Robustness in Single-Audience Value-based Abstract Argumentation: Complexity Results

Bettina Fazzinga[1*] , **Sergio Flesca**[2] and **Filippo Furfaro**[2]
[1]DiCES - University of Calabria
[2]DIMES - University of Calabria
{bettina.fazzinga, sergio.flesca, filippo.furfaro}@unical.it

Abstract

We address the context of *Single-Audience Value-Based Abstract Argumentation Framework* (AVAF), where the arguments are labeled with the social values that they promote and the activation/deactivation of the attacks depends on the audience profile (expressed as a set of preferences between the social values). Herein, we introduce a new notion of *robustness* for measuring the sensitivity of the outcome of the reasoning to the extent of changes in the audience profile. In particular, for a set of arguments S or a single argument a, we define the *robustness degree* of the status of S or a as the maximum number k^* of deletions/insertions of preferences from/into the audience profile that are *tolerable*, in the sense that S remains an extension (or a non-extension) or a accepted (or unaccepted) after performing at most k^* deletions/insertions. We introduce the decision problems related to the computation of the robustness degree and focus on thoroughly investigating their computational complexity.

1 Introduction

The well-known *Abstract Argumentation Framework* (AAF) [Dung, 1995], largely employed for reasoning over disputes and solving problems that can be translated into them [Alfano *et al.*, 2024a; Fazzinga *et al.*, 2022b; Fazzinga *et al.*, 2022a; Fazzinga *et al.*, 2021b] has been extended in a number of ways. Specifically, several efforts have been made in the direction of taking into account the 'strength' of the arguments of the dispute. In particular, in *Preference-based Argumentation Framework* (PAF) [Amgoud and Cayrol, 2002], the fact that some arguments are perceived as stronger than others by the audience to which the arguments are addressed is encoded with a set of preferences, whose effect on the reasoning is the deactivation of any attack where the attacked is preferred to the attacker. Then, in *(single-Audience) Value-based Argumentation Frameworks* (AVAFs) [Bench-Capon, 2003], the difference in strength between the arguments is modeled as a consequence of the fact that the arguments may promote different social values, and the audience has preferences between these social values, as illustrated in the following example.

Example 1 *An analyst is reasoning on a dispute whose arguments promote the following social values:* "Generosity *(G)",* "Fairness *(F)",* "Accountability *(A)". The arguments are:*

- *a (promoting* G*):* "John is unemployed, and he is not been searching for a job for some years. His family is very poor, so we should donate some money to him*";*
- *b (promoting* F*):* "Redistributing wealth without considering who contributed to its creation fosters laziness and leads to impoverishment. So, grant no compensation to people who made no effort to improve their circumstances*";*
- *c (promoting* A*):* "John did not allow us to verify his family's income, so it is not appropriate to make any decision based on the assumption that his family is poor*".*

Independently from the values promoted by the arguments, it is easy to detect, by looking into the meaning of the arguments, that a and b attack each other, and that c attacks a. So, by encoding the dispute in a classical AAF with the above arguments and attacks, the analyst would conclude that $\{b\}$, $\{c\}$, and $\{b, c\}$ are admissible extensions, while $\{a\}$ is not, and that c and b are accepted arguments, while a is not.

Now, the analyst wants to take into account the standpoint of a specific audience, whose perception of the existence of the attacks is implied not only by the meaning of the arguments, but also by the "strength" that the audience attributes to the arguments. So, the analyst turns the AAF into an AVAF, where the audience profile is described by the following preferences between values: G>F *and* G>A*, where the meaning of $X > Y$ is: "the audience will* ***always*** *consider* ***any*** *argument x promoting X stronger than* ***any*** *y promoting Y, and thus* ***any*** *attack from y to x ineffective".*

Thus, in the AVAF obtained this way, the attacks from b to a and from c to a are inactive and are excluded from the reasoning, so the analyst will conclude that, from the standpoint of the audience, b is not an accepted argument, while a is, and $\{a\}$, $\{c\}$ and $\{a, c\}$ are admissible extensions.

Example 1 suggests the reason why AVAFs have proved effective in several scenarios, such as promotional campaigns and trials: they allow the analyst to simulate the subjective views of the people to which the arguments are addressed

[*] Contact Author

(e.g. the target population, in the case of promotional campaigns, or the jury, in the case of trials) and then to reason on how the audience will perceive the status of the arguments.

However, the outcome of the reasoning performed over an AVAF could be affected by reliability issues, since:
1) the audience profile may not be accurate, as the beliefs and social convictions of the involved individuals are not always easy to predict; so, some preferences put in the audience profile by the analyst may be wrong and/or some preferences actually characterizing the audience may be missing;
2) even if, initially, the audience profile is accurately modeled, the audience may change their opinions over time and, therefore, the importance they attribute to values.

A possible way for assessing the reliability of the result of the reasoning is studying its "*robustness*": once the status of a set of arguments (is it an extension or not?) or of a single argument (is it accepted or not?) has been determined over an AVAF, its "**robustness degree**" is the maximum number of changes to the audience profile that are tolerated, in the sense that no modification of the status would be observed if any set of changes of this cardinality were performed.

Intuitively, the higher the degree of robustness, the greater the reliability of the result of the reasoning: if many changes to the profile are required to alter the outcome, it means that, even if there were errors in defining the profile or changes occurred since it was defined, it is unlikely that such errors and changes would be numerous enough to affect the result.

Example 2 *(continuing Example 1) There are ten more arguments $x_1, \cdots, x_{10}$, where: x_{10} attacks and is attacked by c; each x_i promotes a distinguished value v_i, and attacks and is attacked by a. Figure 1 (a) shows the AVAF modeling this situation, when the audience profile has not been defined yet, so all the attacks are active. Suppose that now the analyst models the audience profile as follows: F>G, A>G, A> v_{10} and, for each $i \in [1..10]$, v_i >G. Since all the attacks stemming from a and the attack from x_{10} to c are deactivated by the preferences, the active attacks become those in Figure 1(b).*

Reasoning on this AVAF, the analyst (who adopts the admissible semantics) concludes that b and c are both (credulously) accepted. However, by looking into the robustness of the acceptance of b and c, an evident difference emerges: b's acceptance is way more robust that c's. In fact, making b unaccepted requires at least 23 changes to the set of preferences, while two changes suffice to make c unaccepted. The reason is that invalidating b's acceptance requires deactivating the attack (b, a), activating (a, b), and deactivating all the defenses against a; this can be done by deleting all the preferences except A> v_{10}, and then adding the 'opposite' preferences G>F, G>A and G> v_i, for each $i \in [1..9]$ (note that, this way, G> v_{10} is implied). *This means 12 deletions and 11 insertions of preferences, which yield the AVAF in Figure 1(c). On the other hand, simply inactivating the attack (c, x_{10}) and activating (x_{10}, c) makes c unaccepted, and this requires deleting A> v_{10} and adding $v_{10} >$ A (that are 2 changes in total). It is easy to see that it is not possible to change the status of b and c with less than 23 and 2 insertions/deletions, respectively. Hence, the robustness degrees of b's and c's acceptance are 22 and 1, respectively.*

