Computational Strategies for Trust-aware Abstract Argumentation Frameworks*

Bettina Fazzinga, Sergio Flesca, Filippo Furfaro

ICAR - CNR, email: fazzinga@icar.cnr.it,
DIMES - University of Calabria, email: {flesca, furfaro}@dimes.unical.it

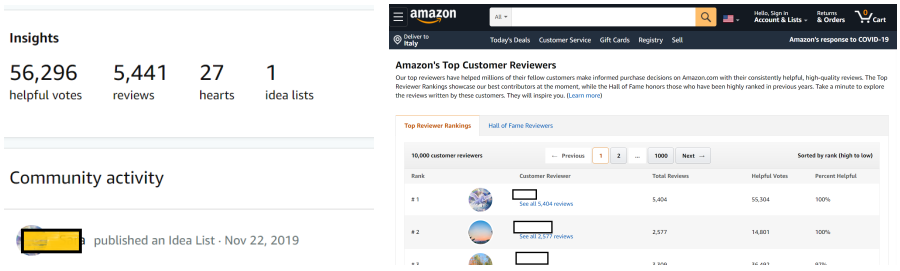
Abstract. The Trust-aware Abstract Argumentation Frameworks (T-AAFs) have been proposed in [18] as a variant of the well-known abstract argumentation frameworks where the trustworthiness of the agents participating the dispute is taken into account. In particular, T-AAFs consist in AAFs where arguments are associated with weights derived from the trust degrees of the agents proposing them. [18] studies the problem MIN-TVER (resp., MIN-TACC) of computing the minimum trust degree τ^* such that, if the arguments said only by agents whose trust degree is not greater than τ^* are discarded, a given set of arguments S (resp., argument a), that is not necessarily an extension (resp., (credulously) accepted) over the original argumentation framework, becomes an extension (resp., (credulously) accepted). We extend the proposal in [18] by devising suitable methods for solving the problems MIN-TVER and MIN-TACC. Specifically, we provide a translation for the intractable cases of MIN-TVER and MIN-TACC into instances of Integer Linear Programming (ILP), so that they can be solved by resorting to standard ILP solvers.

1 Introduction

Abstract Argumentation Frameworks (AAFs) [12] are a paradigm for reasoning on disputes between agents founded on directed graphs, whose nodes are the *arguments* proposed by the agents participating the dispute, and whose edges represent *attack* relationships. Specifically, an attack from an argument a to an argument b represents the fact that a undercuts/rebuts/undermines b. AAFs are used to reason on sets of arguments and/or single arguments to decide whether they are "robust". Herein, in order to decide on the "robustness" of a set of arguments, different semantics have been introduced, such as *admissible*, *preferred*, etc. For instance, a set S is an *admissible* extension if it is "*conflict-free*" (i.e., there is no attack between arguments in S), and every argument attacking arguments in S is counterattacked by an argument in S.

* The research reported in this work was partially supported by the *EU H*2020 *ICT*48 project Humane AI Net under contract #952026. The support is gratefully acknowledged.

In order to make AAFs suitable for modeling disputes in scenarios with different characteristics, several variants have been proposed. In particular, *Weighted AAFs* are a variant of AAFs where the arguments and/or the attacks can be associated with weights. The paper in [18] introduces the *Trust-aware AAFs* (T-AAFs), a form of weighted AAFs where the weights are assigned to arguments and are representative of the trustworthiness of the agents who propose the arguments. A natural application of T-AAFs is the e-commerce scenario, where customers share their reviews about products and get a score based on the quality of their reviews. As an example, consider the Amazon web site, where every product gets reviews, and every reviewer is classified on the basis of the usefulness of her/his reviews. Figure 1(a) shows one page of the Amazon web site, containing the information about one of the reviewers. Each Amazon reviewer has at least three scores: the position in the general ranking of the reviewers, the number of helpful votes and the number of reviews. Moreover, it is possible to devise other statistics about the quality of a reviewer such as the percentage of her/his reviews that are considered useful by other customers (as shown in the "Amazon Top reviewers" page shown in Figure 1(b)). Building a T-AAF starting from the reviews of a certain Amazon product could, then, result in using the content of the reviews as arguments, the contradictions among the reviews' content as attacks and any suitable trustworthiness measure derived from the position in the general ranking or the number of helpful votes (possible weighted with the number of reviews) as trust degree of the reviewers involved in the product review.



(a) (b)

Fig. 1: One of the Amazon's reviewers (a) and the Amazon's top reviewers (b)

The following example is inspired by the above scenario.

Example 1. Ann, Mary, Carl and John are reviewing a notebook. Their reviews contain the following six arguments:
a='Since it contains up-to-date components, it is expensive'
b='Nowadays, it is easy to find cheap up-to-date components. Therefore, that aspect does not imply the price.'
c='Since its brand is not high quality, it does not contain up-to-date components'

d='Since its battery is lightweight, it is lightweight overall'
e='It is heavy'
f='The battery is very heavy'.
Figure 2 shows the corresponding argumentation graph, properly augmented to highlight who-claims-what (for instance, a and d are claimed by Mary, and e is claimed by both Ann and Carl). The numbers in brackets represent the trustworthiness scores, on a scale of 1 to 10, assigned to the agents on the basis of their past reviews.

