Stability in Abstract Argumentation

Jean-Guy Mailly , **Julien Rossit**

LIPADE, Université de Paris

{jean-guy.mailly, julien.rossit}@u-paris.fr

Abstract

The notion of stability in a structured argumentation setup characterizes situations where the acceptance status associated with a given literal will not be impacted by any future evolution of this setup. In this paper, we abstract away from the logical structure of arguments, and we transpose this notion of stability to the context of Dungean argumentation frameworks. In particular, we show how this problem can be translated into reasoning with Argument-Incomplete AFs. Then we provide preliminary complexity results for stability under four prominent semantics, in the case of both credulous and skeptical reasoning. Finally, we illustrate to what extent this notion can be useful with an application to argument-based negotiation.

1 Introduction

Formal argumentation is a family of non-monotonic reasoning approaches with applications to (*e.g.*) multi-agent systems (McBurney, Parsons, and Rahwan 2012), automated negotiation (Dimopoulos, Mailly, and Moraitis 2019) or decision making (Amgoud and Vesic 2012). Roughly speaking, we can group the research in this domain in two families: abstract argumentation (Dung 1995) and structured argumentation (Besnard et al. 2014). The former is mainly based on the seminal paper proposed by Dung, where abstract argumentation frameworks (AFs) are defined as directed graphs where the nodes represent arguments and the edges represent attacks between them. In this setting, the nature of arguments and attacks is not defined, only their interactions are represented in order to determine the acceptability status of arguments. On the opposite, different settings have been proposed where the arguments are built from logical formulas or rules, and the nature of attacks is based on logical conflicts between the elements inside the arguments. See *e.g.* (Baroni, Gabbay, and Giacomin 2018) for a recent overview of abstract and structured argumentation.

In a particular structured argumentation setting, the notion of stability has been defined recently (Testerink, Odekerken, and Bex 2019). Intuitively, it represents a situation where a certain argument of interest will not anymore have the possibility to change its acceptability status. Either it is currently accepted and it will remain so, or on the contrary it is currently rejected, and nothing could make it accepted in the future. In the existing work on this topic, the authors

mention some application to crime investigation (more precisely, Internet trade fraud). We also have in mind some other natural applications, like automated negotiation. For instance, if an agent is certain her argument for supporting her preferred offer cannot be accepted at any future step of the debate, she can switch her offer to another one, that may be less preferred, but at least could be accepted.

In this paper, we adapt the notion of stability to abstract argumentation, and we show that checking stability is equivalent to performing some well-known reasoning tasks in Argument-Incomplete AFs (Baumeister, Rothe, and Schadrack 2015; Baumeister, Neugebauer, and Rothe 2018; Niskanen et al. 2020). While existing work on stability in structured argumentation focuses on a particular semantics (namely the grounded semantics), our approach is generic with respect to the underlying extension-based semantics. Moreover we consider both credulous and skeptical variants of argumentative reasoning.

This paper is organized as follows. Section 2 introduces the basic notions of abstract argumentation literature in which our work takes place, and presents the concept of stability for structured argumentation frameworks. We then propose in Section 3 a counterpart of this notion of stability adapted to abstract argumentation frameworks, and we show how we can reduce it to well-known reasoning tasks. We provide some lower and upper bounds for the computational complexity of checking whether an AF is stable. Section 4 then describes an application scenario in the context of automated negotiation. Finally, Section 5 discusses related work, and Section 6 concludes the paper by highlighting some promising future works.

2 Background

2.1 Abstract Argumentation

Let us first introduce the abstract argumentation framework defined in (Dung 1995).

Definition 1. *An* argumentation framework *(AF) is a pair* $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ *where* $\mathcal{A}$ *is the set of* arguments *and* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is the* attack relation.

In this framework, we are not concerned by the precise nature of arguments (*e.g.* their internal structure or their origin) and attacks (*e.g.* the presence of contradictions between elements on which arguments are built). Only the relations

between arguments (*i.e.* the attacks) are taken into account to evaluate the acceptability of arguments.

We focus on finite AFs, *i.e.* AFs with a finite set of arguments. For $a, b \in \mathcal{A}$, we say that a *attacks* b if $(a, b) \in \mathcal{R}$. Moreover, if b attacks some $c \in \mathcal{A}$, then a *defends* c against b. These notions are extended to sets of arguments: $S \subseteq \mathcal{A}$ attacks (respectively defends) $b \in \mathcal{A}$ if there is some $a \in S$ that attacks (respectively defends) b. The acceptability of arguments is evaluated through a notion of extension, *i.e.* a set of arguments that are jointly acceptable. To be considered as an extension, a set has to satisfy some minimal requirements:

- $S \subseteq \mathcal{A}$ is conflict-free (denoted $S \in \mathsf{cf}(\mathcal{F})$) iff $\forall a, b \in S$, $(a, b) \notin \mathcal{R}$;
- $S \in \mathsf{cf}(\mathcal{F})$ is admissible (denoted $S \in \mathsf{ad}(\mathcal{F})$) iff S defends all its elements against all their attackers.

