Semi-Stable Semantics for Abstract Dialectical Frameworks

Atefeh Keshavarzi Zafarghandi , Rineke Verbrugge , Bart Verheij
Department of Artificial Intelligence, Bernoulli Institute,
University of Groningen, The Netherlands
{A.Keshavarzi.Zafarghandi, L.C.Verbrugge, Bart.Verheij}@rug.nl

Abstract

Abstract dialectical frameworks (ADFs) have been introduced as a formalism for modeling and evaluating argumentation allowing general logical satisfaction conditions. Different criteria that have been used to settle the acceptance of arguments are called semantics. However, the notion of semi-stable semantics as studied for abstract argumentation frameworks has received little attention for ADFs. In the current work, we present the concepts of semi-two-valued models and semi-stable models for ADFs. We show that these two notions satisfy a set of plausible properties required for semi-stable semantics of ADFs. Moreover, we show that semi-two-valued and semi-stable semantics of ADFs form a proper generalization of the semi-stable semantics of AFs, just like two-valued model and stable semantics for ADFs are generalizations of stable semantics for AFs.

1 Introduction

Formalisms of argumentation have been introduced to model and evaluate argumentation. Abstract argumentation frameworks (AFs) as introduced by Dung (1995) are a core formalism in formal argumentation. A popular line of research investigates extensions of Dung's AFs that allow for a richer syntax (see, e.g. Brewka, Polberg, and Woltran 2014).

In this work, we investigate a generalisation of Dung's AFs, namely, abstract dialectical frameworks (ADFs) (Brewka et al. 2018), which are known as an advanced abstract formalism for argumentation covering several generalizations of AFs (Brewka, Polberg, and Woltran 2014; Polberg 2017; Dvořák, Keshavarzi Zafarghandi, and Woltran 2020). This is accomplished by acceptance conditions which specify, for each argument, its relation to its neighbour arguments via propositional formulas. These conditions determine the links between the arguments which can be, in particular, attacking or supporting.

In formal argumentation one is interested in investigating 'How is it possible to evaluate arguments in a given formalism?' Answering this question leads to the introduction of several types of semantics. In AFs, one starts with selecting a set of arguments without any conflicts. Conflict-freeness is a main characteristic of all types of semantics of AFs. Very often a new semantics is an improvement of an already existing one by introducing further restrictions on the set of accepted arguments or possible attackers. A list of semantics of AFs is presented in (Dung 1995), namely conflict-free, admissible, complete, preferred, and stable semantics. Further semantics for AFs have been introduced later on, for instance, stage semantics (Verheij 1996), semi-stable semantics first in (Verheij 1996) (under a different name) then further investigated in (Caminada 2006), ideal semantics (Dung, Mancarella, and Toni 2007), and eager semantics (Caminada 2007b). Each semantics presents a point of view on accepting arguments.

Most of the semantics of AFs have been defined for ADFs and it has been shown that semantics of ADFs are generalizations of semantics of AFs (Brewka et al. 2018; Gaggl, Rudolph, and Straß 2021). In this work, we focus on semi-stable semantics for ADFs, in a way that follows the same idea of semi-stable semantics of AFs. To this end, we first present a weaker version of the two-valued models of ADFs, which we call semi-two-valued models. Then we define semi-stable models for ADFs as a special case of semi-two-valued models of a given ADF. The relation between semi-two-valued and semi-stable models is similar to the relation between two-valued and stable models for ADFs. The difference is that a stable model is chosen among two-valued models, however, a semi-stable model will be chosen among semi-two-valued models of a given ADF.

Some of the semantics have become popular in the domain of argumentation, such as grounded semantics, preferred semantics and stable semantics. Each AF has a unique grounded extension, and one or more preferred extensions. However, it is possible that an AF does not have any stable extension. Because of this shortcoming of stable semantics, in order to pick at least one set of arguments, preferred and grounded semantics become more popular in argumentation. In contrast, stable semantics still enjoys a strong support in logic programming (1988) and answer set programming (1991), since it is preferred to have no outcome as opposed to an imperfect one. On the one hand, in argumentation a grounded extension presents the least amount of information about the acceptance of arguments. That is, a grounded extension collects a set of arguments about which there is no doubt. In other words, the grounded extension of a given AF is very skeptical. On the other hand, it is possible that an AF has a stable extension but the set of preferred extensions and stable extensions are not equal.