As a matter of fact, reasoning on reviews is a hot topic attracting the interest of the research community, owing to the popularity of ecommerce sites. In this context, reasoning on extensions is useful, since the fact that a set of arguments is an extension means that it provides a reasonable summary of the main features and critical aspects of the reviewed object. Analogously, reasoning on the acceptance of an argument helps understand if it can be reasonably considered representative of the object. Now, in the T-AAF F of Example 1, argument a does not belong to any extension. However, a is proposed by Mary, who has a high trust degree. Thus, the analyst can benefit from knowing that, although a is not accepted, it becomes accepted in the AAF F^τ (with $\tau = 2$) obtained from F by discarding what said only by agents whose trust degree is $\leq \tau$. This means that the analyst can choose now to consider a a robust argument, given that F^τ does not contain what said by agents with "low" trust degrees (we recall that we are in a scale from 1 to 10). Analogously, even if $S = \{a, f\}$ is not an (admissible) extension in F, it can be somehow considered a reasonable summary of the reviews, since it is an extension over the same F^τ. In general, denoting as "τ-extension" (resp., "τ-accepted") a set (resp., an argument) that is an extension (resp., accepted) over F^τ, the following two problems over a T-AAF F are of interest to the analyst:

- MIN-TVER$^\sigma(F, S)$: What is the minimum trust degree τ such that the set S is a τ-extension over F under σ?
- MIN-TACC$^\sigma(F, a)$: What is the minimum trust degree τ such that the argument a is τ-accepted over F under σ?

The complexity of MIN-TVER$^\sigma(F, S)$ and MIN-TACC$^\sigma(F, a)$ has been characterized in [18]: in particular, it has been proved that computing MIN-TVER$^\sigma(F, S)$ is intractable for the preferred semantics and that MIN-TACC$^\sigma(F, a)$ is intractable

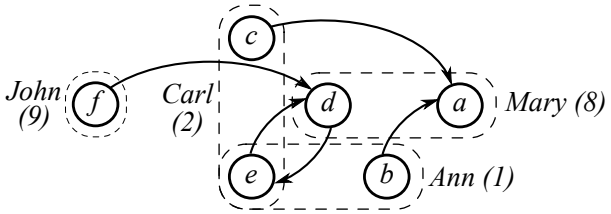


Fig. 2: A T-AAF F, where the agents are assigned a trust degree

for the admissible, complete, stable and preferred semantics. In this paper, we provide methods for computing MIN-TVER$^\sigma(F,S)$ and MIN-TACC$^\sigma(F,a)$ for the semantics for which they resulted to be intractable (see Table 1). Our strategy is based on the translation of MIN-TVER$^\sigma(F,S)$ and MIN-TACC$^\sigma(F,a)$ into Integer Linear Programming (ILP), so that they can be efficiently computed by exploiting the many heuristics already implemented in the commercial solvers.

σ	VER$^\sigma$ and TVER$^\sigma$	ACC$^\sigma$ and TACC$^\sigma$	MIN-TVER$^\sigma$	MIN-TACC$^\sigma$
ad,*st*,*co*	*P*	*NP-c*	*FP*	$FP^{NP[\log n]}$-*c*
gr	*P*	*P*	*FP*	*FP*
pr	*coNP-c*	*NP-c*	$FP^{NP[\log n]}$-*c*	$FP^{NP[\log n]}$-*c*

Table 1: Summary of the computational complexities

2 Preliminaries

Abstract Argumentation Framework [12]. An *Abstract Argumentation Framework* (*AAF*) F is a pair $\langle A, D\rangle$, where A is a finite and non-empty set, whose elements are called *arguments*, and $D \subseteq A \times A$ is a binary relation over A, whose elements are called *attacks*. The graph having A and D as set of nodes and edges, respectively, is called *argumentation graph* of F. Given $a, b \in A$, we say that a *attacks* b iff $(a, b) \in D$. A set $S \subseteq A$ *attacks* an argument $b \in A$ iff there is $a \in S$ that *attacks* b. An argument a *attacks* S iff $\exists b \in S$ attacked by a.

A set $S \subseteq A$ of arguments is said to be *conflict-free* if there are no $a, b \in S$ such that a *attacks* b. An argument a is said to be *acceptable w.r.t.* $S \subseteq A$ iff $\forall b \in A$ such that b *attacks* a, there is $c \in S$ such that c *attacks* b.

