of arguments.

It is up to the knowledge engineer defining, case by case, what $\mathcal{S}(\mathcal{A})$ is⁴, and how it affects the assessment of the ‘intrinsic’ reliability of arguments. For those who are not comfortable with the use of contextual information in an AF, $\mathcal{S}(\mathcal{A})$ can simply be empty. They might still accept the use of $w_{\mathcal{A}}$ for expressing some kind of ‘intrinsic’ strength of the arguments, or ignore $w_{\mathcal{A}}$ as well.

Note that, differently from all previous models, the relationship between arguments is implicit in the GAF model. This is because we consider a complete graph, where any pair of arguments has a weighted relationship. For practical purposes, weight 0 can be interpreted as the absence of any (attack or support) relationship, and ignored when drawing the argumentation graph. The bipolar relationship considered in BWAFs can be easily extracted as

$$\hat{\mathcal{R}} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\mathcal{R}}(\alpha, \beta) \neq 0\}$$

Not only using negative weights for attacks and positive weights for supports is quite intuitive (attacking an argument subtracts to its credibility, supporting it adds to its credibility) and comfortable (the kind of relationship can be immediately distinguished by its sign). Using negative weights for attacks also allows to straightforwardly translate the traditional assumptions for the bipolar case:

1. attacking the attacker of an argument amounts to defending (i.e., somehow supporting) that argument (known as *reinstatement*);

³ This allows us to embed existing proposals in the literature in GAFs. In principle, a system that plugs external information also in the definition of attack and support strength might be added, as well.

⁴ E.g., $\mathcal{S}(\mathcal{A}) = (\mathcal{K}, \text{conf})$ in T-BWAFs.

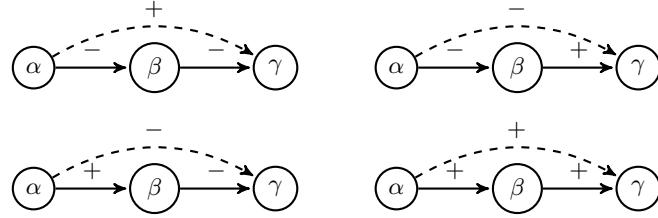


Fig. 1. Sign rule for attacks and supports

2. attacking the supporter of an argument amounts to attacking that argument;
3. supporting the attacker of an argument amounts to attacking that argument;
4. supporting the supporter of an argument amounts to supporting that argument;

into mathematical computations, since they clearly correspond to the sign rule used in mathematics:

followed by	Support	Attack	\times	+	-
Support	Support	Attack	+	+	-
Attack	Attack	Support	-	-	+

(see Figure 1 for a graphical representation).

This rule also allows to immediately turn the notions of indirect attacks and defenses into mathematical operations. Indeed, just like a path of relationships including an even number of attacks amounts to a defense, so the product of an even number of minus signs gets a plus sign; *vice versa*, just like a path of relationships including an odd number of attacks still amounts to an attack, so the product of an odd number of minus signs gets a minus sign. So, we can report to GAFs the notions of bw-attacks and bw-defense defined for BWAFs⁵:

Definition 6 Given a GAF $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$ and a sequence of arguments $\langle x_0, x_1, \dots, x_n \rangle$ such that $\forall i = 0, \dots, n : x_i \in \mathcal{A}$, we say that:

- x_0 g-defends x_n iff $\prod_{i=1}^n w_{\mathcal{R}}(x_{i-1}, x_i) > 0$
- x_0 g-attacks x_n iff $\prod_{i=1}^n w_{\mathcal{R}}(x_{i-1}, x_i) < 0$

Note that, while in BWAFs these notions are defined only for sequences of arguments which made up a path in the argumentation graph, in GAFs we may consider any sequence of arguments, since when it is not associated to a path, the missing links would have weight 0 and thus would bring the product at 0.