To overcome this deficiency, semi-stable semantics have been introduced for AFs. Semi-stable semantics is a way of approximating stable semantics when a given AF does not have any stable extension. Key characteristics of semi-stable semantics in AFs are: 1. It is placed between stable semantics and preferred semantics; 2. If an AF has at least one stable extension, then the set of stable extensions and the set of semi-stable extensions coincide; 3. Each finite AF has at least one semi-stable extension.[1] Computational complexity of semi-stable semantics is studied in (Dunne and Caminada 2008). Furthermore, (Caminada 2007a) presents an algorithm to compute semi-stable semantics of AFs.

In this paper we propose a notion of semi-stable semantics for ADFs. First we discuss required properties for such a semantics in order to ensure that our notion is a proper generalization of the notion of semi-stable semantics for AFs. Then we define our notion of semi-stable semantics for ADFs and study its properties. It turns out that our notion fulfills the required properties presented in Section 1.1.

1.1 Requirements of Semi-Stable Semantics

For AFs, the property holds that a semi-stable extension is stable in the AF restricted to the arguments that have a truth value (accepted/rejected, in/out). This holds in general, and in particular also for AFs that have no stable extension. In the current work we follow this same idea to extend the notion of semi-stable semantics of AFs for ADFs.

In ADFs, the notion of stable model is defined based on the notion of two-valued model. An ADF may have no stable model. On the one hand, if a given ADF does not have any two-valued model, then it does not have any stable model. On the other hand, an ADF may have two-valued models, while none of them is a stable model. We focus on the first issue here. To define the notion of semi-stable semantics for ADFs, we follow the same method as for stable semantics of ADFs. That is, first we introduce the notion of semi-two-valued semantics. Subsequently, we pick semi-stable models among semi-two-valued models of a given ADF. A *semi-two-valued model* is a complete interpretation, that is, the number of arguments that are assigned to unknown is $\subseteq$-minimal among all complete interpretations. Further, a *semi-stable model* is a semi-two-valued model v that has a constructive proof for arguments that are assigned to $\mathbf{t}$ in v. We show that the semi-stable semantics/semi-two-valued model presented in this work will satisfy the following conditions, which are akin to the properties of the notion of semi-stable semantics of AFs.

1. A semi-stable/semi-two-valued model of a given ADF should maximize the union of the sets of the accepted and of the rejected/denied arguments among all complete interpretations, with respect to subset inclusion;
2. Each semi-stable/semi-two-valued model is a preferred interpretation;
3. Each stable model is a semi-stable/semi-two-valued model;
4. Each finite ADF has at least one semi-two-valued model;
5. If an ADF has a stable model, then the set of stable models coincides with the set of semi-stable models;
6. The notion of semi-stable/semi-two-valued semantics for ADFs is a proper generalization of semi-stable semantics for AFs.

[1](Verheij 2003, Example 5.8) shows that existence is not guaranteed for infinite AFs. See also (Caminada and Verheij 2010).

This paper is structured as follows. In Section 2, we present the relevant background of AFs and ADFs. Then, in Section 3, we present definitions of semi-two-valued/semi-stable semantics for ADFs. In this section, we show that the notion of semi-stable semantics and semi-two-valued semantics satisfy the required properties, items $1-5$, presented above in this section. Further, in Section 4 we show that the notion of semi-stable/semi-two-valued semantics of ADFs is a proper generalization of the concept of semi-stable semantics of AFs, cf. the 6th property. In Section 5, we present the conclusion of our work. Furthermore, we briefly discuss a related research, in particular, (Alcântara and Sá 2018) has also proposed a notion of semi-stable semantics for ADFs.

2 Background

2.1 Abstract Argumentation Frameworks

We recall the basic notions of Dung's abstract argumentation frameworks (AFs) (Dung 1995).

Definition 1. (Dung 1995) An abstract argumentation framework (AF) is a pair (A, R) in which A is a set of arguments and $R \subseteq A \times A$ is a binary relation representing attacks among arguments.