Then, Dung defines several semantics:

Definition 2. *Given $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ an AF, a set $S \subseteq \mathcal{A}$ is:*

- *a* complete *extension ($S \in \mathsf{co}(\mathcal{F})$) iff $S \in \mathsf{ad}(\mathcal{F})$ and S contains all the arguments that it defends;*
- *a* preferred *extension ($S \in \mathsf{pr}(\mathcal{F})$) iff S is a $\subseteq$-maximal complete extension;*
- *the unique* grounded *extension ($S \in \mathsf{gr}(\mathcal{F})$) iff S is the $\subseteq$-minimal complete extension;*
- *a* stable *extension ($S \in \mathsf{st}(\mathcal{F})$) iff $S \in \mathsf{cf}(\mathcal{F})$ and S attacks each $a \in \mathcal{A} \setminus S$,*

where $\subseteq$-maximal and $\subseteq$-minimal denote respectively the maximal and the minimal elements for classical set inclusion.

Example 1. *Let $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ be the AF depicted in Figure 1. Nodes in the graph represent the arguments $\mathcal{A}$, while the edges correspond to the attacks $\mathcal{R}$. Its extensions for $\sigma \in \{\mathsf{gr}, \mathsf{st}, \mathsf{pr}, \mathsf{co}\}$ are given in Table 1.*
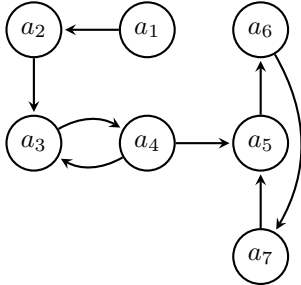

Figure 1: An Example of AF $\mathcal{F}$

Semantics σ	σ-extensions
grounded	$\{\{a_1\}\}$
stable	$\{\{a_1, a_4, a_6\}\}$
preferred	$\{\{a_1, a_4, a_6\}, \{a_1, a_3\}\}$
complete	$\{\{a_1, a_4, a_6\}, \{a_1, a_3\}, \{a_1\}\}$

Table 1: σ-Extensions of $\mathcal{F}$

We refer the interested reader to (Baroni, Caminada, and Giacomin 2018) for more details about these semantics, as well as other ones defined after Dung's initial work. From the set of extensions $\sigma(\mathcal{F})$ (for $\sigma \in \{\mathsf{co}, \mathsf{pr}, \mathsf{gr}, \mathsf{st}\}$), we define two reasoning modes:

- an argument $a \in \mathcal{A}$ is credulously accepted with respect to σ iff $a \in S$ for some $S \in \sigma(\mathcal{F})$;
- an argument $a \in \mathcal{A}$ is skeptically accepted with respect to σ iff $a \in S$ for each $S \in \sigma(\mathcal{F})$.

Then, a possible enrichment of Dung's framework consists in taking into account some uncertainty in the AF. This yields the notion of *Incomplete AFs*, studied *e.g.* in (Baumeister, Rothe, and Schadrack 2015; Baumeister, Neugebauer, and Rothe 2018; Niskanen et al. 2020). Here, we focus on a particular type, namely *Argument-Incomplete AFs*, but for a matter of simplicity we just refer to them as Incomplete AFs.

Definition 3. *An incomplete argumentation framework (IAF) is a tuple $\mathcal{I} = \langle \mathcal{A}, \mathcal{A}^? \mathcal{R} \rangle$ where*

- *$\mathcal{A}$ is the set of* certain *arguments;*
- *$\mathcal{A}^?$ is the set of* uncertain *arguments;*
- *$\mathcal{R} \subseteq (\mathcal{A} \cup \mathcal{A}^?) \times (\mathcal{A} \cup \mathcal{A}^?)$ is the* attack relation;

and $\mathcal{A}, \mathcal{A}^?$ are disjoint sets of arguments.