⁵ Given a BWAF $G = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$, two arguments $x_0, x_n \in \mathcal{A}$ and a path $\langle x_0, x_1, \dots, x_n \rangle$ from x_0 to x_n :

- x_0 bw-defends x_n iff $\prod_{i=1}^n w_{\hat{\mathcal{R}}}(x_{i-1}, x_i) > 0$
- x_0 bw-attacks x_n iff $\prod_{i=1}^n w_{\hat{\mathcal{R}}}(x_{i-1}, x_i) < 0$

So, the GAF formalization also provides a computational means to determine whether or not two arguments affect each other along a given path.

Also, GAFs allow to easily compute statistics on the direct attacks and supports for an argument:

Definition 7 *Given a GAF $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$ and an argument $x_0 \in \mathcal{A}$, we can compute:*

- the number of attacks received by x_0 as $\sum_{x \in \mathcal{A}, w_{\mathcal{R}}(x, x_0) < 0} 1$
- the number of supports received by x_0 as $\sum_{x \in \mathcal{A}, w_{\mathcal{R}}(x, x_0) > 0} 1$
- the direct justification balance of x_0 as $\sum_{x \in \mathcal{A}} 1 \cdot \text{sign}(w_{\mathcal{R}}(x, x_0))$
- the cumulative weighted attack received by x_0 as $\sum_{x \in \mathcal{A}, w_{\mathcal{R}}(x, x_0) < 0} -w_{\mathcal{R}}(x, x_0)$
- the cumulative weighted support received by x_0 as $\sum_{x \in \mathcal{A}, w_{\mathcal{R}}(x, x_0) > 0} w_{\mathcal{R}}(x, x_0)$
- the weighted direct justification balance of x_0 as $\sum_{x \in \mathcal{A}} w_{\mathcal{R}}(x, x_0)$

Compared to traditional weighted frameworks (WAFs), where the weight of an attack could be any number, bounding the absolute weights within fixed minimum and maximum values intuitively allows one to identify a level of strength at which the attacking argument ‘fully’ defeats the attacked one (or the supporting argument ‘fully’ supports the supported one). The specific $[0, 1]$ range also helps intuition due to its wide use in probability theory.

3.1 Mapping From and To Classical Frameworks

Since one stated objective of our proposal is that it should be able to encompass, combine and extend less expressive models, a basic requirement is that GAFs can at least simulate the established models in the literature, namely BWAFs, WAFs, BAFs, and AFs. The following proposition confirms that our generality hypothesis holds.

Proposition 1 *Given an argumentation framework in any of the less expressive models (BWAF, WAF, BAF, AF), a corresponding GAF $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$ can be defined, including only the portion of information that they are able to express.*

Intuitively, the GAF can be defined by setting:

- $\mathcal{S}(\mathcal{A}) = \{\perp\}$, i.e., a single uninformative item;
- $w_{\mathcal{A}} = 1$, i.e., the constant function returning 1 for any argument, meaning full reliability

and $w_{\mathcal{R}}$ as follows for the different models:

BWAF $\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$:

$$w_{\mathcal{R}}(\alpha, \beta) = \begin{cases} w_{\hat{\mathcal{R}}}(\alpha, \beta) & \text{if } (\alpha, \beta) \in \hat{\mathcal{R}} \\ 0 & \text{otherwise} \end{cases}$$

WAF $\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$:

$$w_{\mathcal{R}}(\alpha, \beta) = \begin{cases} -\frac{w_{\hat{\mathcal{R}}}(\alpha, \beta)}{\max_{\alpha, \beta \in \mathcal{A}} w_{\hat{\mathcal{R}}}(\alpha, \beta)} & \text{if } (\alpha, \beta) \in \hat{\mathcal{R}} \\ 0 & \text{otherwise} \end{cases}$$

(by normalizing the attack weights into $[0, 1]$ —of course, also the justification thresholds used in this model must be normalized into the same range)