Extension. An *extension* is a set of arguments that is considered "reasonable" according to some semantics. In particular, we consider the following semantics from the literature:

- *admissible* (*ad*): S is an *admissible extension* iff S is conflict-free and its arguments are acceptable w.r.t. S;
- *stable* (*st*): S is a *stable extension* iff S is conflict-free and S attacks each argument in $A \setminus S$;
- *complete* (*co*): S is a *complete extension* iff S is admissible and every argument acceptable w.r.t. S is in S;
- *grounded* (*gr*): S is a *grounded extension* iff S is a minimal (w.r.t. $\subseteq$) complete set of arguments;
- *preferred* (*pr*): S is a *preferred extension* iff S is a maximal (w.r.t. $\subseteq$) complete set of arguments.

Accepted arguments. An argument a is (credulously) *accepted* under a semantics σ iff a belongs to some σ extension of F. In some sense, checking the acceptability of an argument is a way of deciding whether a represents a robust point of view in the discussion modeled by F.

Classical problems: VER ***and*** ACC. Given an AAF F, a semantics σ, a set of arguments S and an argument a, the fundamental problems of verifying whether S is a σ extension and whether a is (credulously) accepted (under σ) will be denoted as $\text{VER}^\sigma(F, S)$ and $\text{ACC}^\sigma(F, a)$, respectively. The complexity of these problems, widely studied in the literature [15, 10, 13, 8], is reported in Table 1.

3 Trust-aware AAFs

We here recall the *Trust-aware AAFs* (proposed in [18]), a form of weighted AAFs where weights are associated with the arguments and represent the trust degree of the agents proposing them. We start introducing an agent trust function assigning a trust degree to each agent, from which an argument trust function assigning a trust degree to each argument is derived. Next, we formally recall the definitions of the Trust-aware Abstract Argumentation Framework and of the τ-restrictions, that are T-AAFs derived from the original T-AAF by retaining only those arguments whose trust degree is greater than a certain threshold τ. Finally, we recall how the concepts of extensions and acceptance are adapted to the τ-restrictions, that is by reporting the definitions of τ-extensions and τ-acceptance, and of the problems for which we provide the computational strategies.

Let $F = \langle A, D \rangle$ be an AAF and U the set of agents proposing the arguments in A. Function $\omega : U \rightarrow 2^A$ returns, for each agent u, the set of arguments proposed by u. We assume that every argument is proposed by at least one agent, and the same argument can be proposed by several agents. The set of agents proposing an argument a is denoted as $\omega^{-1}(a)$.

We assume the presence of an *agent trust function* τ^U assigning to each agent $u \in U$ a trust degree $\tau^U(u)$, i.e., a positive integer providing a measure of how trustworthy u is considered. Regarding the trustworthiness of an argument a, it seems natural to derive the trust degree of a from the trust degrees of the agents who propose a. In this regard, the trustworthiness of arguments is modeled with the *argument trust function* T^{U,ω,τ^U} (or, more simply, T) assigning to each argument a the positive integer equal to the maximum trust degree of the agents that propose, i.e., $T(a) = \max_{u \in \omega^{-1}(a)} \tau^U(u)$.

For the sake of simplicity, and without loss of generality, from now on we will only implicitly consider the set of users U and the functions ω and τ^U, and we will explicitly consider only the argument trust function T implied by them.

We now recall the definition of the Trust-aware Abstract Argumentation Frameworks.

Definition 1 (T-AAF). *Given an abstract argumentation framework $\langle A, D\rangle$ and an argument trust function T over A, the triple $F = \langle A, D, T\rangle$ is called* Trust-aware Abstract Argumentation Framework *(T-AAF).*

We denote as $\mathcal{T}(F)$ the set of distinct trust degrees of F's arguments augmented with 0.

Example 2. (Continuing Example 1 - Fig. 2) From the users' trust degrees, we have $T(e) = \max(\tau^U(Ann), \tau^U(Carl)) = 2$, $T(a) = T(d) = 8$, $T(c) = 2$, $T(b) = 1$, $T(f) = 9$. Moreover, we have: $\mathcal{T}(F) = \{0, 1, 2, 8, 9\}$.

We now recall the definition of the concept F^τ of *τ-restrictions*, that is the T-AAF consisting of all and only the arguments of the original T-AAF F with trust greater than τ and of all and only the attacks in F between these arguments. Let $F = \langle A, D, T\rangle$ be a T-AAF, τ a trust value, and σ a semantics. The **τ-restriction** of F is the T-AAF $F^\tau = \langle A', D', T'\rangle$ where $A' = \{a |\, a \in A \wedge T(a) > \tau\}$, $D' = D \cap (A' \times A')$, and T' is the restriction of T over A'. The T-AAF F^τ will be also called the "*τ-restriction of F*". Basically, considering the τ-restriction of F means considering τ as a threshold, and then taking into account only what said by the agents whose trust degree is greater than τ, while discarding what said *only* by agents whose trust degree is $\leq \tau$. Observe that $F^\tau = F$ when $\tau = 0$, since the trust function assigns only positive values.