Example 2. *The IAF $\mathcal{I} = \langle \mathcal{A}, \mathcal{A}^?, \mathcal{R} \rangle$ is shown on Figure 2. The dotted nodes represent the uncertain arguments $\mathcal{A}^?$. Plain nodes and arrows have the same meaning as previously.*
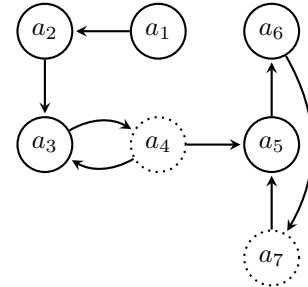

Figure 2: An Example of IAF $\mathcal{I}$

Uncertain arguments are those that may not actually belong to the system (for instance because of some uncertainty about the agent's environment). There are different ways to "solve" the uncertainty in an IAF, that correspond to different completions:

Definition 4. *Given $\mathcal{I} = \langle \mathcal{A}, \mathcal{A}^? \mathcal{R} \rangle$ an IAF, a completion is an AF $\mathcal{F} = \langle \mathcal{A}', \mathcal{R}' \rangle$ where*

- *$\mathcal{A} \subseteq \mathcal{A}' \subseteq \mathcal{A} \cup \mathcal{A}^?$;*
- *$\mathcal{R}' = \mathcal{R} \cap (\mathcal{A}' \times \mathcal{A}')$.*

Example 3. *Considering again $\mathcal{I}$ from the previous example, we show all its completions at Figure 3. For each uncertain argument in $\mathcal{A}^? = \{a_4, a_7\}$, there are two possibilities: either the argument is present, or it is not. Thus, there are four completions.*
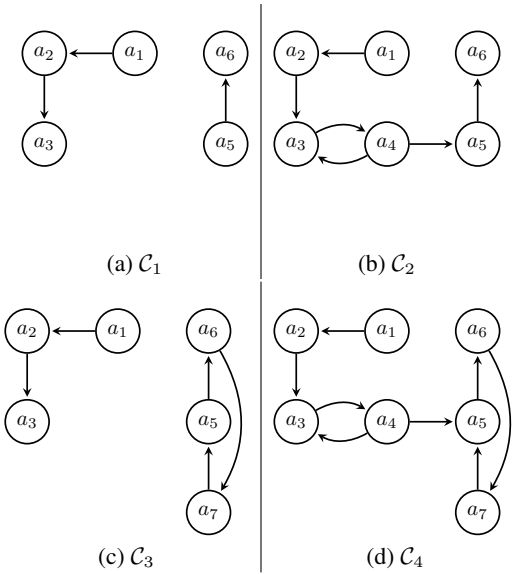
(a) $\mathcal{C}_1$ (b) $\mathcal{C}_2$

(c) $\mathcal{C}_3$ (d) $\mathcal{C}_4$

Figure 3: The Completions of $\mathcal{I}$

This means that a completion is a "classical" AF made of all the certain arguments, some of the uncertain elements, and all the attacks that concern the selected arguments. Reasoning with an IAF generalizes reasoning with an AF, by taking into account either some or each completion. Formally, given $\mathcal{I}$ an IAF and σ a semantics, the status of an argument $a \in \mathcal{A}$ is:

- *possibly credulously* accepted with respect to σ iff a belongs to some σ-extension of some completion of $\mathcal{I}$;

- *possibly skeptically* accepted with respect to σ iff a belongs to each σ-extension of some completion of $\mathcal{I}$;

- *necessarily credulously* accepted with respect to σ iff a belongs to some σ-extension of each completion of $\mathcal{I}$;

- *necessarily skeptically* accepted with respect to σ iff a belongs to each σ-extension of each completion of $\mathcal{I}$.

Example 4. *Let us consider again $\mathcal{I}$ from the previous example, and its completions $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$. We observe that a_1 is necessarily skeptically accepted for any semantics, since it appears unattacked in every completion (thus, it belongs to every extension of every completion).*

On the opposite, a_6 is possibly credulously accepted with respect to the preferred semantics: it belongs to some extension of $\mathcal{C}_4$. It is not skeptically accepted (because $\{a_1, a_3\}$ is a preferred extension of $\mathcal{C}_4$ as well), and it is not necessarily accepted (because in $\mathcal{C}_1$, it is not defended against a_5, thus it cannot belong to any extension).

2.2 Stability in Structured Argumentation

Now we briefly introduce the argumentation setting from (Testerink, Odekerken, and Bex 2019), based on *ASPIC*$^+$ (Modgil and Prakken 2014).

Let us start with the notation that is used to represent the negation of a literal, *i.e.* for a propositional variable p, $-p =$

$\neg p$ and $-(\neg p) = p$, with $\neg$ the classical negation. We call p (respectively $\neg p$) a positive (respectively negative) literal.

Definition 5. *An argumentation setup is a tuple $AS = \langle L, R, Q, K, \tau \rangle$ where:*

- L *is a set of* literals *s.t.* $l \in L$ *implies* $-l \in L$;

- R *is a set of* defeasible rules $p_1, \ldots, p_m \Rightarrow q$ *s.t.* $p_1, \ldots, p_m, q \in L$. *Such a rule is called "a rule for q".*

- $Q \subseteq L$ *is a set of* queryable literals, *s.t. no* $q \in Q$ *is a negative literal;*

- $K \subseteq L$ *is the agent's (consistent)* knowledge base;

- $\tau \in L$ *is a particular literal called the* topic.