BAF $\langle \mathcal{A}, \hat{\mathcal{R}}_{att}, \hat{\mathcal{R}}_{sup} \rangle$:

$$w_{\mathcal{R}}(\alpha, \beta) = \begin{cases} -1 & \text{if } (\alpha, \beta) \in \hat{\mathcal{R}}_{att} \\ 1 & \text{if } (\alpha, \beta) \in \hat{\mathcal{R}}_{sup} \\ 0 & \text{otherwise} \end{cases}$$

AF $\langle \mathcal{A}, \hat{\mathcal{R}} \rangle$:

$$w_{\mathcal{R}}(\alpha, \beta) = \begin{cases} 1 & \text{if } (\alpha, \beta) \in \hat{\mathcal{R}} \\ 0 & \text{otherwise} \end{cases}$$

Conversely, when the additional information provided by GAFs is not needed for the current purposes, one might be interested in working in one of the simpler models (e.g., for using existing argument evaluation strategies and tools). The following proposition shows how a GAF can be reduced to each those models, by stripping the information they cannot convey and keeping only the portion that they can express.

Proposition 2 *Given a GAF $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$, corresponding frameworks can be defined for each of the less expressive models (BWAF, WAF, BAF, AF) by extracting from F only the portion of information that they are able to express.*

Indeed, the less expressive frameworks are extracted from GAFs as follows:

BWAF $\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ with $\hat{\mathcal{R}} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\mathcal{R}}(\alpha, \beta) \neq 0\} \subseteq \mathcal{A} \times \mathcal{A}$ and $w_{\hat{\mathcal{R}}} = w_{\mathcal{R}}|_{\hat{\mathcal{R}}}$

WAF $\langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ with $\hat{\mathcal{R}} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\mathcal{R}}(\alpha, \beta) < 0\} \subseteq \mathcal{A} \times \mathcal{A}$ and $w_{\hat{\mathcal{R}}} = -w_{\mathcal{R}}|_{\hat{\mathcal{R}}}$

BAF $\langle \mathcal{A}, \hat{\mathcal{R}}_{att}, \hat{\mathcal{R}}_{sup} \rangle$ with $\hat{\mathcal{R}}_{att} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\hat{\mathcal{R}}}(\alpha, \beta) < 0\}$ and $\hat{\mathcal{R}}_{sup} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\hat{\mathcal{R}}}(\alpha, \beta) > 0\}$

AF $\langle \mathcal{A}, \hat{\mathcal{R}} \rangle$ with $\hat{\mathcal{R}} = \{(\alpha, \beta) \in \mathcal{A} \times \mathcal{A} \mid w_{\hat{\mathcal{R}}}(\alpha, \beta) > 0\} \subseteq \mathcal{A} \times \mathcal{A}$

4 Adding User and Topic Information

To fully exploit the extended expressive power of GAFs, the two components $\mathcal{S}(\mathcal{A})$ and $w_{\mathcal{A}}$ must be defined. In particular, $\mathcal{S}(\mathcal{A})$ must be preliminarily defined, since it is also used in the definition of $w_{\mathcal{A}}$. While the knowledge engineer is totally free in defining such component, we still believe that the features proposed

in [12] are sensible and useful. Indeed, we expect the interrelations existing in the community in which the argumentation takes place, and the topic about which the claims are made, to almost always significantly affect the evaluation of arguments.

For this reason, having defined the overall GAF model, we propose here a first specialization of it, which is still very general and abstract, but introduces some fundamental features that would probably be relevant to most practical cases of argumentation: community and topics. To model these features, we define T-GAFs⁶, that introduce these first two components in $\mathcal{S}(\mathcal{A})$:

- \mathcal{U} the finite set of members of the community, possibly including the entities who put forward the arguments, and
- \mathcal{T} a finite set of topics that may be involved in an argumentation.