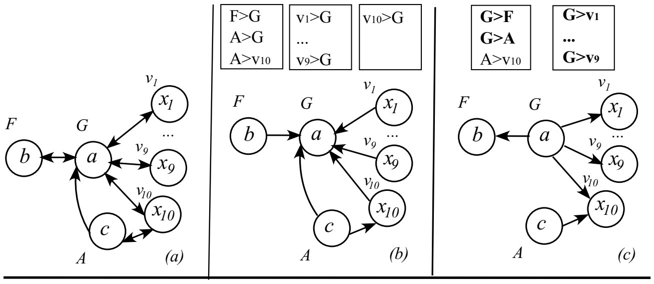

Figure 1: The arguments and the active attacks of the AVAF of Example 2: (a) initially, when the audience profile is empty (so all the attacks are active); (b) after an audience profile has been specified; (c) after the audience profile has been changed. The symbol beside each argument is the promoted value. The boxes on the top of each graph contain the preferences describing the audience.

Given this, based on the different robustness degrees, the analyst may conclude that, in light of the risk that the audience profile might contain inaccuracies, relying on the acceptance of b is much more cautious than on the acceptance of c.

Using the same reasoning, it is easy to see that both x_{10} and a are unaccepted in the AVAF in Figure 1(b), but these non-acceptance statuses have different robustness (0 for x_{10} and 12 for a): only one change to the profile (the deletion of A> v_{10}) makes x_{10} accepted, while making a accepted requires at least 13 changes: deleting all the preferences but A $>$ v_{10}, and adding G>A. These different robustness degrees suggest to the analyst that, even if the audience turns out to be somewhat different from what was specified in the model, it is unlikely that the status of a returned by the AVAF would differ significantly from what is perceived by the real audience, because that would mean the real audience is much different from what was modeled. On the other hand, with x_{10}, this risk is much more real.

The notion of robustness can be applied also to extensions and non-extensions: for instance, $S_1 = \{c\}$ and $S_2 = \{b, x_1, \ldots, x_9\}$ are admissible extensions in the AVAF of Figure 1(b), and their robustness degrees are 1 and 19, respectively (deleting A $>$ v_{10} and inserting $v_{10} >$ A makes S_1 a non-extension, while performing the 20 updates needed to invert the preference F $>$ G and every v_i $>$ G, with $i \in [1..9]$, makes S_2 a non-extension).

The contribution of this paper is the introduction of a new paradigm for reasoning on the robustness of the status of (sets of) arguments, which is a relevant issue since, as shown in Example 2, it can effectively help assess the reliability of the outcome of the reasoning over AVAFs. In particular, we introduce the decision problems underlying the computation of the degree of robustness, and thoroughly investigate their computational complexity. The novelty of the contribution lies in the fact that robustness has never been studied in the literature of abstract argumentation with regard to audience profiling. In fact, it was addressed only in the case of standard AAFs (thus disregarding values and preferences): specifically, in [Rienstra *et al.*, 2020], where the resilience of the acceptance status w.r.t. the addition/removal of attacks has been used as a principle for comparing different admissibility-based semantics;

in [Rapberger and Toni, 2024], in terms of the property of explanations of being still valid after changes of the argumentation graph; as well as, indirectly, in the framework of enforcement of extensions and of acceptability constraints [Doutre and Mailly, 2018].

2 Preliminaries

Abstract Argumentation Frameworks (AAFs). An *abstract argumentation framework* (*AAF*) is a pair $F = \langle A, D \rangle$, where A is a finite set of *arguments* and $D \subseteq A \times A$ is a binary relation, whose elements are called *attacks*. Given a set of arguments S and an argument a, we say that "*S attacks a*" if there is b in S such that b attacks a, and that "*a attacks S*" if there is $b \in S$ such that a attacks b. We say that an argument a (resp., a set of arguments S) "*defends b against c's attack*" if c attacks b, while a (resp., S) attacks c, and that *a is acceptable w.r.t. S* if S defends a against every attack. The "*argumentation graph*" of F is the directed graph whose nodes and edges are the arguments and the attacks of F.

Several semantics for AAFs have been proposed to identify "reasonable" sets of arguments, called *extensions* [Dung, 1995]. A set $S \subseteq A$ is an extension of type: *conflict-free* ($\mathtt{cf}$) if there is no attack between its arguments; *admissible* ($\mathtt{ad}$) if S is conflict-free and its arguments are acceptable w.r.t. S; *stable* ($\mathtt{st}$) if S is conflict-free and S attacks every $a \in A \setminus S$; *complete* ($\mathtt{co}$) if S is admissible and contains all the arguments acceptable w.r.t. S; *grounded* ($\mathtt{gr}$) if S is a minimal (w.r.t. $\subseteq$) complete extension; *preferred* ($\mathtt{pr}$) if S is a maximal (w.r.t. $\subseteq$) complete extension. $\mathtt{cf}$, $\mathtt{ad}$, $\mathtt{st}$, $\mathtt{gr}$, $\mathtt{co}$, $\mathtt{pr}$ will be referred to as *Dungean semantics*. Arguments belonging to at least one (resp., every) extension are said to be *Credulously* (*Cr*-) accepted (resp., *Skeptically* (*Sk*-) accepted). The fundamental problems supporting the reasoning over AAFs are the *verification* and the *acceptance* problems $\text{VER}^{\sigma}(F, S)$ and $\text{ACC}^{\sigma}(F, a, X)$, asking if S is an extension and if a is X-accepted, with $X \in \{\text{Cr}, \text{Sk}\}$, respectively.

Preferences, Preference Relation, and Preference-based Argumentation Framework (PAF). A *preference* between two elements x_1, x_2 of a set X is an expression "$x_1 > x_2$", to be read as "*x_1 is preferred to x_2*". A (strict) *preference relation* on a set X is a partial, transitive, asymmetric, irreflexive relation $P \subset X \times X$. Each $(x_1, x_2) \in P$ is interpreted as a preference $x_1 > x_2$. If $(x_1, x_2) \notin P$, we write $x_1 \not> x_2$.

A *PAF* [Amgoud and Cayrol, 2002] is a triplet $PF = \langle A, D, P \rangle$, where $\langle A, D \rangle$ is an AAF and P is a (strict) preference relation on A. The effect of preferences is the *inactivation* of any attack (a, b) where $b > a$. Thus, the extensions and the accepted arguments of $PF = \langle A, D, P \rangle$ are those of the "*implied*" AAF $F = \langle A, D' \rangle$, where D' is the subset of D containing only its *active* attacks, i.e. the attacks (a, b) with $b \not> a$.

Single-Audience Value-based Argumentation Framework (AVAF). AVAF [Bench-Capon, 2003] extends the traditional reasoning paradigm over AAF by taking into account the audience profile. This profile is a set of preferences on the set of social values involved in the dispute to be modeled, and implies a set of preferences on the arguments, so that reasoning on an AVAF reduces to reasoning over the PAF where the preference relation over the arguments is that implied by the preferences specified in the AVAF.