Usual mechanisms are used to define arguments and attacks. An argument for a literal q is an inference tree rooted in a rule $p_1, \ldots, p_m \Rightarrow q$, such that for each p_i, there is a child node that is either an argument for p_i, or an element of the knowledge base. Then, an argument A attacks an argument B if the literal supported by A is the negation of some literal involved in the construction of B. From the sets of arguments and attacks built in this way, the grounded extension is defined as usual (see Definition 2).

Given an argumentation setup AS, the status of the topic τ may be:

- *unsatisfiable* if there is no argument for τ in AS;

- *defended* if there is an argument for τ in the grounded extension of AS;

- *out* if there are some arguments for τ in AS, and all of them are attacked by the grounded extension;

- *blocked* in the remaining case.

Then, stability can be defined, based on the following notion of future setups:

Definition 6. *Let $AS = \langle L, R, Q, K, \tau \rangle$ be an argumentation setup. The set of future setups of AS, denoted by $F(AS)$, is defined by $F(AS) = \{\langle L, R, Q, K', \tau \rangle \mid K \subseteq K'\}$. AS is called stable if for each $AS' \in F(AS)$, the status of τ is the same as in AS.*

Intuitively, a future setup is built by adding new literals to the knowledge base (keeping the consistency property, of course). Then, new arguments and attacks may be built thanks to these new literals. The setup is stable if these new arguments and attacks do not change the status of the topic.

To conclude this section, let us mention that (Testerink, Odekerken, and Bex 2019) provides a sound algorithm that approximates the reasoning task of checking the stability of the setup. This algorithm is however not complete, *i.e.* AS is actually stable if the algorithm outcome is a positive answer, but there are stable setups that are not identified by the algorithm. This algorithm has the interest of being polynomially computable (more precisely, it stops in $\mathcal{O}(n^2)$ steps, where $n = |L| + |R|$).

3 Stability in Abstract Argumentation

In this section, we describe how we adapt the notion of stability to abstract argumentation. Contrary to previous works, we do not focus on a specific semantics, and thus we consider both credulous and skeptical reasoning. Moreover, we

provide a translation of the stability problem into reasoning with AFs and IAFs. Despite being theoretically intractable, efficient algorithms exist for solving these problems in practice. So it paves the way to future implementations of an exact algorithm for checking stability, and its applications to concrete scenarios.

3.1 Formal Definition of Stability in AFs

From now on, we consider a finite *argumentation universe* that is represented by an AF $\mathcal{F}_U = \langle \mathcal{A}_U, \mathcal{R}_U \rangle$. We suppose that any "valid" AF is made of arguments and attacks in $\mathcal{F}_U$, *i.e.* $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ s.t. $\mathcal{A} \subseteq \mathcal{A}_U$ and $\mathcal{R} = \mathcal{R}_U \cap (\mathcal{A} \times \mathcal{A})$.

Definition 7. *Given an AF $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$, we call the* future *AFs the set of AFs $F(\mathcal{F}) = \{\mathcal{F}' = \langle \mathcal{A}', \mathcal{R}' \rangle \mid \mathcal{A} \subseteq \mathcal{A}'\}$.*

Intuitively speaking, this means that a future AF represents a possible way to continue the argumentative process (by adding arguments and attacks), accordingly to $\mathcal{F}_U$. This corresponds to some kind of expansions of $\mathcal{F}$ (Baumann and Brewka 2010), where the authorized expansions are constrained by $\mathcal{F}_U$. This is reminiscent of the set of authorized updates defined in (de Saint-Cyr et al. 2016). Notice that $\mathcal{F}$ is a particular future AF.

Now we have all the elements to define stability.

Definition 8. *Given an AF $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$, $a \in \mathcal{A}$ an argument, and σ a semantics, we say that $\mathcal{F}$ is credulously (respectively skeptically) σ-stable with respect to a iff*

- *either $\forall \mathcal{F}' \in F(\mathcal{F})$, a is credulously (respectively skeptically) accepted with respect to σ;*
- *or $\forall \mathcal{F}' \in F(\mathcal{F})$, a is not credulously (respectively skeptically) accepted with respect to σ.*

Although in this paper we focus on $\sigma \in \{\mathsf{gr}, \mathsf{st}, \mathsf{pr}, \mathsf{co}\}$, the definition of stability is generic, and the concept can be applied when any extension semantics is used (Baroni, Caminada, and Giacomin 2018).