For practical purposes, we propose to consider \mathcal{T} as always including an additional dummy topic \top associated to the general authority and trust of a user, independent of specific topics. So, formally, $\mathcal{T} = \overline{\mathcal{T}} \cup \{\top\}$, where $\overline{\mathcal{T}}$ is the set of specific topics that may be involved in an argumentation.

Now, also based on these components in $\mathcal{S}(\mathcal{A})$, some components to be used in $w_{\mathcal{A}}$ can be defined, as well. Possible features to combine in $w_{\mathcal{A}}$ are:

1. the subjective *confidence* that the members of the community (including the entity which posits the argument) have in an argument;
2. the recognized *authority* degree of the entity putting forward an argument on the topic of the argument⁷;
3. the *trust* that the community of entities involved in the argumentation have in the entity putting forward an argument, relative to the topic of the argument (indeed, not just the quality of evidence, but also the credibility of the entity positing it is important).

While (2) expresses the degree of expertise of an entity about a topic (e.g., medicine), (1) expresses the degree of confidence about a specific claim, and (3) the degree of confidence by which a user's opinions about a topic are taken into consideration by other users. E.g., if Joe, a MD, posits the argument “I am *quite confident* that a vaccine for COVID-19 disease will be available before the end of year 2020”, we may consider: via (1), a degree of uncertainty expressed by Joe himself about the validity of the argument, in the phrase “quite confident” (which might be translated into an entity's confidence degree on that argument of 0.7), and different degrees of confidence of the various members of the community with respect to that claim (some will more or less agree with Joe, some will more or less disagree); via (2), a degree of authority of Joe about medicine (let's

⁶ We call it T-GAF for analogy to T-BWAF, since it is aimed at introducing the same components as T-BWAF, but again at generalizing them so that several frameworks, including T-BWAFs, can be expressed in the model.

⁷ E.g., the education or skill level of the user on that topic —opinions of experts in a topic are typically more convincing than those of novices or outsiders of the topic.

say it's 0.8, since Joe is a MD); and via (3), a degree of trust of the community for Joe as a doctor (many people might consider him not a very good doctor).

The 3 features above are formalized by the following functions:

1. $w_c : \mathcal{U} \times \mathcal{A} \mapsto [0, 1]$ where 1 means certainty, according to the classical probabilistic interpretation.
2. $w_a : \mathcal{U} \times \mathcal{T} \mapsto [0, 1]$ where 1 means maximum authority of the user in the topic, and 0 absolutely no authority.
3. $w_t : \mathcal{U} \times \mathcal{T} \mapsto [-1, 1]$ where -1 means total distrust, 0 means no opinion, and 1 means full trust

Functions 1 and 2 might be defined extensionally, by directly associating a value to each input pair based on the available information. E.g., feature 1 is quite subjective, and the values might be obtained by asking the single members of the community; feature 2 might be assessed based on the formal certifications owned by the arguer about the given topic (e.g., BSc, MS, PhD, etc.) Feature 3 is more complex, because it must be based on a formal model of trust that might involve many direct and indirect trust evaluations between the members of the community. We propose a graph-based formal model of trust based on the following definition⁸:

Definition 8 (Community Trust Graph) *Given community \mathcal{U} , a Community Trust Graph (or CTG) for \mathcal{U} is a complete directed weighted graph $G = \langle \mathcal{U}, \mathcal{E}, w_{\mathcal{E}} \rangle$ where:*

- \mathcal{U} is the set of members in the community,
- $\mathcal{E} = \mathcal{U} \times \mathcal{U}$ is the complete set of edges,
- $w_{\mathcal{E}} : \mathcal{E} \mapsto [-1, 1]$ is a function that, given two members $u_1, u_2 \in \mathcal{U}$, expresses the trust $w_{\mathcal{E}}(u_1, u_2)$ that member u_1 has for member u_2 (where -1 means total distrust, 0 means no opinion, and 1 means full trust).