Formally, given a set of values V, an *(audience) profile* (or *preference specification*) π over V is a set of *preferences* of the form $v_1 > v_2$, with $v_1, v_2 \in V$. In graph terms, π is represented by the digraph, called *preference graph*, having V and $\{(v_1, v_2) | v_1 > v_2\}$ as sets of nodes and arcs. We denote as π^* the transitive closure of π. A profile π is said to be *consistent* iff it is acyclic, meaning that π^* is a (strict) preference relation over V. For instance, the profile $\{v_1 > v_2, v_2 > v_3, v_3 > v_1\}$ is not consistent: observe that its transitive closure is not a strict preference relation, since it contains the preferences $v_1 > v_3$ and $v_3 > v_1$ (thus violating the asymmetry).

An AVAF is a tuple $VF = \langle A, D, V, val, \pi \rangle$, where $\langle A, D \rangle$ is an AAF, V a set of values, π a consistent preference specification over V, and $val : A \to V$ a total function associating arguments with values. The semantics of an AVAF $VF = \langle A, D, V, val, \pi \rangle$ is given by the PAF $PF = \langle A, D, P(\pi) \rangle$, where $P(\pi)$ is the strict preference relation implied by π, i.e. $P(\pi) = \{(a, b) | (val(a) > val(b)) \in \pi^*\}$. Thus, the extensions and the accepted arguments of *VF* are those of *PF*, or, equivalently, of the AAF implied by *PF*, that are called the PAF and the AAF "implied" by *VF*, respectively.

2.1 Further Notions and Notations

Given a preference specification π over a set of values V, we consider two primitive update operations over π: the insertion $ins(v_1 > v_2)$ and the deletion $del(v_1 > v_2)$, which inserts and removes the preference $v_1 > v_2$ into and from π, respectively. In turn, we call "*preference update*" any set U of primitive update operations, and define the application of U to π as $U(\pi) = \pi \cup \{v_1 > v_2 | ins(v_1 > v_2) \in U\} \setminus \{v_1 > v_2 | del(v_1 > v_2) \in U\}$. In turn, we define the application of a preference update U to an AVAF $VF = \langle A, D, V, val, \pi \rangle$ as the AVAF $U(VF) = \langle A, D, V, val, U(\pi) \rangle$.

A preference update U over π is said to be *consistent* if $U(\pi)$ is consistent. U is said to be "del-only" (resp., "ins-only") if it contains only deletions (resp., insertions). Obviously, a del-only preference update over a consistent profile is always consistent.

3 Problem Statement: Reasoning on the Robustness over AVAFs

We now introduce the fundamental problems that support the analysis on the "robustness" of the outcome of the reasoning over an AVAF. We address the case where an AVAF has been already defined, and the analyst wants to investigate the consequences of some modifications of the audience profile, due to changes in mind of the audience or to the need of fixing inaccuracies w.r.t. the actual subjective view of the audience. Herein, for a set of arguments S and an argument a, "robustness" means the number of changes to the audience profile that have no impact on the status of S and a, in terms of being or not an extension and accepted, respectively.

Definition 1 (**k-robustness problems**) $\text{RE}^{\sigma}(VF, S, k)$ *(resp.,* $\text{RNE}^{\sigma}(VF, S, k)$*) is the problem of checking if the set S, that is (resp., is not) an extension of the AVAF VF under σ,* *is still (resp., is still not) an extension of $U(VF)$ under σ for every consistent preference update U over π with $|U| \leq k$.*

$\mathrm{RA}^\sigma(VF, a, X, k)$ *(resp.,* $\mathrm{RNA}^\sigma(VF, a, X, k)$*), where $X \in \{Cr, Sk\}$, is the problem of checking if the argument a, that is (resp., is not) X-accepted for the AVAF VF under σ, is still (resp., is still not) X-accepted for $U(VF)$ under σ for every consistent preference update U over π with $|U| \leq k$.*

The suffixes -IO and -DO will denote the variants of the robustness problems restricted to ins-only and del-only preference updates. Restricting to ins-only (resp., del-only) updates means assessing the robustness in the case where the audience profile may be strictly more specific (resp., strictly more general) than what initially specified by the analyst.

Remark 1: on the encoding of the audience. It is worth noting that reasoning on the robustness over AVAFs has required a modification of the definition of AVAF introduced in [Bench-Capon, 2003], where the subjective view of the audience is encoded in terms of a (strict) preference relation, while, in our framework, it is encoded in terms of a (consistent) set of preferences. In fact, the preference relation in [Bench-Capon, 2003] is the transitive closure of the preference specification occurring in "our" AVAF. The point is that these two encodings are equally suitable when no change of preference occurs, as in [Bench-Capon, 2003]. But the encoding based on a set of preferences is more suitable for measuring the extents of the modifications of what is believed about the subjective view of the audience. The reason is that two sets of preferences π_1, π_2 may correspond to the same preference relation (that is, π_1^* and π_2^* coincide), but they may not coincide. For instance, consider $\pi_1 = \{v_1 > v_2, v_2 > v_3\}$ and $\pi_2 = \{v_1 > v_2, v_2 > v_3, v_1 > v_3\}$. Since preferences are implied by transitivity, we have $\pi_1^* = \pi_2^*$, but the fact that $v_1 > v_3$ occurs in π_2 means that the analyst who models the audience with π_2 explicitly believes that the audience prefers value v_1 to v_3. Thus, when counting the changes to the profile of the audience needed to remove $v_1 > v_3$, if π_1 is adopted, we obtain that only one deletion is necessary (either $del(v_1 > v_2)$ or $del(v_2 > v_3)$, since this way $v_1 > v_3$ is no more implied), while if π_2 is adopted, at least two deletions are needed, since also $del(v_1, v_3)$ should be performed. And this difference reflects the fact that in π_2 the preference $v_1 > v_3$ is explicitly stated, while in π_1 it is a consequence of what is believed. Given this, if the audience profile were encoded in terms of the preference relation, it would be impossible to distinguish the scenarios corresponding to π_1 and π_2.

Remark 2: on assuming the initial status known. In Definition 1, the "initial" status of S and a is assumed known. In fact, from a practical perspective, it is natural that the analyst starts to reason on the robustness of the status of (a set of) arguments after the initial status has been determined. From a technical perspective, this has no impact on the importance of our results on the computational complexity presented in the following section. On the one hand, Definition 1 allows us to "isolate" the source of complexity related to assessing the initial status from the complexity of reasoning on the effects of changing the audience profile. On the other hand, the computational complexity of the variant of the k-robustness problems where the initial status is not known can be obtained by properly combining the complexities of the standard verification and acceptance problems Ver and Acc with those of the k-robustness problems studied in this paper.