Example 5. *Let us consider the argumentation universe $\mathcal{F}_U$ and the AF $\mathcal{F}$, both depicted in Figure 4. The argument a_3 is not credulously σ-stable for $\sigma = \mathsf{st}$, since it is credulously accepted in $\mathcal{F}$, but not in the future AF where a_2 is added. On the contrary, it is skeptically σ-stable since it is not skeptically accepted in $\mathcal{F}$, nor in any future AF.*

a_6 is skeptically σ-stable as well, but for another reason: indeed we observe that in $\mathcal{F}$ (and in any future AF), a_6 is defended by the (unattacked) argument a_7, thus it belongs to every extension.

On this simple example, it may seem obvious to determine that a_5, a_6 and a_7 will keep their status. However, let us notice that determining whether an argument keeps its status when an AF is updated has been studied, and is not a trivial question in the general case (Baroni, Giacomin, and Liao 2014; Alfano, Greco, and Parisi 2019).

3.2 Computational Issues

We now provide a method for checking the stability of an AF with respect to some argument. The method is generic regarding the underlying extension semantics. It is based on the observation that the set of future AFs can be encoded into a single IAF (see Definition 3).
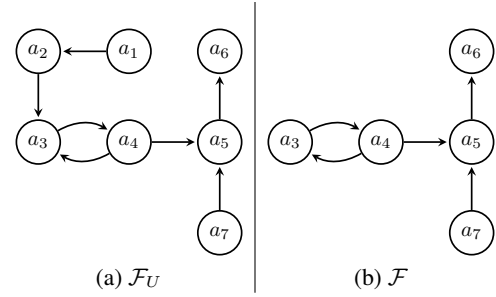
Figure 4: The Argumentation Universe $\mathcal{F}_U$ and a Possible AF $\mathcal{F}$

Definition 9. *Given $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$, the corresponding IAF is $\mathcal{I}_\mathcal{F} = \langle \mathcal{A}, \mathcal{A}_U \setminus \mathcal{A}, \mathcal{R}_U \rangle$.*

The corresponding IAF is built from the whole set of arguments that appear in the universe. The ones that belong to $\mathcal{F}$ are the certain arguments, while the other ones are uncertain. Then of course, all the attacks from the universe appear in the IAF. The set of completions of $\mathcal{I}_\mathcal{F}$ is actually $F(\mathcal{F})$.

Example 6. *Figure 5 shows the IAF corresponding to $\mathcal{F}$. The arguments that belong to the universe but not to $\mathcal{F}$ (namely, a_1 and a_2) appear as uncertain arguments. This means that the four completions of this IAF correspond to $F(\mathcal{F})$.*
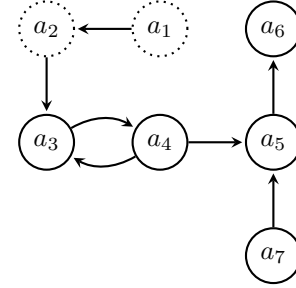
Figure 5: The IAF $\mathcal{I}_\mathcal{F}$ Corresponding to $\mathcal{F}$

We give a characterization of stability based on the IAF corresponding to an AF.

Proposition 1. *Given an AF $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$, $a \in \mathcal{A}$ an argument, and σ a semantics, $\mathcal{F}$ is credulously (respectively skeptically) σ-stable with respect to a iff*

- *either a is necessarily credulously (respectively skeptically) accepted in $\mathcal{I}_\mathcal{F}$ with respect to σ;*
- *or a is not possibly credulously (respectively skeptically) accepted in $\mathcal{I}_\mathcal{F}$ with respect to σ.*

This result shows that solving efficiently the stability problem is possible, using for instance the SAT-based piece of software `taeydennae` (Niskanen et al. 2020) for reasoning in $\mathcal{I}_\mathcal{F}$.

Now, we provide preliminary complexity results. We start with upper bounds for the computational complexity of stability.

Proposition 2. *The upper bound complexity of checking whether an AF is (credulously or skeptically) σ-stable with respect to an argument is as presented in Table 2.*

σ	Credulous	Skeptical
st	$\in \Pi_2^P$	$\in \Sigma_2^P$
co	$\in \Pi_2^P$	$\in$ coNP
gr	$\in$ coNP	$\in$ coNP
pr	$\in \Pi_3^P$	$\in \Sigma_3^P$

Table 2: Upper Bound Complexity of Checking Stability

Sketch of proof. Non-deterministically guess a pair of future AFs $\mathcal{F}'$ and $\mathcal{F}''$. Check that a is credulously (respectively skeptically) accepted in $\mathcal{F}'$, and a is not credulously (respectively skeptically) accepted in $\mathcal{F}''$. The complexity of credulous (respectively skeptical) acceptance in AFs (Dvořák and Dunne 2018) allows to deduce an upper bound for credulous (respectively skeptical) stability. □

Now we also identify lower bounds for the computational complexity of stability.