Like for the GAF definition, we consider a complete graph for the sake of formalization simplicity and for allowing a more straightforward translation of the graph into matrix representation. Again, for practical purposes, edges having 0 weight can be ignored and removed from the graphical representation. Using a $[-1, 1]$ range for trust provides the same computational advantages as in the case of GAF. Indeed, the sign rule can again leveraged to handle the fact that, if u distrusts v and in turn v distrusts s , then this might be taken as a hint that u might somehow trust s .

Given a community \mathcal{U} , and a CTG G for \mathcal{U} , the overall trust for each member of \mathcal{U} according to G , possibly based on the direct and indirect trust information expressed by G , can be determined by evaluating a function, say $t(u, G) \in [-1, 1]$ (where the range $[-1, 1]$ was chosen for compliance with $w_{\mathcal{E}}$).

⁸ Compared to [12], here we adopt a $[-1, 1]$ range for trust, which is more intuitive and provides computational advantages, and dismiss the weights on the nodes, that in our model are recovered by function w_{α} . Indeed, the authority of a user is not necessarily related to the subjective trust the community has in the user.

So, a T-GAF includes a CTG for each topic $T \in \mathcal{T}$ (let us call it G^T), and assessing the trust of a user u for T corresponds to computing

$$w_t(u, T) = t(u, G^T) \in [-1, 1]$$

Finally, given specific definitions for functions w_c , w_a and w_t (for the various topics), an overall assessment $w_{\mathcal{A}}$ of the ‘intrinsic’ reliability of an argument in the GAF can be obtained by applying a function that combines all these perspectives together.

Example 1. To show a possible practical application of the GAF, let us express in the GAF model the aggregation function proposed in [12] for T-BWAFs. For reference and for the sake of comparison, we will report in footnotes the definitions of the various functions in T-BWAFs. Consider an argument α , posited by user u and concerning topic T . Then,

$$w_{\mathcal{A}}(\alpha) = \beta \cdot w_c(u, \alpha) \cdot \max(\min_{v \neq u} w_{\mathcal{E}}^T(v, u), w_a(u)) + (1 - \beta) \cdot ca(\alpha)$$

where $\beta \in [0, 1]$ and the following notational correspondence was applied:

- $w_c(u, \alpha) \equiv \text{conf}(u, \alpha)$, called the ‘User Argument Confidence’ in [12]⁹
- $\max(\min_{v \neq u} \{w_{\mathcal{E}}^T(v, u)\}, w_a(u)) \equiv \text{authority}(u)$, called the ‘Authority Degree’ in [12]¹⁰, and also based on the Trust Users Graph as in GAFs, where $t(u, G^T) = \min_{v \neq u} \{w_{\mathcal{E}}^T(v, u)\}$
- $w_{\mathcal{A}}(\alpha) \equiv \text{strength}(\alpha)$, called the ‘Argument Strength’ in [12]¹¹
- $ca(\alpha)$, called the ‘Crowd’s Agreement’ of the community in [12], is implemented, following [9], as the Simple Vote Aggregation function¹²:

$$ca(\alpha) = \begin{cases} 0 & \text{if } V^+(\alpha) = V^-(\alpha) = 0 \\ \frac{V^+(\alpha)}{V^+(\alpha) + V^-(\alpha)} & \text{otherwise} \end{cases}$$

⁹ Note that $\text{conf}: \mathcal{U} \times \mathcal{A} \mapsto [0, 1]$, whereas w_c ranges in $[-1, 1]$.