Verification Problem

σ	Ver	RE	RE	RE	RNE	RNE	RNE
			-IO	-DO		-IO	-DO
cf	P	P	trivial	P	coNP	coNP	trivial
ad	P	coNP	coNP	P	coNP	coNP	coNP
st	P	coNP	coNP	P	coNP	coNP	coNP
co, gr	P	coNP	coNP	coNP	coNP	coNP	coNP
pr	coNP	coNP	coNP	coNP	Π_2^p	Π_2^p	Π_2^p

Credulous acceptance

σ	Acc	RA	RA	RA	RNA	RNA	RNA
			-IO	-DO		-IO	-DO
ad	NP	Π_2^p	Π_2^p	Π_2^p	coNP	coNP	coNP
st	NP	Π_2^p	Π_2^p	Π_2^p	coNP	coNP	coNP
co	NP	Π_2^p	Π_2^p	Π_2^p	coNP	coNP	coNP
gr	P	coNP	coNP	coNP	coNP	coNP	coNP
pr	NP	Π_2^p	Π_2^p	Π_2^p	coNP	coNP	coNP

Skeptical acceptance

σ	Acc	RA	RA	RA	RNA	RNA	RNA
			-IO	-DO		-IO	-DO
st	coNP	coNP	coNP	coNP	Π_2^p	Π_2^p	Π_2^p
co	P	coNP	coNP	coNP	coNP	coNP	coNP
gr	P	coNP	coNP	coNP	coNP	coNP	coNP
pr	Π_p^2	Π_2^p	Π_2^p	Π_2^p	Π_3^p	Π_3^p	Π_3^p

Table 1: Complexity of the verification and acceptance problems over AAFs/PAFs/AVAFs and of the k-robustness problems over AVAFs. When a class beyond P is reported, it means that the problem is complete for the class

4 The Complexity of the k-Robustness Problems

We here present the main technical contribution of this paper, that is the characterization of the computational complexity of the k-robustness problems. The results are summarized in Table 1, which also reports the complexity of the classical verification and acceptance problems over AAFs, PAFs, and AVAFs (obviously, the semantics cf is not considered for the acceptance and the related robustness problems, and ad is not considered for the skeptical acceptance and the related robustness problems, since these cases are trivial).

We start with a preliminary result: verifying if a specific attack can be deactivated by a certain number of preference deletions is in P. This result is used for proving the tractability of the polynomial-time cases of the k-robustness problems.

Lemma 1 *The problem* singleAttackActivation-do$(VF, (a, b), k)$, *where $VF = \langle A, D, V, \mathrm{val}, \pi\rangle$ is an AVAF, (a, b) an attack in D, $k \geq 0$ an integer, of checking if there is a consistent del-only preference update U with $|U| \leq k$ such that the attack (a, b) is active in $U(VF)$ is in P.*

Lemma 1 is implied by the fact that the deactivation of an attack (a, b) via del-only updates is equivalent to making, in the preference graph, the node $val(b)$ unreachable from $val(a)$ via edge removals. In turn, this check is a particular instance of the well-known directedCut problem: "*Given a* *directed graph G and a set of pairs of nodes $\mathcal{T}$, check if there are no more than k edges whose removal from G makes, for every pair (n_1, n_2) in $\mathcal{T}$, n_2 unreachable from n_1*". Now, although DIRECTEDCUT is NP-hard in general, it is in P if $|\mathcal{T}| = 1$ (see Theorem 3.1 in [Bang-Jensen and Yeo, 2014]). This also explains why Lemma 1 cannot be extended to the deactivation of multiple attacks.

Theorem 1 *The computational complexity of* RE, RNE, RA *and* RNA *are reported in Table 1.*

Proof. Memberships: cases beyond P. The memberships in classes beyond P can be all proved with a similar scheme. As for RE, RNE, credulous RA and skeptical RNA, the scheme for the complement of these problems is: guess a preference update U with $|U| \leq k$ and then check if S is an extension (for RE and RNE) or a is accepted (for RA and or RNA) in $U(\mathit{VF})$. An exception is the case RE under $\sigma = \texttt{pr}$: the guess must also contain a set of arguments $S' \supset S$, and the check consists in verifying whether S and S' are admissible extensions. In every case, the final complexity class is coNP$^{\mathcal{C}}$, where $\mathcal{C}$ is the complexity class of the verification or acceptance problem over AVAFs solved in the checking phase.

As for credulous RNA and skeptical RA, the scheme for the complement of these problems is: guess a preference update U with $|U| \leq k$ and a set of arguments S such that $a \notin S$; then, check if S is an extension of $U(\mathit{VF})$ (under $\sigma = \texttt{pr}$ the verification is done under the admissible semantics). Hence, the final complexity class is coNP$^{\mathcal{C}}$, where $\mathcal{C}$ is the class characterizing the verification problem over AVAFs solved in the check.

Memberships: trivial cases. RNE-DO (resp., RE-IO) is trivial since del-only (resp., ins-only) updates cannot deactivate (resp., activate) attacks, thus they cannot make conflict-free (resp., conflicting) a set of arguments that is conflicting (resp., conflict-free).

Memberships: polynomial-time cases. RE-DO$^{\sigma}(\mathit{VF}, S, k)$ is in P under $\sigma \in \{\texttt{cf}, \texttt{ad}, \texttt{st}\}$, where $\mathit{VF} = \langle A, D, V, \mathit{val}, \pi \rangle$.

Under $\sigma = \texttt{cf}$, let $D_S = D \cap S \times S$ be the set of (inactive) attacks between nodes in S. If $D_S = \emptyset$, the answer to RE-DO$^{\sigma}(\mathit{VF}, S, k)$ is trivially TRUE, as no attack can be created (implying conflicts inside S) by modifying the preferences. Otherwise, if $D_S \neq \emptyset$, in order to make S non conflict-free it is necessary and sufficient to make at least one attack in D_S active. Therefore, RE-DO$^{\sigma}(\mathit{VF}, S, k)$ is TRUE iff there is at least one attack α in D_s such that SINGLEATTACKACTIVATION$(\mathit{VF}, \alpha_i, k)$ is TRUE. Performing this check requires solving at most $|D_S|$ instances of SINGLEATTACKACTIVATION, so Lemma 1 implies that RE-DO is in P.

Under $\sigma = \texttt{st}$, RE-DO coincides with the case $\sigma = \texttt{cf}$, as preference deletions cannot inactivate attacks from S against nodes outside S (so del-only updates can invalidate a stable extension only by making it conflicting).

Under $\sigma = \texttt{ad}$, solving RE-DO means deciding if either *Property* 1: "*there is a del-only preference update U with $|U| \leq k$ making S non conflict-free*", or *Property* 2: "*there is a del-only preference update U with $|U| \leq k$ making S undefended against attacks from outside S*", or both. Now, checking *Property* 1 is in P, as it means solving RE-DO$^{\sigma}(\mathit{VF}, S, k)$ under $\sigma = \texttt{cf}$. We show that also checking *Property* 2 is in P, thus concluding the proof. Since S is an extension of VF, every attack $\alpha = (a, s)$ towards S with $a \notin S$ satisfies one of the two conditions:

C_1 : α is inactive and no attack from S to a is active, or
C_2 : α is either active or inactive, and at least one attack from S to a is active.