Let $F = (A, R)$ be an AF. For each $a, b \in A$, the relation $(a, b) \in R$ is used to represent that a is an argument attacking the argument b. An argument $a \in A$ is, on the other hand, defended by a set $S \subseteq A$ of arguments (alternatively, the argument is acceptable with respect to S) (in F) if for each argument $c \in A$, it holds that if $(c, a) \in R$, then there is a $s \in S$ such that $(s, c) \in R$ (s is called a defender of a).

Different semantics of AFs present which sets of arguments in a given AF can be accepted jointly. In (Dung 1995), extension-based semantics of AFs are presented; we recall them in Definition 2. An extension is a set of arguments of a given AFs. Set $S \subseteq A$ is called a *conflict-free* set (extension) (in F) if there is no pair $a, b \in S$ such that $(a, b) \in R$. The *characteristic function* $\Gamma_F : 2^A \mapsto 2^A$ is defined as $\Gamma_F(S) = \{a \mid a \text{ is defended by } S\}$.

Definition 2. Let $F = (A, R)$ be an AF. A set $S \in cf(F)$ is

- admissible in F if $S \subseteq \Gamma_F(S)$;
- preferred in F if S is $\subseteq$-maximal admissible;
- complete in F if $S = \Gamma_F(S)$;
- grounded in F if S is the $\subseteq$-least fixed point of $\Gamma_F(S)$;
- stable in F if $\forall a \in A \setminus S$: $\exists b \in S$ s.t. $(b, a) \in R$.

Let $\sigma \in \{cf, adm, grd, prf, com, stb\}$ be different semantics in the obvious manner. Now we define that the set of all σ extensions for an AF F is denoted by $\sigma(F)$.

Definition 3. (Caminada 2006) Let $F = (A, R)$ be an AF and let S be an extension of F. For $a \in A$, we write $a^+ = \{b \mid (a, b) \in R\}$ and $S^+ = \cup\{a^+ \mid a \in S\}$. Set S is called a semi-stable extension iff S is a complete extension where $S \cup S^+$ is maximal.

The set of semi-stable extensions of F is denoted by *semi-stb*(F). We recall some of the main properties of semi-stable semantics of AFs in Theorem 1.

Theorem 1. (Caminada 2006) Let $F = (A, R)$ be a finite AF, i.e., $|A|$ is finite, and let S be an extension of F.

- *semi-stb*$(F) \neq \emptyset$;
- if $S \in$ *semi-stb*(F), then $S \in$ *prf*(F);
- if $S \in$ *stb*(F), then $S \in$ *semi-stb*(F);
- if *stb*$(F) \neq \emptyset$, then *stb*$(F) =$ *semi-stb*(F).

2.2 Abstract Dialectical Frameworks

In the current section, we briefly restate some of the key concepts of abstract dialectical frameworks that are derived from those given in (Brewka et al. 2018; Brewka et al. 2017b; Brewka et al. 2013; Brewka and Woltran 2010).

Definition 4. An abstract dialectical framework (ADF) is a tuple $D = (A, L, C)$ where:

- A is a finite set of arguments (statements, positions), denoted by letters;
- $L \subseteq A \times A$ is a set of links among arguments;
- $C = \{\varphi_a\}_{a \in A}$ is a collection of propositional formulas over arguments, called acceptance conditions.

An ADF can be represented by a graph in which nodes indicate arguments and links show the relation among arguments. Each argument a in an ADF is labelled by a propositional formula, called acceptance condition, φ_a over $par(a)$ where $par(a) = \{b \mid (b, a) \in L\}$. The acceptance condition of each argument clarifies under which condition it can be accepted. Acceptance conditions indicate the set of links implicitly, thus, there is no need to explicitly present L in the example ADFs. We restrict to the finite setting, thereby excluding the complication mentioned in footnote 1.

An argument a is called an *initial argument* if $par(a) = \{\}$. A *three-valued interpretation* v (for D) is a function $v : A \mapsto \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$, that maps arguments to one of the three truth values true ($\mathbf{t}$), false ($\mathbf{f}$), or undecided ($\mathbf{u}$). Interpretation v is called *trivial*, and v is denoted by $v_{\mathbf{u}}$, if $v(a) = \mathbf{u}$ for each $a \in A$. Furthermore, v is called a two-valued interpretation if for each $a \in A$ either $v(a) = \mathbf{t}$ or $v(a) = \mathbf{f}$. For $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ we write $v^x = \{a \in A \mid v(a) = x\}$.