¹⁰ Given a Trust Users Graph $T = \langle \mathcal{U}, \mathcal{E}, w_{\mathcal{U}}, w_{\mathcal{E}} \rangle$, the *Authority Degree* of a user $u \in \mathcal{U}$ is defined as:

$$\text{authority}(u) = \max(\min_{v \neq u} \{w_{\mathcal{E}}(v, u), w_{\mathcal{U}}(u)\})$$

¹¹ Let $F = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\mathcal{A}}, w_{\hat{\mathcal{R}}}, \mathcal{K}, \text{conf} \rangle$ be a T-BWAF, $a \in \mathcal{A}$ an argument, $u \in \mathcal{U}$ a user, $i \in \mathcal{T}$ a topic, the intrinsic *Argument Strength* is defined as

$$\text{strength}(a) = \alpha \cdot \text{conf}(u, a) \cdot \text{authority}_i(u) + (1 - \alpha) \cdot w_{\mathcal{A}}(a), \text{ with } \alpha \in [0, 1].$$

¹² As defined in [12], $ca(\alpha)$ takes values in $[0, 1]$, and only considers positive votes. In the T-GAF framework, it would be more consistent (and a more appropriate approach in general) to consider also negative votes, using the following formula:

$$\frac{V^+(\alpha) - V^-(\alpha)}{V^+(\alpha) + V^-(\alpha)}$$

which takes values in $[-1, 1]$ (specifically, -1 means that all votes are negative, and $+1$ means that all votes are positive).

where $V^+(\alpha)$ and $V^-(\alpha)$ denote, respectively, the number of positive and negative votes for argument $\alpha \in \mathcal{A}$. In the T-GAF model, they can be expressed in terms of w_c as follows:

- $V^+(\alpha) = |\{u \in \mathcal{U} \mid w_c(u, \alpha) > 0\}|$
- $V^-(\alpha) = |\{u \in \mathcal{U} \mid w_c(u, \alpha) < 0\}|$

5 Matrix Representation

As for BWAFs in [13]¹³, we propose a matrix representation for GAFs. Indeed, in addition to providing a comfortable representation that is also consistent with intuition, matrices also provide an efficient computational tool for supporting many argument evaluation-related tasks, and may even suggest new semantics, especially in the extended framework where computations on argument and relationship weights are needed.

Definition 9 Let $F = \langle \mathcal{A}, \mathcal{S}(\mathcal{A}), w_{\mathcal{A}}, w_{\mathcal{R}} \rangle$ be a GAF with $|\mathcal{A}| = n$. Then, the General Argumentation Matrix of F is an $n \times n$ matrix $\mathbf{M}_F = [m_{ij}]$ such that

$$\forall \alpha_i, \alpha_j \in \mathcal{A} : m_{ij} = w_{\mathcal{R}}(\alpha_i, \alpha_j)$$

Note that this representation for GAFs is even more straightforward than for BWAFs, since the 0 value for pairs of arguments having no relationship is explicit in the formalization of GAFs, while in BWAFs it must be handled as a default case.

For the same reasons as for GAFs, we propose to use the same matrix representation also for Community Trust Graphs:

Definition 10 Let $G = \langle \mathcal{U}, \mathcal{E}, w_{\mathcal{E}} \rangle$ be a CTG with $|\mathcal{U}| = n$. Then, the Community Trust Matrix of G is an $n \times n$ matrix $\mathbf{M}_G = [m_{ij}]$ such that

$$\forall u_i, u_j \in \mathcal{U} : m_{ij} = w_{\mathcal{E}}(u_i, u_j)$$

Example 2. The GAF G in Figure 2-a has the following matrix representation:

$$\mathbf{M}_G = \begin{matrix} & \alpha & \beta & \gamma & \delta & \epsilon \\ \alpha & \begin{bmatrix} 0 & 0.4 & 0 & 0 & -0.7 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 & 0.3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \beta & & & & & \\ \gamma & & & & & \\ \delta & & & & & \\ \epsilon & & & & & \end{matrix}$$

¹³ In BWAFs, given $F = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ a BWAF with $|\mathcal{A}| = n$, the *Signed Weighted Argumentation Matrix* of F is defined as a $n \times n$ matrix $\mathbf{M}_F = [m_{ij}]$ such that

$$\forall \alpha_i, \alpha_j \in \mathcal{A} : m_{ij} = \begin{cases} w_{\hat{\mathcal{R}}}(\langle \alpha_i, \alpha_j \rangle) & \text{if } \langle \alpha_i, \alpha_j \rangle \in \hat{\mathcal{R}} \\ 0 & \text{otherwise} \end{cases}$$