We now recall how the classical notions of *extension* and *accepted argument* (reviewed in Section 2) are adapted to the case of T-AAFs. Given a T-AAF F and a trust degree τ, a **τ-extension** of F" (shorthand for "*trusted extension with trust level τ*") under the semantics σ is any set of arguments that is an extension of F^τ under σ. Basically, a τ-extension for F is a set of arguments that meets the conditions of the semantics σ in the T-AAF obtained from the original one by discarding the arguments proposed by agents whose trust degree is $\leq \tau$. In turn, an argument a of F is said to be ***τ-accepted*** (shorthand for "*trustingly accepted with trust level τ*") under σ if a belongs to at least one τ-extension under σ. The rationale of τ-acceptance is analogous to τ-extension: An argument a may not be accepted in the original T-AAF, but it can still be τ-accepted for some τ, meaning that a turns out to be a "robust" argument when discarding what said by users not sufficiently trustworthy (w.r.t. the threshold τ). The reason is that the removal of arguments (and the consequent removal of the attacks involving the removed arguments) can change the number of extensions and their composition.

Example 3. (Continuing examples 1, 2) Under $\sigma = ad$, $\{c, f\}$ is a τ-extension even with $\tau = 0$, while $\{a, f\}$ is a τ-extension for $\tau = 2$ but not for lower degrees in $\mathcal{T}(F)$. Under all the considered semantics, there is no $\tau \in \mathcal{T}(F)$ such that d is τ-accepted, while a is τ-accepted for $\tau = 2$, but not for any lower $\tau \in \mathcal{T}(F)$.

Now we recall the definitions of the fundamental problems MIN-TVER$^\sigma(F, S)$ and MIN-TACC$^\sigma(F, a)$.

Definition 2 (min-Tver$^\sigma(F,S)$). MIN-TVER$^\sigma(F,S)$*: Given a T-AAF F, a semantics σ, and a set S of arguments of F, what is the minimum trust degree τ in $\mathcal{T}(F)$ (if exists) such that S is a τ-extension of F under σ?*

Definition 3 (min-Tacc$^\sigma(F,a)$). MIN-TACC$^\sigma(F,a)$*: Given a T-AAF F, a semantics σ, and an argument a of F, what is the minimum trust degree τ in $\mathcal{T}(F)$ (if exists) such that a is τ-accepted under σ?*

The choice of minimizing the value of τ required to make S a τ-extension and a τ-accepted goes in the direction of discarding as few agents as possible from the dispute. This way, what the the agents said is tried to be preserved as much as possible, and agents with low trust degrees are discarded at first, coherently with the assumption that low values of trust degrees correspond to less reliable agents. In fact, if the output τ^* of MIN-TVER and MIN-TACC is "low", it means that considering S as an extension and a as accepted is quite reasonable: indeed, only users with low trust degree must be discarded to make S extension and a accepted. The case that τ^* is "high", instead, represents a clue of a possible risky situation: considering S as an extension and a as accepted requires to discard some/many trustworthy agents, thus it could be the case of questioning the robustness of S and a.

Example 4. From the discussion in Example 3 regarding the set $\{c, f\}$ and the argument a, it follows that MIN-TVER$^{ad}(F, \{c, f\}) = 0$ and MIN-TACC$^{ad}(F, a) = 2$.

MIN-TVER and MIN-TACC are the natural optimization counterparts of the following decision problems over a given T-AAF F and under a semantics σ:

- TVER$^\sigma(F, S, \tau^*)$: Is S a τ-extension of F under σ for some $\tau \leq \tau^*$?
- TACC$^\sigma(F, a, \tau^*)$: Is a τ-accepted for F under σ for some $\tau \leq \tau^*$?

The complexity of these problems was studied in [18] before of the complexity of MIN-TVER and MIN-TACC, since the complexity characterization of the optimization counterparts is simplified by the knowledge of the complexity of the decisional counterparts. In the next section, we report the results.

4 Complexity Characterization

We first recall the characterization of the complexity of the decisional variants TVER$^\sigma(F, S, \tau^*)$ and TACC$^\sigma(F, a, \tau^*)$.

Theorem 1. *[18]* TVER$^\sigma(F, S, \tau^*)$ *is in FP for $\sigma \in \{ad, co, st, gr\}$ and is coNP-complete for $\sigma = pr$.*

The PTIME results for $\sigma \in \{ad, co, st, gr\}$ straightforwardly follows from the fact that TVER$^\sigma(F, S, \tau^*)$ can be decided by iteratively invoking an algorithm solving VER$^\sigma(F^\tau, S)$ (that is in P), for each $\tau \in \mathcal{T}(F)$ smaller than or equal to τ^*. Furthermore, the fact that TVER$^{pr}(F, S, \tau^*)$ is *coNP*-complete can be proved by observing that a polynomial size witness for the answer "*false*" consists of

x supersets $S_1, \ldots, S_x$ of S witnessing that S is not maximally admissible in $F^{\tau_1}, \ldots, F^{\tau_x}$, respectively, and the *coNP*-hardness straightforwardly follows from the fact that VER^{pr} is *coNP*-hard.