In fact, the combination "*α is active and no attack from S to a is active*" is not possible, as S is an extension in *VF*. Let D_1 and D_2 be the disjoint subsets of D consisting of the attacks α satisfying C_1 and C_2, respectively.

Applying any del-only update over *VF* keeps S defended against the attacks in D_2, since removing preferences cannot inactivate active attacks from S. So, checking *Property* 2 means checking if there is a del-only preference update U with $|U| \leq k$ such that, for at least one attack $\alpha = (a, s)$ in D_1: i. U makes α active (i.e. $(\mathit{val}(s) > \mathit{val}(a)) \notin U(\pi)^*$), and ii. U leaves S undefended against α. We now prove that the existence of a preference update U with $|U| \leq k$ satisfying i. implies the existence of a preference update U' with $|U'| \leq k$ satisfying both i. and ii.. By contradiction, if this were false, every del-only preference update U with $|U| \leq k$ activating α would also activate some attack from some $b \in S$ to a; thus, some edges removed by U from the preference graph encoding π^* to make $\mathit{val}(a)$ unreachable from $\mathit{val}(s)$ would also make $\mathit{val}(b)$ unreachable from $\mathit{val}(a)$. This means that, in the preference graph encoding π^*, there is at least a path p_{sa} from $\mathit{val}(s)$ to $\mathit{val}(a)$ containing an edge $e = (\mathit{val}(x), \mathit{val}(y))$ in common with a path p_{ab} from $\mathit{val}(a)$ to $\mathit{val}(b)$. Thus, p_{sa} and p_{ab} are of the form $p_{sa} = p_{sx} \cdot e \cdot p_{ya}$ and $p_{ab} = p_{ax} \cdot e \cdot p_{ya}$, where p_{wz} denotes a path from $\mathit{val}(w)$ to $\mathit{val}(z)$. This implies that π^* contains the path $p_{ax} \cdot e \cdot p_{ya}$, which contradicts that π is consistent.

What said so far allows us to check *Property* 2 by merely checking the existence of a del-only preference update U with $|U| \leq k$ satisfying i. In turn, this means solving SINGLEATTACKACTIVATION$(\mathit{VF}, (a, s), k)$, so Lemma 1 implies that checking *Property* 2 is in P.

Hardness: RNE-DO with $\sigma \in \{\texttt{ad}, \texttt{st}, \texttt{co}, \texttt{gr}\}$. Under $\sigma = \texttt{ad}$, we show a reduction from the NP-hard problem 2-SKEWMULTICUT (a special case of SKEWMULTICUT [Kratsch *et al.*, 2015]) to the complement of RNE-DO. An instance of SKEWMULTICUT is a tuple $(G, k, \mathcal{T})$, where $G = \langle N, E \rangle$ is a directed acyclic graph, $k \geq 0$ is an integer, and $\mathcal{T}$ a set of pairs of nodes of the following form: there are two disjoint subsets $\mathcal{V} = \{v_1, \ldots, v_m\}$ and $\mathcal{W} = \{w_1, \ldots, w_m\}$ of N (with $|\mathcal{V}| = |\mathcal{W}| = m$), such that $\mathcal{T} = \{(v_i, w_j) \in \mathcal{V} \times \mathcal{W} \mid i \leq j\}$. Without loss of generality, it is assumed that for each $(v_i, w_j) \in \mathcal{T}$, w_j is reachable from v_i in G. The answer of SKEWMULTICUT$(G, k, \mathcal{T})$ is *Yes* if there is a set of edges E' of G with $|E'| \leq k$ such that, for every $(v, w) \in \mathcal{T}$, w is not reachable from v in $G \setminus E'$.

2-SKEWMULTICUT is SKEWMULTICUT where $m = 2$, i.e. $\mathcal{T}$ has the specific form: $\mathcal{T} = \{(v_1, w_1), (v_1, w_2), (v_2, w_2)\}$.

Given an instance 2-SKEWMULTICUT$(G, k, \mathcal{T})$, where we denote the sets of nodes and arcs of G as N and E, respectively, we build the AVAF $\mathit{VF} = \langle A, D, V, \mathit{val}, \pi \rangle$, where:

– $V = N \cup \{v_b\}$, where v_b is a fresh value and $\pi = E$;

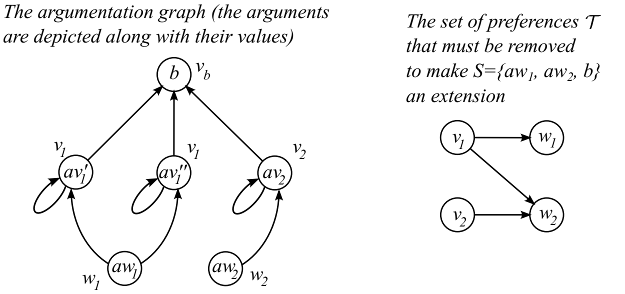
The argumentation graph (the arguments are depicted along with their values)
b
v_b
v_1
av_1'
v_1
av_1''
v_2
av_2
w_1
aw_1
aw_2
w_2
The set of preferences $\mathcal{T}$ that must be removed to make S={aw_1, aw_2, b} an extension
v_1
w_1
v_2
w_2

Figure 2: The construction used to prove the hardness of RNE-DO$^{\mathtt{ad}}$

- $A = \{av_1', av_1'', av_2, aw_1, aw_2, b\}$;
- $D = \{(av_1', av_1'), (av_1'', av_1''), (av_2, av_2), (av_1', b), (av_1'', b), (av_2, b), (aw_1, av_1'), (aw_1, av_1''), (aw_2, av_2)\}$;
- *val* is defined as follows: $\mathit{val}(aw_1) = w_1$, $\mathit{val}(aw_2) = w_2$, $\mathit{val}(av_1') = \mathit{val}(av_1'') = v_1$, $\mathit{val}(av_2) = v_2$, $\mathit{val}(b) = v_b$.

The argumentation graph and the graph of preferences underlying *VF* are depicted in Figure 2.

We set $S = \{aw_1, aw_2, b\}$. Observe that, in *VF*, the attacks from aw_1 and aw_2 are inactive (by hypothesis, every pair in $\mathcal{T}$ denotes the existence of a path in G, so every pair in $\mathcal{T}$ is also a preference in π^*). Hence, S is not an extension of *VF*.

Therefore, the problem RNE-DO$^\sigma(\mathit{VF}, S, k)$ makes sense, and we conclude the proof by showing the following equivalence: "*The answer of* 2-SKEWMULTICUT$(G, k, \mathcal{T})$ *is* Yes" $\Leftrightarrow$ "*The answer of* RNE-DO$^\sigma(\mathit{VF}, S, k)$ *is* No"

$\Rightarrow$: Let E' be a set of edges with $|E'| \leq k$ such that, for every pair $(v_i, w_j) \in \mathcal{T}$, w_j is not reachable from v_i in $G \setminus E'$. Consider the update: $U = \{\mathit{del}(x > y) | (x, y) \in E'\}$. As the preference graph of $U(\pi)$ (after removing the node v_b) coincides with $G \setminus E'$, we have that, for every pair $(v_i, w_j) \in \mathcal{T}$, $(v_i, w_j) \notin \pi^*$. Hence, the attacks from aw_1, aw_2 to av_1', av_1'', av_2 are active in $U(\mathit{VF})$, so S is an extension of $U(\mathit{VF})$, and this proves the implication, as $|U| = |E'| \leq k$.