Proposition 3. *The lower bound complexity of checking whether an AF is (credulously or skeptically) σ-stable with respect to an argument is as presented in Table 3.*

σ	Credulous	Skeptical
st	NP-hard	coNP-hard
co	NP-hard	P-hard
gr	P-hard	P-hard
pr	NP-hard	Π_2^P-hard

Table 3: Lower Bound Complexity of Checking Stability

Sketch of proof. Credulous (respectively skeptical) acceptance in an AF $\mathcal{F}$ can be reduced to credulous (respectively skeptical) stability, such that the current AF is $\mathcal{F}$, and the argumentation universe is $\mathcal{F}_U = \mathcal{F}$. Thus, $\mathcal{F}$ is credulously (respectively skeptically) σ-stable with respect to some argument a iff a is credulously (respectively skeptically) accepted in $\mathcal{F}$ with respect to σ. The nature of the reduction (its computation is bounded with logarithmic space and polynomial time) makes it suitable for determining both P-hardness and C-hardness, for $C \in \{NP, coNP, \Pi_2^P\}$. Thus, we can conclude that stability is at least as hard as acceptance in AFs. From known complexity results for AFs (Dvořák and Dunne 2018), we deduce the lower bounds given in Table 3. □

4 Applying Stability to Automated Negotiation

Now, we discuss the benefit of stability in a concrete application scenario, namely automated negotiation. Let us consider a simple negotiation framework, where practical arguments (*i.e.* those that support some offers) are mutually exclusive, and for each agent there is a preference relation between the offers supported by these arguments (for instance, these preferences can be obtained from a notion of utility associated with each offer). So, each agent's goal is to make her preferred practical argument (*i.e.* the one that supports the preferred offer) accepted at the end of the debate. Each agent, in turn, can add one (or more) argument(s) that defend her preferred argument. In this first version of the negotiation framework, agents have a total ignorance about their opponent.

Then, an enriched version of this protocol can be defined, where the agents use the notion of argumentation universe to model their (uncertain) knowledge about the opponent. Then, stability can help the agent to obtain a better outcome: if at some point, the agent's preferred practical argument is rejected and stable, this means that this argument will not be accepted at the end of the debate, whatever the actual moves of the other agents. It is then profitable to the agent to change her goal, defending now the argument that supports her second preferred offer instead of the first one. This can reduce the number of rounds in the negotiation (and thus, any communication cost associated with these rounds), and even improve the outcome of the negotiation for the agent.

Let us now provide a concrete example. We suppose that the offers $O = \{o_1, o_2, o_3\}$ are supported by one practical argument each, *i.e.* $\{p_1, p_2, p_3\}$ with p_i supporting o_i. The practical arguments are mutually exclusive. The preferences of the agents are opposed: agent 1 has a preference ranking $o_3 >_1 o_2 >_1 o_1$, while the preferences of agent 2 are $o_1 >_2 o_2 >_2 o_3$. So, at the beginning of the debate, the goal of agent 1 (respectively agent 2) is to accept the argument p_3 (respectively p_1). Let us suppose that the first round consists in agent 1 attacking the argument p_1 with three arguments a_1, a_2 and a_3, thus defending p_3. This situation is depicted in Figure 6.
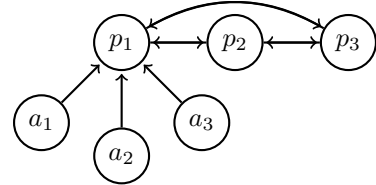

Figure 6: The Negotiation Debate $\mathcal{F}_1$

In $\mathcal{F}_1$, that represents the state of the debate after agent 1's move, the argument p_1 is clearly rejected under the stable semantics,[1] since it is not defended against a_1, a_2 and a_3. Consider that agent 2 has one argument at her disposal, a_4, with the corresponding attacks (a_4, a_3) and (a_4, p_2). Without a possibility to anticipate the evolution of the debate, the best action for agent 2 is to utter this argument, thus defending p_1 against a_3 and p_2.

Now, let us suppose that agent 2 has an opponent modelling in the form of the argumentation universe $\mathcal{F}_U$, described at Figure 7.