Similar arguments were exploited in [18] for proving the following theorem regarding TACC$^{\sigma}(F, a, \tau^*)$.

Theorem 2. *[18]* TACC$^{\sigma}(F, a, \tau^*)$ *is NP-complete for every* $\sigma \in \{ad, co, st, pr\}$ *and is in FP for* $\sigma = gr$.

As regards MIN-TVER$^{\sigma}(F, S)$ and MIN-TACC$^{\sigma}(F, a)$, from Theorems 1 and 2 it can be proved that MIN-TVER$^{\sigma}(F, S)$ is in *FP* for $\sigma \in \{ad, co, st, gr\}$ and MIN-TACC$^{\sigma}(F, a)$ is in *FP* for $\sigma = gr$. Specifically, both for MIN-TVER$^{\sigma}(F, S)$ and MIN-TACC$^{\sigma}(F, a)$ we can reason as done for TVER$^{\sigma}(F, S, \tau^*)$ and TACC$^{\sigma}(F, a, \tau^*)$, that is by trying the trust degrees in $\mathcal{T}(F)$ in ascending order. Moreover, reasoning analogously ot the case of TVER$^{\sigma}(F, S, \tau^*)$ and TACC$^{\sigma}(F, a, \tau^*)$, in [18] it was shown that MIN-TVER$^{\sigma}(F, S)$ is in $FP^{NP[\log n]}$ for $\sigma = pr$ and that MIN-TACC$^{\sigma}(F, a)$ is in $FP^{NP[\log n]}$ for $\sigma = \{ad, co, st, pr\}$. Finally in [18] it was shown that the $FP^{NP[\log n]}$ upper bounds are tight. We report below the Theorems proved in that paper.

Theorem 3. *[18]* MIN-TVER$^{\sigma}(F, S)$ *is in FP for* $\sigma \in \{ad, co, st, gr\}$ *and is* $FP^{NP[\log n]}$*-complete for* $\sigma = pr$.

Theorem 4. *[18]* MIN-TACC$^{\sigma}(F, a)$ *is in FP for* $\sigma = gr$ *and* $FP^{NP[\log n]}$*-complete for* $\sigma \in \{ad, co, st, pr\}$.

5 From Theory to Practice: Evaluating min-Tver$^{\sigma}(F, S)$ and min-Tacc$^{\sigma}(F, a)$

The characterization of the computational complexity of MIN-TVER$^{\sigma}(F, S)$ and MIN-TACC$^{\sigma}(F, a)$ is relevant not only from a theoretical standpoint, but also in a practical perspective, since it suggests suitable computational strategies for these problems. We consider the two problems separately.

***Solving* min-Tver$^{\sigma}(F, S)$.** The proof of Theorem 3 in [18] contains the details of polynomial- time algorithms solving MIN-TVER$^{\sigma}(F, S)$ under $\sigma \in \{ad, co, st, gr\}$.

Hence, we focus on $\sigma = pr$. Theorem 3 states that MIN-TVER$^{pr}(F, S)$ is in FP^{NP}, and this suggests to try a solution based on *Integer Linear Programming* (ILP), that is well-suited for research problems inside this complexity class. Generally speaking, resorting to ILP solvers (such as CPLEX) is a reasonable choice (if allowed by the expressiveness of ILP), as this exploits a number of heuristics implemented in the commercial solvers that in many cases enhance the efficiency of evaluating even hard instances.

The core of our approach is a system of linear inequalities $I^{\text{MINTVER}}(F, S)$ over binary variables that is parametric on the T-AAF $F = \langle A, D, T \rangle$ and the set S, and that tests whether S is a preferred extension in F^{τ_z}, for different values of

τ'_z. In particular, τ'_z ranges over the ordered sequence $\tau'_1, \ldots, \tau'_m$ of trust degrees extracted from $\mathcal{T}(F)$ where:

1. τ'_1 is the result of MIN-TVER$^{ad}(F, S)$, i.e., the minimum trust degree such that S is admissible for $F^{\tau'_1}$;
2. $\tau'_m = \min_{a \in S} T(a)$;
3. $\tau'_2, \ldots, \tau'_{m-1}$ are the trust degrees in $\mathcal{T}(F)$ between τ_1 and τ_m.

The reason for this restriction of the search space is that S cannot be a preferred extension in any F^τ with $\tau < \tau'_1$ (since S would not be admissible) or $\tau \geq \tau'_m$ (since some of its arguments would not be present in F^τ).