$\Leftarrow$: Let U be a del-only update over *VF* such that $|U| \leq k$ and S is an extension of $U(\mathit{VF})$. As the value v_b is involved in no preference in *VF*, it is still involved in no preference in $U(\mathit{VF})$, so the attacks towards b are active in both *VF* and $U(\mathit{VF})$, as deleting preferences cannot inactivate attacks. Hence, the fact that S is not an extension of *VF* but is an extension of $U(\mathit{VF})$ means that the attacks from aw_1 and aw_2 are inactive in *VF* but active in $U(\mathit{VF})$. This means that, for every pair $(v_i, w_j) \in \mathcal{T}$, we have that $(v_i, w_j) \notin U(\pi)^*$. It is straightforward to see that this is the same as saying that the set of edges E' consisting of the same pairs as U is such that, for every pair $(v_i, w_j) \in \mathcal{T}$, w_j is not reachable from v_i in $G \setminus E'$, that proves the implication, as $|E'| = |U| \leq k$.

The proof used in the case $\sigma = \mathtt{ad}$ trivially extends to $\sigma \in \{\mathtt{st}, \mathtt{co}, \mathtt{gr}\}$, as, for any consistent preference update U, $S = \{aw_1, aw_2, b\}$ is an admissible extension of $U(\mathit{VF})$ iff S is an extension of $U(\mathit{VF})$ under any $\sigma \in \{\mathtt{st}, \mathtt{co}, \mathtt{gr}\}$.

Hardness: RE-DO under $\sigma \in \{\mathtt{co}, \mathtt{gr}, \mathtt{pr}\}$. We show a reduction from 2-SKEWMULTICUT. Given an instance 2-SKEWMULTICUT$(G, k, \mathcal{T})$, we consider the instance RE-DO(VF, S', k), where *VF* is the same AVAF defined in the reduction used above to prove the coNP-hardness of RNE-DO under $\sigma = \mathtt{ad}$, while $S' = \{aw_1, aw_2\}$. The same reasoning used above proves the following equivalence, that proves the correctness of the reduction: "*The answer of* 2-SKEWMULTICUT$(G, k, \mathcal{T})$ *is* Yes" $\Leftrightarrow$ "*The answer of* RE-DO(VF, S', k) *is* No". The difference, here, is that S' is complete in *VF*, and activating the attacks from aw_1, aw_2 against av_1', av_1'', av_2 makes it no longer complete.

Hardness: other cases. *(Sketch)* As for RNE-IO under $\sigma \in \{\mathtt{cf}, \mathtt{ad}, \mathtt{st}, \mathtt{co}, \mathtt{gr}, \mathtt{pr}\}$, it is easy to see that the complement of RNE-IO is more general than checking the existence of an update consisting of no more than k insertions that inactivates a given set of attacks in an AVAF. In turn, this problem can be shown to be able to encode any instance of the well-known NP-hard problem SET-COVER.

As for RNE (only under $\sigma = \mathtt{pr}$), RNA-IO and RNA-DO, reductions from verification or acceptance problems in *attack incomplete AAFs* (aiAAFs) can be defined (the relationship with aiAAFs is also discussed in Section 5).

As for RA with $\sigma \in \{\mathtt{ad}, \mathtt{st}, \mathtt{co}, \mathtt{gr}, \mathtt{pr}\}$ and $X = \mathit{Cr}$, reductions from the satisfiability problem over (quantified) propositional formulas can be defined.

As for RA with $\sigma \in \{\mathtt{st}, \mathtt{pr}\}$ and $X = \mathit{Sk}$, reductions from the skeptical acceptance problem over AAFs can be defined. The hardness for RA with $\sigma \in \{\mathtt{co}, \mathtt{gr}\}$ and $X = \mathit{Sk}$ follows from the case of RA with $\sigma = \mathtt{gr}$ and $X = \mathit{Cr}$. □

The following proposition states that, if the number of values is constant w.r.t. the size of the AVAF, several k-robustness problems become solvable in polynomial time, i.e. they are *fixed parameter tractable* w.r.t. the number of values.

Proposition 1 RE, RNE *under* $\sigma \in \{\mathtt{cf}, \mathtt{ad}, \mathtt{st}, \mathtt{co}, \mathtt{gr}\}$*, as well as* RA *and* RNA *under* $\sigma = \mathtt{gr}$ *and* $X \in \{\mathit{Cr}, \mathit{Sk}\}$ *or* $\sigma = \mathtt{co}$ *and* $X = \mathit{Sk}$ *are fixed parameter tractable w.r.t the size of the set of values.*

Proof. Let $\mathit{VF} = \langle A, D, V, \mathit{val}, \pi \rangle$ be an AVAF. The k-robustness problems can be solved by generating all the preference updates U and checking the verification or the acceptance over $U(\mathit{VF})$. Since: i. the number of preference updates is bounded by $2^{|V| \times |V|}$, ii. the cost of applying an update is polynomial in the size of the update and the AVAF, iii. the verification or acceptance test over the implied AAF are in P for the cases in the statement, the overall cost is $O\big(g(|V|) \cdot f(|A| + |D|)\big)$, where f is a polynomial function. □

5 Discussion of the Results and Related Work

This study is related to the works investigating how the outcome of the reasoning over AAFs changes when attacks are inserted/removed. In this research context, the following two topics are particularly related:

- *attack-incomplete AAFs (aiAAFs)* [Baumeister *et al.*, 2018], i.e. AAFs where some attacks are uncertain, so the reasoning takes into account all the alternative "completions", i.e. the alternative AAFs containing all the arguments and the certain attacks, along with a subset of the set of uncertain attacks;
- *the strict (resp., non-strict) argument-fixed extension enforcement problem over AAFs*, that asks for minimizing the number of *local updates* (i.e. insertions/deletions

into/from the set of attacks) needed to make a set an extension (resp., a subset of an extension) [Baumann *et al.*, 2021; Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2017].