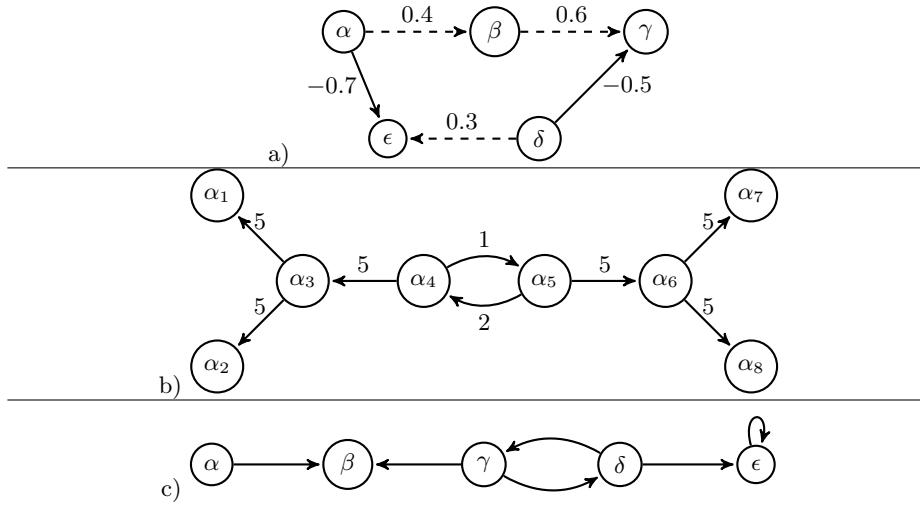


Fig. 2. Sample argumentation frameworks: (a) a GAF G , (b) a WAF W , (c) an AF F

Note that the graph for G can also be interpreted as a Community Trust Graph, where $\{\alpha, \beta, \gamma, \delta, \epsilon\}$ are the members in the community, solid edges denote negative trust between members, and dashed edges represent positive trust between members (and the weight represent the magnitude of the trust). Under this interpretation, \mathbf{M}_G is the matrix representing the community trust.

The GAF G is clearly also a BWAF. Ignoring the weights in G , we have a BAF B with the following GAF matrix representation:

$$\mathbf{M}_B = \begin{matrix} & \alpha & \beta & \gamma & \delta & \epsilon \\ \alpha & \begin{bmatrix} 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \beta & \\ \gamma & \\ \delta & \\ \epsilon & \end{matrix}$$

The GAF matrix representation for the WAF W in Figure 2-b is:

$$\mathbf{M}_W = \begin{matrix} & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ \alpha_1 & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & -0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.4 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \alpha_2 & \\ \alpha_3 & \\ \alpha_4 & \\ \alpha_5 & \\ \alpha_6 & \\ \alpha_7 & \\ \alpha_8 & \end{matrix}$$

where weights were normalized with respect to $\max_{\alpha, \beta \in \mathcal{A}} w_{\mathcal{R}}(\alpha, \beta) = 5$.

Finally, the GAF matrix representation of the AF F in Figure 2-c is:

$$\mathbf{M}_F = \begin{matrix} & \alpha & \beta & \gamma & \delta & \epsilon \\ \alpha & \begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \\ \beta & \\ \gamma & \\ \delta & \\ \epsilon & \end{matrix}$$

As regards the argument weights assigned by $w_{\mathcal{A}}$, they can be collected in a vector, indexed exactly like the argumentation matrix rows and columns, which allows their easy combination through standard matrix operators. E.g., for the GAF G in Example 2:

$$\mathbf{M}_G^{\mathcal{A}} = \begin{bmatrix} \alpha & \beta & \gamma & \delta & \epsilon \\ w_{\alpha} & w_{\beta} & w_{\gamma} & w_{\delta} & w_{\epsilon} \end{bmatrix}$$