Given this, for each argument a_i of F, we represent the membership of a_i to S with the boolean constant s_i. Moreover, for each $z \in [1..m]$ we use suitable variables and inequalities for testing whether S is a preferred extension in F^{τ_z}. The fact that an argument a_i is maintained or discarded in F^{τ_z} (corresponding to the fact that its trust degree $T(a_i)$ is higher or lower than τ_z) is represented with a boolean variable x_{iz} ($x_{iz} = 1$ means that a_i is NOT discarded in F^{τ_z}). Then, in order to test whether S is a preferred extension in F^{τ_z}, we search for a superset S'_z of S that is admissible and such that $|S'_z| > |S|$. We encode the membership of an argument a_i to S'_z using a binary variable s'_{iz}, where $s'_{iz} = 1$ iff $a_i \in S'_z$. Moreover, for every $z \in [1..m]$, we use a boolean variable y_z to express the result of the comparison $|S'_z| - |S|$. Then, for every z, we enforce $|S'_z| - |S|$ to be as large as possible by means of the objective function. This leads to the following ILP instance

$$I^{\text{MINTVER}}(F, S): \begin{cases} \max \sum_{z \in [1..m]} (2^z \cdot y_z) & \\ (0)\ s'_{iz} - s_i \leq y_z & \\ (1)\ s_i \leq s'_{iz} & \forall i \in [1..n], \\ (2)\ x_{iz} \geq s'_{iz} & z \in [1..m] \\ (3)\ M \cdot x_{iz} \geq T(a_i) - \tau_z & \\ (4)\ M \cdot (1 - x_{iz}) \geq \tau_z - T(a_i) + 1 & \\ (5)\ s'_{iz} + s'_{jz} \leq 1 & \\ (6)\ x_{iz} + s_j \leq \sum_{l \mid (a_l, a_i) \in D} s_l + 1 & \forall i, j \mid (a_i, a_j) \in D, \\ (7)\ x_{iz} + s'_{jz} \leq \sum_{l \mid (a_l, a_i) \in D} s'_l + 1 & z \in [1..m] \end{cases}$$

where $M = \max_{a_i \in A} T(a_i)$.

The semantics of the inequalities is the following:

(0) $y_z = 1$ iff $|S'_z| > |S|$, for each $z \in [1..m]$;
(1) every S'_z is a superset of S;
(2) every S'_z (and thus S) contains only non-discarded arguments;
(3, 4) an argument a_i is discarded in F^{τ_z} iff $T(a_i) \leq \tau_z$;
(5) every S'_z (and thus S) is conflict free;
(6) S is admissible;
(7) every S'_z is admissible.

The result R of $I^{\text{MINTVER}}(F, S)$ can be easily translated into what asked by MIN-TVER$^{pr}(F, S)$: the position of the leftmost bit 0 in R (if any) is the index of the minimum trust degree τ^* in $\mathcal{T}(F)$ such that S is a preferred extension over F^{τ^*}. Obviously, if all the bits of R are 1, it means that there is no way of

making S a preferred extension by removing all the arguments less trustworthy than some threshold.

***Solving* min-Tacc$^\sigma(F,a)$.** The case $\sigma = gr$ can be solved by the polynomial-time algorithm described in the proof of Theorem 4 in [18].

As for the other semantics, analogously to what said for MIN-TVER$^\sigma(F,S)$, the $FP^{NP[\log n]}$-completeness backs an ILP-based approach. Our formulation of MIN-TACC$^\sigma(F,a)$ as an ILP instance $I_\sigma^{\text{MINTACC}}(F,a)$ is based on searching an extension S that contains a. We represent the membership of an argument a_i to S with a boolean variable s_i (where the variable s_j corresponding to a is constrained to be 1), and the fact that a_i is maintained or discarded (owing to the threshold τ) with a boolean variable x_i ($x_i = 1$ means "a_i is NOT discarded"). The objective function consists in minimizing τ. Observe that we can resort to the same ILP instance to solve MIN-TACC$^\sigma(F,a)$ under any $\sigma \in \{ad, co, pr\}$, since an argument belongs to a complete or a preferred extension if and only if it belongs to an admissible extension. This leads to the following formulation for $I_\sigma^{\text{MINTACC}}(F,a)$ under any $\sigma \in \{ad, co, pr\}$:

$$\begin{cases} \min \tau & \\ (0)\ \ 0 \le \tau \le T(a) - 1 & \\ (1)\ \ s_j = 1 \text{ (where } j \text{ is the index of } a \text{ in } A) & \\ (2)\ \ x_i \ge s_i & \forall i \in [1..n] \\ (3)\ \ M \cdot x_i \ge T(a_i) - \tau & \forall i \in [1..n] \\ (4)\ \ M \cdot (1 - x_i) \ge \tau - T(a_i) + 1 & \forall i \in [1..n] \\ (5)\ \ s_i + s_j \le 1 & \forall i,j \,|\, (a_i, a_j) \in D \\ (6)\ \ x_i + s_j \le \sum_{l|(a_l,a_i)\in D} s_l + 1 & \forall i,j \,|\, (a_i, a_j) \in D \end{cases}$$

where inequalities $(1) - (6)$ have the following meaning:

(0) the threshold τ ranges from 0 to $T(a)$–1 (a threshold $\ge T(a)$ would discard a);
(1) a belongs to S;
(2) an argument can belong to S only if it is not discarded;
(3, 4) an argument a_i is discarded iff $T(a_i) \le \tau$;
(5) S is conflict-free;
(6) S is admissible.