As for the relationship with aiAAFs, the problems RNE and RNA resemble the complements of the problems IVER and IACC of deciding if a set is an extension and an argument is accepted in a completion of an aiAAF, studied in [Fazzinga *et al.*, 2020; Baumeister *et al.*, 2021]. However, compared with IVER and IACC, RNE and RNA exhibit two further sources of complexity: the necessity for counting (as in RNE and RNA a condition is specified on the number of preference updates) and the fact that the preferences between values encode correlations between activations/deactivations of distinct attacks, thus making these problems more similar to the variants of IVER and IACC addressed in [Fazzinga *et al.*, 2021a], where correlations can be expressed over the uncertain attacks. Indeed, the correlations expressible via preferences between values do not coincide with any of the single forms of dependencies in [Fazzinga *et al.*, 2021a]. From a computational complexity standpoint, it is interesting to observe that, for the semantics under which IVER and IACC are hard for some class $\mathcal{C}$ beyond P, the problems RNE and RNA are complete (both in the ins-only and del-only variants) for the complement of $\mathcal{C}$; on the other hand, when IVER is in P (i.e. under $\sigma \in \{\texttt{cf}, \texttt{ad}, \texttt{st}, \texttt{co}, \texttt{gr}\}$), RNE is coNP-complete (except for the trivial case of del-only updates under $\sigma = \texttt{cf}$).

As for the relationship with the enforcement problems above, RNE and RNA are close to (the complements of) the strict and the non-strict enforcement problems. However, analogously to what was said for aiAAFs, the enforcement problem imposes no constraint on which attacks can be added/deleted. On the contrary, when assessing the robustness, it is not possible to consider inconsistent updates, so, for instance, it is not possible to simultaneously deactivate two attacks of the form (a, b), (b, a) to make $\{a, b\}$ conflict-free (the removal of (a, b), (b, a) is instead allowed when enforcing the conflict-freeness of $\{a, b\}$). Comparing our results on RNE and RNA with those on the enforcement in [Wallner *et al.*, 2017], the main difference is in the cases where the enforcement is decidable in P (i.e. under $\sigma \in \{\texttt{cf}, \texttt{ad}, \texttt{st}\}$), as in these cases RNE is harder (coNP-complete), while in the other cases the complexity classes are complementary.

Observe also that our fine-grained study of the computational complexity, where we separately consider the cases of ins-only and del-only updates, has no counterpart in the literature of aiAAFs and argument-fixed enforcements, where it was not investigated what happens when attacks can be only added or only removed (for instance, the reductions in [Wallner *et al.*, 2017] proving the NP-hardness of the strict enforcement under $\sigma \in \{\texttt{co}, \texttt{gr}, \texttt{pr}\}$ simultaneously exploit the insertions and deletions of attacks).

As for the other two k-robustness problems studied in this paper, RE and RA, they are in some sense analogous to enforcement of general acceptability constraints [Doutre and Mailly, 2018] (however, the setting where this enforcement was studied disregards values and preferences).

This work is also closely related to the generalizations of AAFs where preferences are used for filtering the extensions [Alfano *et al.*, 2023; Amgoud and Vesic, 2011] (similarly to what done in [Alfano *et al.*, 2024b; Alfano *et al.*, 2024c] by means of constraints) or for "revising" the attacks (as done in PAFs [Amgoud and Cayrol, 2002], that underlie the semantics of AVAFs, or in [Kaci and Labreuche, 2014], where preferences are valued and are used to obtain varied-strength attacks) or for both these aspects [Amgoud and Vesic, 2014]. Preferences have been employed also in other forms of argument systems: they are specified over arguments to extend *Claim-augmented AAFs* in [Bernreiter *et al.*, 2024], over structured arguments in *ASPIC+* [S. Mogdil, 2013], and over assumptions to extend *Assumption Based Argumentation* [Cyras and Toni, 2016]. In [Modgil, 2009], a framework integrating metalevel argumentation that encodes the reasoning over preferences and generalizes PAFs and AVAFs has been introduced. In [Kaci *et al.*, 2018; Amgoud and Vesic, 2014], three translations from PAFs to AAFs alternative to the one used in this paper (that was defined in [Amgoud and Cayrol, 2002]) have been introduced. The notion of robustness and the definitions of the related problems studied in this paper are orthogonal to the specific translation used. As for the complexity results, it is easy to see that the memberships in classes beyond P in Table 1 hold for the other translations; the same is true, up to minor changes, for most of the reductions proving the hardness results. However, formally proving the correctness of the modified reductions and extending the tractability results of Table 1 to the other translations is a matter of future work.

As robustness deals with the uncertainty of the audience profile, our work is related with the literature on uncertainty in abstract argumentation such as *incomplete AAFs* [Fazzinga *et al.*, 2020] (generalizing aiAAFs, as they also consider uncertain arguments), *Control Argumentation Frameworks* [Dimopoulos *et al.*, 2018], and *probabilistic AAFs* (following the *constellations* [Li *et al.*, 2011; Fazzinga *et al.*, 2018; Fazzinga *et al.*, 2019; Polberg and Hunter, 2018] or the *epistemic* [Hunter and Thimm, 2014; Thimm, 2012] approach), where the uncertainty is expressed probabilistically.

6 Conclusions and Future Work

In the context of AVAF, we have introduced a framework for reasoning on the robustness of the status of sets of arguments and of single arguments with respect to changes of the audience profile. This profile is a set of preferences over the social valued promoted by the arguments, and its modifications cause the activation/deactivation of the attacks. We have characterized the computational complexity of the decision problems at the core of the reasoning paradigm. Future work will be devoted to addressing the following extensions of the framework: 1) modeling arguments promoting multiple values, as done in [Kaci and van der Torre, 2008]; 2) simultaneously considering multiple audiences, and reasoning on the robustness of the status of the arguments for at least one or all the audiences; 3) introducing a weighted model for the preferences, so that it can be taken into account that a preference insertion/deletion/inversion is more or less likely than another; 4) defining new gradual semantics for AVAFs, where the acceptability degrees of the arguments are defined in terms of their robustness degrees.

Acknowledgments

We acknowledge financial support from PNRR MUR project PE0000013-FAIR and project Tech4You ECS0000009, and MUR project PRIN 2022 EPICA (CUP H53D23003660006).

References

[Alfano *et al.*, 2023] Gianvincenzo Alfano, Sergio Greco, Francesco Parisi, and Irina Trubitsyna. Preferences and constraints in abstract argumentation. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 3095–3103, 2023.

[Alfano *et al.*, 2024a] Gianvincenzo Alfano, Andrea Cohen, Sebastian Gottifredi, Sergio Greco, Francesco Parisi, and Guillermo Ricardo Simari. Credulous acceptance in high-order argumentation frameworks with necessities: An incremental approach. *Artif. Intell.*, 333:104159, 2024.

[Alfano *et al.*, 2024b] Gianvincenzo Alfano, Sergio Greco, Domenico Mandaglio, Francesco Parisi, and Irina Trubitsyna. Abstract argumentation frameworks with strong and weak constraints. *Artif. Intell.*, 336:104205, 2024.

[Alfano *et al.*, 2024c] Gianvincenzo Alfano, Sergio Greco, Francesco Parisi, and Irina Trubitsyna. Complexity of credulous and skeptical acceptance in epistemic argumentation framework. In *Proc. Int. Conf. Artificial Intelligence (AAAI)*, pages 10423–10432, 2024.

[Amgoud and Cayrol, 2002] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *A. Math. Artif. Intell.*, 34(1-3):197–215, 2002.