Now, we observe that p_1 does not appear in any extension of any future framework. Indeed, it is obvious that, if one of a_4, a_5 and a_6 is not present at the end of the debate, then

---

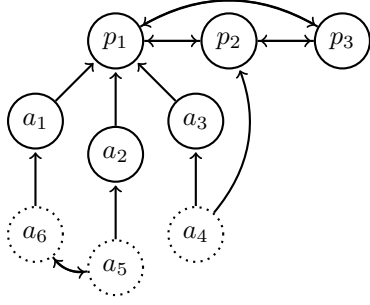[1] As well as any semantics considered in this paper.

Figure 7: The Argumentation Universe $\mathcal{F}_U$

p_1 is not defended against (respectively) a_3, a_2 or a_1. Otherwise, if a_4, a_5 and a_6 appear together, the mutual attack between a_5 and a_6 will be at the origin of two extensions, one where a_6 appears with a_2 (then defeating p_1), and the other one containing a_5 and a_1 (thus defeating again p_1). This means that p_1 is rejected in $\mathcal{F}_1$, and it is (both credulously and skeptically) σ-stable. In this situation, it is in the interest of agent 2 to stop arguing, and proposing instead the option supported by the argument p_2. Indeed, according to the agent's preferences, p_2 is the best option if p_1 is not available anymore. Not only using the notion of stability in the argumentation universe allows to stop the debate earlier, but it also allows the agent 2 to propose her second best option, which would not be possible if she had uttered a_4.

5 Related Work

Dynamics of abstract argumentation frameworks (Doutre and Mailly 2018) has received much attention in the last decade. We can summarize this field in two kinds of approaches: the goal either is to modify an AF to enforce some (set of) arguments as accepted, or to determine to what extent the acceptability of arguments is impacted by some changes in the AF. In the first family, we can mention extension enforcement (Baumann and Brewka 2010), that is somehow dual to stability. Enforcement is exactly the operation that consists in finding whether it is possible to modify an AF to ensure that a set of arguments becomes (included in) an extension, while stability is the property of an argument that will keep its acceptance status, whatever the future evolution of the AF. Control Argumentation Frameworks (Dimopoulos, Mailly, and Moraitis 2018) are also somehow related to stability, since they are a generalization of Dung's AFs that permit to realize extension enforcement under uncertainty.

The second family of works in the field of argumentation dynamics are those that propose efficient approaches to recompute the extensions or the set of (credulously/skeptically) accepted arguments when the AF is modified (Baroni, Giacomin, and Liao 2014; Alfano, Greco, and Parisi 2019). Although related to stability, these approaches do not provide an algorithmic solution to the problem studied in our work, since they focus on one update of the AF at once, instead of the set of all the future AFs.

6 Conclusion

In this paper, we have addressed a first study which investigates to what extent the notion of stability can be adapted to abstract argumentation frameworks. In particular, we have shown how it relates with Incomplete AFs, that are a model that integrates uncertainty in abstract argumentation. Our preliminary complexity results, as well as the translation of stability into reasoning with IAFs pave the way to the development of efficient computational approaches for stability, taking benefit from SAT-based techniques. Finally, we have shown that, besides the existing application of stability to Internet fraud inquiry (Testerink, Odekerken, and Bex 2019), this concept has other potential applications, like automated negotiation.

This paper opens the way for several promising research tracks. First of all, we plan to study more in depth complexity issues in order to determine tight results for the semantics that were studied here. Other direct future works include the investigation of other semantics, and the implementation of our stability solving technique in order to experimentally evaluate its impact in a context of automated negotiation.

We have focused on stability in extension semantics, which means that an argument will either remain accepted, or remain unaccepted. However, in some cases, it is important to deal more finely with unaccepted arguments. It is possible with 3-valued labellings (Caminada 2006). Studying the notion of stability when such labellings are used to evaluate the acceptability of arguments is a natural extension of our work.

In some contexts, the assumption of a completely known argumentation universe is too strong. For such cases, it seems that using arbitrary IAFs (with also uncertainty on the attack relation) is a potential solution. Uncertainty on the existence (or direction) of attacks makes sense, for instance, when preferences are at play. Indeed, dealing with preferences in abstract argumentation usually involves a notion of defeat relation, that is a combination of the attacks and preferences. This defeat relation may somehow "cancel" or "reverse" the initial attack (Amgoud and Cayrol 2002; Amgoud and Vesic 2014; Kaci, van der Torre, and Villata 2018), thus some uncertainty or ignorance about the other agents' preferences can be represented as uncertainty in the attack relation of the argumentation universe.

We are also interested in stability for other abstract argumentation frameworks. Besides preference-based argumentation that we have already mentioned, Dung's AFs has been generalized by adding a support relation (Amgoud et al. 2008), or associating quantitative weights with attacks (Dunne et al. 2011) or arguments (Rossit et al. 2020), or associating values with arguments (Bench-Capon 2002). But adapting the notion of stability to these frameworks may require different techniques than the one used in this paper. Also, the recent claim-based argumentation (Dvořák and Woltran 2020) provides an interesting bridge between structured argumentation and purely abstract frameworks. It makes sense to study stability in this setting, as a step that would make our results for different semantics and reasoning modes available for structured argumentation frameworks.