Under $\sigma = st$, (6) must be replaced with:

$$x_i - s_i \le \sum_{l|(a_l,a_i)\in D} s_l \quad \forall i \in [1..n],$$

stating that every argument in F^τ outside S must be attacked by some argument in S.

6 Related Work

There are a lot of works extending AAFs: most of them have the aim of handling uncertainty [19, 27, 22, 17, 21, 20, 23, 16], or the aim of representing the

"strength" of arguments and/or attacks via preferences [3], degrees of beliefs [30] and importance of the values the arguments pertain to [6, 4].

The reasonability of associating weights with arguments or attacks has been widely discussed in the literature, and, as observed in [14], depending on the scenarios and the semantics of the weights, there are cases where assigning weights to arguments is more reasonable than to attacks, and vice versa. An example of weighted AAF where weights represent trust degrees and are associated with the arguments is [11], where a fuzzy reasoning mechanism is embedded in *SMACk*, a system for analyzing arguments taken from disputes available in online commercial websites. The latter work, along with [5, 24, 26, 7, 29], belongs to the family of approaches where the reasoning yields acceptability degrees for the arguments, obtained by suitably revising the "initial" arguments' strengths. A second family of approaches [2, 6, 28, 27, 31, 25], instead, eventually produces a binary result for each argument, stating whether it is acceptable or not. In this regard, the framework in [18] can be viewed in between these two families: on the one hand, the mechanisms invoked to decide if S is an extension and a accepted produce a binary result; on the other hand, the results of MIN-TVER$^{\sigma}(F, S)$, MIN-TACC$^{\sigma}(F, a)$ and their variants could be also viewed as "strengths" of S and a. However, these strengths are not revisions of the initial weights. For instance, consider an argument a with the highest trust degree in $\mathcal{T}(A)$. If the answer of MIN-TACC$^{\sigma}(F, a)$ is 0, it means that even discarding no argument, a is accepted, that is a positive characteristics, and not a downgrading of $T(a)$. Thus, several properties listed in [1] regarding the output strength of arguments (such as *Weakening* and *Maximality*) make no sense on the semantics of T-AAFs, as they are better tailored at reasoning paradigms belonging to the first family.

It is worth noting that the results in [18] still hold if the weights are associated to attacks: the difference in semantics does not correspond to a difference in computational complexity and solution strategies. Thus, in particular, the work in [18] completes the framework in [14] (where the problem MIN-BUDGET, dual to MIN-TACC$^{gr}(F, a)$, was addressed). In fact, the results on MIN-TVER$^{\sigma}(F, S)$ and MIN-TACC$^{\sigma}(F, a)$ can be used over the framework of [14] to use a different threshold-based mechanism tailored at the case where the weights denote levels instead of additive measures.

In this regard, the interest of the research community to extending the framework in [14] in the direction of T-AAFs is witnessed by [9], where the use of aggregate operators other than *sum* (including *min* and *max*) for reasoning on attacks to be discarded was formalized. However, no result on the computational complexity and no computational method has been proposed in [9] for these extensions.

7 Conclusions

We have provided some computational strategies for the intractable cases of the problems MIN-TVER and MIN-TACC proposed in [18]. Those problems are extensions of the verification and acceptance problems for reasoning over AAFs

where the trustworthiness of the agents is encoded as a weight function over the arguments.

We have provided a translation of the cases of MIN-TVER and MIN-TACC that have been shown to be inside the class FP^{NP} into ILP instances so that a well-established ILP solver can be invoked. Generally speaking, resorting to ILP solvers (such as CPLEX) is a reasonable choice (if allowed by the expressiveness of ILP, that is bounded by FP^{NP}), as this exploits a number of heuristics implemented in the commercial solvers that in many cases enhance the efficiency of evaluating even hard instances. Future work will be devoted to implement ILP-based strategies and compare them with the usage of SAT-solvers, that are commonly used as tools for verifying/generating the extensions and deciding the acceptance of arguments in "classical" abstract argumentation.