[Amgoud and Vesic, 2011] Leila Amgoud and Srdjan Vesic. A new approach for preference-based argumentation frameworks. *A. Math. Artif. Intell.*, 63(2):149–183, 2011.

[Amgoud and Vesic, 2014] Leila Amgoud and Srdjan Vesic. Rich preference-based argumentation frameworks. *Int. J. Approx. Reason.*, 55(2):585–606, 2014.

[Bang-Jensen and Yeo, 2014] Jørgen Bang-Jensen and Anders Yeo. The complexity of multicut and mixed multicut problems in (di)graphs. *Theor. Comput. Sci.*, 520:87–96, 2014.

[Baumann *et al.*, 2021] Ringo Baumann, Sylvie Doutre, Jean-Guy Mailly, and Johannes Peter Wallner. Enforcement in formal argumentation. *FLAP*, 8(6):1623–1678, 2021.

[Baumeister *et al.*, 2018] Dorothea Baumeister, Daniel Neugebauer, Jörg Rothe, and Hilmar Schadrack. Verification in incomplete argumentation frameworks. *Artif. Intell.*, 264:1–26, 2018.

[Baumeister *et al.*, 2021] Dorothea Baumeister, Matti Järvisalo, Daniel Neugebauer, Andreas Niskanen, and Jörg Rothe. Acceptance in incomplete argumentation frameworks. *Artif. Intell.*, 295:103470, 2021.

[Bench-Capon, 2003] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. Log. Comput.*, 13(3):429–448, 2003.

[Bernreiter *et al.*, 2024] Michael Bernreiter, Wolfgang Dvorák, Anna Rapberger, and Stefan Woltran. The effect of preferences in abstract argumentation under a claim-centric view. *J. Artif. Intell. Res.*, 81:203–262, 2024.

[Coste-Marquis *et al.*, 2015] Sylvie Coste-Marquis, Sébastien Konieczny, Jean-Guy Mailly, and Pierre Marquis. Extension enforcement in abstract argumentation as an optimization problem. In *Proc. Int. Conf. on Artificial Intelligence (IJCAI)*, pages 2876—-2882, 2015.

[Cyras and Toni, 2016] Kristijonas Cyras and Francesca Toni. ABA+: assumption-based argumentation with preferences. In *Proc. Int. Conf. Principles of Knowledge Representation and Reasoning (KR)*, pages 553–556, 2016.

[Dimopoulos *et al.*, 2018] Yannis Dimopoulos, Jean-Guy Mailly, and Pavlos Moraitis. Control argumentation frameworks. In *Proc. Conf. on Artificial Intelligence (AAAI), New Orleans, USA*, pages 4678–4685, 2018.

[Doutre and Mailly, 2018] Sylvie Doutre and Jean-Guy Mailly. Constraints and changes: A survey of abstract argumentation dynamics. *Argument Comput.*, 9(3):223–248, 2018.

[Dung, 1995] Phan Minh Dung. 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.

[Fazzinga *et al.*, 2018] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Probabilistic bipolar abstract argumentation frameworks: complexity results. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1803–1809, 2018.

[Fazzinga *et al.*, 2019] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Complexity of fundamental problems in probabilistic abstract argumentation: Beyond independence. *Artif. Intell.*, 268:1–29, 2019.

[Fazzinga *et al.*, 2020] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Revisiting the notion of extension over incomplete abstract argumentation frameworks. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1712–1718, 2020.

[Fazzinga *et al.*, 2021a] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Reasoning over argument-incomplete aafs in the presence of correlations. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), Virtual Event / Montreal, Canada*, pages 189–195, 2021.

[Fazzinga *et al.*, 2021b] Bettina Fazzinga, Sergio Flesca, and Filippo Furfaro. Reasoning over attack-incomplete aafs in the presence of correlations. In *Proc. Int. Conf. on Principles of Knowledge Representation and Reasoning (KR), Online event*, pages 301–311, 2021.

[Fazzinga *et al.*, 2022a] Bettina Fazzinga, Sergio Flesca, Filippo Furfaro, and Luigi Pontieri. Process mining meets argumentation: Explainable interpretations of low-level event logs via abstract argumentation. *Inf. Syst.*, 107:101987, 2022.

[Fazzinga *et al.*, 2022b] Bettina Fazzinga, Andrea Galassi, and Paolo Torroni. A privacy-preserving dialogue system based on argumentation. *Intell. Syst. Appl.*, 16:200113, 2022.

[Hunter and Thimm, 2014] Anthony Hunter and Matthias Thimm. Probabilistic argumentation with epistemic extensions. In *Proc. of DARe@ECAI*, 2014.

[Kaci and Labreuche, 2014] Souhila Kaci and Christophe Labreuche. Valued preference-based instantiation of argumentation frameworks with varied strength defeats. *Int. J. Approx. Reason.*, 55(9):2004–2027, 2014.

[Kaci and van der Torre, 2008] Souhila Kaci and Leendert W. N. van der Torre. Preference-based argumentation: Arguments supporting multiple values. *Int. J. Approx. Reason.*, 48(3):730–751, 2008.

[Kaci *et al.*, 2018] Souhila Kaci, Leendert W. N. van der Torre, and Serena Villata. Preference in abstract argumentation. In *Proc. Computational Models of Argument (COMMA)*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, pages 405–412, 2018.

[Kratsch *et al.*, 2015] Stefan Kratsch, Marcin Pilipczuk, Michał Pilipczuk, and Magnus Wahlström. Fixed-parameter tractability of multicut in directed acyclic graphs. *SIAM J. Discret. Math.*, 29(1):122–144, January 2015.

[Li *et al.*, 2011] Hengfei Li, Nir Oren, and Timothy J. Norman. Probabilistic argumentation frameworks. In *Proc.of TAFA*, 2011.

[Modgil, 2009] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173(9-10):901–934, 2009.

[Polberg and Hunter, 2018] Sylwia Polberg and Antony Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *Int. J. Approx. Reasoning*, 93:487–543, 2018.

[Rapberger and Toni, 2024] Anna Rapberger and Francesca Toni. On the robustness of argumentative explanations. In *Proc. Computational Models of Argument (COMMA)*, volume 388 of *Frontiers in Artificial Intelligence and Applications*, pages 217–228, 2024.

[Rienstra *et al.*, 2020] Tjitze Rienstra, Chiaki Sakama, Leendert van der Torre, and Beishui Liao. A principle-based robustness analysis of admissibility-based argumentation semantics. *Argument Comput.*, 11(3):305–339, 2020.

[S. Mogdil, 2013] H. Prakken S. Mogdil. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.

[Thimm, 2012] Matthias Thimm. A probabilistic semantics for abstract argumentation. In *Proc. of ECAI*, pages 750–755, 2012.

[Wallner *et al.*, 2017] Johannes Peter Wallner, Andreas Niskanen, and Matti Järvisalo. Complexity results and algorithms for extension enforcement in abstract argumentation. *J. Artif. Intell. Res.*, 60:1–40, 2017.