References

Alfano, G.; Greco, S.; and Parisi, F. 2019. An efficient algorithm for skeptical preferred acceptance in dynamic argumentation frameworks. In *Proc. of IJCAI'19*, 18–24.

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

Amgoud, L., and Vesic, S. 2012. On the use of argumentation for multiple criteria decision making. In *Proc. of IPMU'12*, 480–489.

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

Amgoud, L.; Cayrol, C.; Lagasquie-Schiex, M.; and Livet, P. 2008. On bipolarity in argumentation frameworks. *Int. J. Intell. Syst.* 23(10):1062–1093.

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*. College Publications. 159–236.

Baroni, P.; Gabbay, D. M.; and Giacomin, M. 2018. *Handbook of Formal Argumentation*. College Publications.

Baroni, P.; Giacomin, M.; and Liao, B. 2014. On topology-related properties of abstract argumentation semantics. A correction and extension to dynamics of argumentation systems: A division-based method. *Artif. Intell.* 212:104–115.

Baumann, R., and Brewka, G. 2010. Expanding argumentation frameworks: Enforcing and monotonicity results. In *Proc. of COMMA'10*, 75–86.

Baumeister, D.; Neugebauer, D.; and Rothe, J. 2018. Credulous and skeptical acceptance in incomplete argumentation frameworks. In *Proc. of COMMA'18*, 181–192.

Baumeister, D.; Rothe, J.; and Schadrack, H. 2015. Verification in argument-incomplete argumentation frameworks. In *Proc. of ADT'15*, 359–376.

Bench-Capon, T. J. M. 2002. Value-based argumentation frameworks. In *Proc. of NMR'02*, 443–454.

Besnard, P.; García, A. J.; Hunter, A.; Modgil, S.; Prakken, H.; Simari, G. R.; and Toni, F. 2014. Introduction to structured argumentation. *Argument & Computation* 5(1):1–4.

Caminada, M. 2006. On the issue of reinstatement in argumentation. In *Proc. of JELIA'06*, 111–123.

de Saint-Cyr, F. D.; Bisquert, P.; Cayrol, C.; and Lagasquie-Schiex, M. 2016. Argumentation update in YALLA (yet another logic language for argumentation). *Int. J. Approx. Reason.* 75:57–92.

Dimopoulos, Y.; Mailly, J.-G.; and Moraitis, P. 2018. Control argumentation frameworks. In *Proc. of AAAI'18*, 4678–4685.

Dimopoulos, Y.; Mailly, J.-G.; and Moraitis, P. 2019. Argumentation-based negotiation with incomplete opponent profiles. In *Proc. of AAMAS'19*, 1252–1260.

Doutre, S., and Mailly, J.-G. 2018. Constraints and changes: A survey of abstract argumentation dynamics. *Argument & Computation* 9(3):223–248.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–358.

Dunne, P. E.; Hunter, A.; McBurney, P.; Parsons, S.; and Wooldridge, M. J. 2011. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.* 175(2):457–486.

Dvořák, W., and Dunne, P. E. 2018. Computational problems in formal argumentation and their complexity. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*. College Publications. 631–688.

Dvořák, W., and Woltran, S. 2020. Complexity of abstract argumentation under a claim-centric view. *Artif. Intell.* 285:103290.

Kaci, S.; van der Torre, L. W. N.; and Villata, S. 2018. Preference in abstract argumentation. In *Proc. of COMMA'18*, 405–412.

McBurney, P.; Parsons, S.; and Rahwan, I., eds. 2012. *Proc. of ArgMAS'11*, volume 7543 of *Lecture Notes in Computer Science*. Springer.

Modgil, S., and Prakken, H. 2014. The *ASPIC$^+$* framework for structured argumentation: a tutorial. *Argument & Computation* 5(1):31–62.

Niskanen, A.; Neugebauer, D.; Järvisalo, M.; and Rothe, J. 2020. Deciding acceptance in incomplete argumentation frameworks. In *Proc. of AAAI'20*, 2942–2949.

Rossit, J.; Mailly, J.-G.; Dimopoulos, Y.; and Moraitis, P. 2020. United we stand: Accruals in strength-based argumentation. *Argument & Computation*. To appear.

Testerink, B.; Odekerken, D.; and Bex, F. 2019. A method for efficient argument-based inquiry. In *Proc. of FQAS'19*, 114–125.