References

1. Amgoud, L., Ben-Naim, J., Doder, D., Vesic, S.: Acceptability semantics for weighted argumentation frameworks. In: Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), Melbourne, Australia, Aug. 19-25, 2017. pp. 56–62 (2017)
2. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Ann. Math. Artif. Intell. **34**(1-3), 197–215 (2002)
3. Amgoud, L., Vesic, S.: A new approach for preference-based argumentation frameworks. Ann. Math. Artif. Intell. **63**(2), 149–183 (2011)
4. Atkinson, K., Bench-Capon, T.: Value based reasoning and the actions of others. In: Proc. European Conf. on Artificial Intelligence (ECAI), The Hague, The Netherlands. pp. 680–688 (2016)
5. Baroni, P., Romano, M., Toni, F., Aurisicchio, M., Bertanza, G.: Automatic evaluation of design alternatives with quantitative argumentation. Argument & Computation **6**(1), 24–49 (2015)
6. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. J. Log. Comput. **13**(3), 429–448 (2003)
7. da Costa Pereira, C., Tettamanzi, A., Villata, S.: Changing one's mind: Erase or rewind? In: Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), Barcelona, Catalonia, Spain, July 16-22. pp. 164–171 (2011)
8. Coste-Marquis, S., Devred, C., Marquis, P.: Symmetric argumentation frameworks. In: Proc. of Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Barcelona, Spain. pp. 317–328 (2005)
9. Coste-Marquis, S., Konieczny, S., Marquis, P., Ouali, M.A.: Weighted attacks in argumentation frameworks. In: Proc. Int. Conf. on Knowledge Representation and Reasoning (KR), Rome, Italy. pp. 593–597 (2012)
10. Dimopoulos, Y., Torres, A.: Graph theoretical structures in logic programs and default theories. Theor. Comput. Sci. **170**(1-2), 209–244 (1996)
11. Dragoni, M., da Costa Pereira, C., Tettamanzi, A.G.B., Villata, S.: Combining argumentation and aspect-based opinion mining: The smack system. AI Commun. **31**(1), 75–95 (2018)
12. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. **77**(2), 321–358 (1995)

13. Dunne, P.E., Bench-Capon, T.J.M.: Coherence in finite argument systems. Artif. Intell. **141**(1/2), 187–203 (2002)
14. Dunne, P.E., Hunter, A., McBurney, P., Parsons, S., Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artif. Intell. **175**(2) (2011)
15. Dunne, P.E., Wooldridge, M.: Complexity of abstract argumentation. In: Argumentation in Artificial Intelligence, pp. 85–104 (2009)
16. Fazzinga, B., Flesca, S., Furfaro, F.: Probabilistic bipolar abstract argumentation frameworks: complexity results. In: Proc. Int. Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden. pp. 1803–1809. ijcai.org (2018)
17. Fazzinga, B., Flesca, S., Furfaro, F.: Complexity of fundamental problems in probabilistic abstract argumentation: Beyond independence. Artif. Intell. **268**, 1–29 (2019)
18. Fazzinga, B., Flesca, S., Furfaro, F.: Embedding the trust degrees of agents in abstract argumentation. In: Proc. European Conf. on Artificial Intelligence (ECAI) 2020. Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 737–744. IOS Press (2020)
19. Fazzinga, B., Flesca, S., Furfaro, F.: Revisiting the notion of extension over incomplete abstract argumentation frameworks. In: Proc. Int. Joint Conference on Artificial Intelligence, IJCAI 2020. pp. 1712–1718. ijcai.org (2020)
20. Fazzinga, B., Flesca, S., Parisi, F.: Efficiently estimating the probability of extensions in abstract argumentation. In: Proc. Int Conf. on Scalable Uncertainty Management (SUM). pp. 106–119 (2013)
21. Fazzinga, B., Flesca, S., Parisi, F.: On the complexity of probabilistic abstract argumentation. In: Proc. Int. Joint Conference on Artificial Intelligence (IJCAI) (2013)
22. Fazzinga, B., Flesca, S., Parisi, F.: On the complexity of probabilistic abstract argumentation frameworks. ACM Trans. Comput. Log. (TOCL) **16**(3), 22 (2015)
23. Fazzinga, B., Flesca, S., Parisi, F., Pietramala, A.: PARTY: A mobile system for efficiently assessing the probability of extensions in a debate. In: Proc. Int. Conf. on Database and Expert Systems Applications (DEXA). pp. 220–235 (2015)
24. Gabbay, D.M., Rodrigues, O.: Equilibrium states in numerical argumentation networks. Logica Universalis **9**(4), 411–473 (2015)
25. Hunter, A.: Probabilistic qualification of attack in abstract argumentation. Int. J. Approx. Reasoning **55**(2), 607–638 (2014)
26. Leite, J., Martins, J.: Social abstract argumentation. In: Proc. Int. Joint Conf. on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011. pp. 2287–2292 (2011)
27. Li, H., Oren, N., Norman, T.J.: Probabilistic argumentation frameworks. In: Proc. Int. Workshop on Theory and Applications of Formal Argumentation (TAFA), Barcelona, Spain. pp. 1–16 (2011)
28. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. **173**(9-10), 901–934 (2009)
29. Rago, A., Toni, F., Aurisicchio, M., Baroni, P.: Discontinuity-free decision support with quantitative argumentation debates. In: Proc. Int. Conf. on Principles of Knowledge Representation and Reasoning (KR), Cape Town, South Africa, April 25-29. pp. 63–73 (2016)
30. Santini, F., Jøsang, A., Pini, M.S.: Are my arguments trustworthy? abstract argumentation with subjective logic. In: Proc. Int. Conf. on Information Fusion (FUSION), Cambridge, UK, July 10-13. pp. 1982–1989 (2018)

31. Thimm, M.: A probabilistic semantics for abstract argumentation. In: Proc. European Conf. on Artificial Intelligence (ECAI), Montpellier, France. pp. 750–755 (2012)