6 Conclusion

The classical definition of Abstract Argumentation Frameworks considers only attacks between arguments, based on which different evaluation strategies ('semantics') have been proposed to identify the subsets of arguments that consistently support the same position in a dispute or debate ('extensions'), and possibly determine the winning position. However, in complex situations, additional information may be important to properly describe the debate and take better decisions. This led to the definition of extended frameworks, among which bipolar (considering also supports among arguments), and weighted ones (allowing to assign different importance to the attacks). Since some of these extended frameworks are partly incompatible, or anyhow disjoint, this paper proposed GAFs, a general model that encompasses all of them, and further extends them by allowing to express weights also on arguments, based on contextual information. In particular, we propose that the extended framework includes at least information about authority of users, their subjective confidence in the arguments, and the mutual truth of members in the community. Here we do not propose specific semantics for GAFs. However, since the previous models can be represented as GAFs, the semantics defined for the previous models can be also applied to GAFs. Moreover, new ones can be defined that exploit its extended expressiveness. The definition of GAFs allows a straightforward matrix representation, that allows the use of matrix operations to improve efficiency in the evaluation of arguments, and perhaps to define new semantics.

In the future, we will define new semantics that can exploit the full expressive power of GAFs. We also would like to investigate its relationships to other AFs proposed in the literature, and to identify other relevant specializations of it.

References

1. Amgoud, L., Ben-Naim, J.: Ranking-based semantics for argumentation frameworks. In: SUM. pp. 134–147. Springer (2013)
2. Amgoud, L., Ben-Naim, J.: Axiomatic foundations of acceptability semantics. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning, KR. vol. 16 (2016)
3. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Annals of Mathematics and Artificial Intelligence **34**(1-3), 197–215 (2002)
4. Amgoud, L., en Naim, J.: Weighted bipolar argumentation graphs: axioms and semantics. In: IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence (2018)
5. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. Journal of Logic and Computation **13**(3), 429–448 (2003)
6. Cayrol, C., Lagasquie-Schiex, M.: On the acceptability of arguments in bipolar argumentation frameworks. In: ECSQARU. vol. 3571, pp. 378–389 (2005)
7. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. Artificial intelligence **77**(2), 321–357 (1995)
8. Dunne, P.E., Hunter, A., McBurney, P., Parsons, S., Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artificial Intelligence **175**(2), 457 – 486 (2011)
9. Leite, J., Martins, J.: Social abstract argumentation. In: International Joint Conference on Artificial Intelligence. vol. 11, pp. 2287–2292 (2011)
10. Parsons, S., Tang, Y., Sklar, E., McBurney, P., Cai, K.: Argumentation-based reasoning in agents with varying degrees of trust. In: The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2. pp. 879–886. International Foundation for Autonomous Agents and Multiagent Systems (2011)
11. Pazienza, A., Ferilli, S., Esposito, F.: Constructing and evaluating bipolar weighted argumentation frameworks for online debating systems. In: Proceedings of the 1st Workshop on Advances In Argumentation In Artificial Intelligence, AI^3 2017 @ AI*IA 2017. pp. 111–125 (2017)
12. Pazienza, A., Ferilli, S., Esposito, F.: On the gradual acceptability of arguments in bipolar weighted argumentation frameworks with degrees of trust. In: Foundations of Intelligent Systems - 23rd International Symposium, ISMIS. pp. 195–204 (2017)
13. Pazienza, A., Ferilli, S., Esposito, F.: Synthesis of argumentation graphs by matrix factorization. In: Proceedings of the 1st Workshop on Advances In Argumentation In Artificial Intelligence, AI^3 2017. pp. 1–6 (2017)
14. Toulmin, S.E.: The Uses of Argument. University Press (1958), <https://books.google.it/books?id=WffWAAAAMAAJ>