Truth values can be ordered via the information ordering relation $<_i$ given by $\mathbf{u} <_i \mathbf{t}$ and $\mathbf{u} <_i \mathbf{f}$ and no other pair of truth values are related by $<_i$. Relation $\leq_i$ is the reflexive closure of $<_i$. Let $\mathcal{V}$ be the set of all interpretations for an ADF D. Interpretations can be ordered via $\leq_i$ with respect to their information content, i.e., for $v, w \in \mathcal{V}$, if $w(a) \leq_i v(a)$ for each $a \in A$, it is said that v is an extension of w, denoted by $w \leq_i v$.

The *characteristic operator* Γ_D maps interpretations to interpretations. Given an interpretation v (for D), the partial valuation of φ_a by v, is $v(\varphi_a) = \varphi_a^v = \varphi_a[b/\top : v(b) = \mathbf{t}][b/\bot : v(b) = \mathbf{f}]$, for $b \in par(a)$.

Definition 5. Let D be an ADF and let v be an interpretation of D. Applying Γ_D on v leads to v' such that for each $a \in A$, v' is as follows:

$$v'(a) = \begin{cases} \mathbf{t} & \text{if } \varphi_a^v \text{ is irrefutable (i.e., } \varphi_a^v \text{ is a tautology) ,} \\ \mathbf{f} & \text{if } \varphi_a^v \text{ is unsatisfiable,} \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

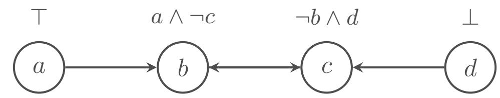


Figure 1: ADF of Example 1

The semantics of ADFs are defined via the characteristic operator as in Definition 6.

Definition 6. Given an ADF D, an interpretation v is:

- conflict-free in D iff $v(s) = \mathbf{t}$ implies φ_s^v is satisfiable and $v(s) = \mathbf{f}$ implies φ_s^v is unsatisfiable;
- admissible in D iff $v \leq_i \Gamma_D(v)$;
- preferred in D iff v is $\leq_i$-maximal admissible;
- complete in D iff $v = \Gamma_D(v)$;
- a (two-valued) model in D iff v is two-valued and $\Gamma_D(v) = v$;
- the grounded interpretation in D iff v is the least fixed point of Γ_D.

Let $\sigma(D)$ where $\sigma \in \{cf, adm, grd, prf, com, mod\}$ be the different semantics in the obvious manner. The set of all σ interpretations for an ADF D is denoted by $\sigma(D)$.

Example 1. An example of an ADF $D = (S, L, C)$ is shown in Figure 1. To each argument a propositional formula is associated, the acceptance condition of the argument. For instance, the acceptance condition of c, namely, $\varphi_c : \neg b \wedge d$, states that c can be accepted in an interpretation where b is denied and d is accepted. In D the interpretation $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}\}$ is conflict-free, since $\varphi_b^v = a \wedge \neg c$ is satisfiable. However, v is not an admissible interpretation, because $\Gamma_D(v) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$, that is, $v \not\leq_i \Gamma_D(v)$.

The interpretation $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$, on the other hand, is an admissible interpretation, since $\Gamma_D(v_1) = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$ and $v_1 \leq_i \Gamma_D(v_1)$. Moreover, in D the unique grounded interpretation $v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}\}$ is a preferred interpretation of D. In addition, *com*$(D) = \{v_2\}$.

The notion of stable semantics for ADFs is defined following similar ideas from logic programming. Stable models extend the concept of minimal model in logic programming by excluding self-justifying cycles of atoms. The concept of stable semantics of ADFs has been presented in (Brewka et al. 2013, Definition 6) and in (Brewka et al. 2017a, Definition 18); we recall it in Definition 7.

Definition 7. Let D be an ADF and let v be a two-valued model of D. Then v is a stable model of D if $v^{\mathbf{t}} = w^{\mathbf{t}}$, where w is the grounded interpretation of the *stb*-reduct $D^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}}$, $L^v = L \cap (A^v \times A^v)$, and $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$.

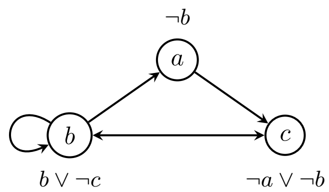
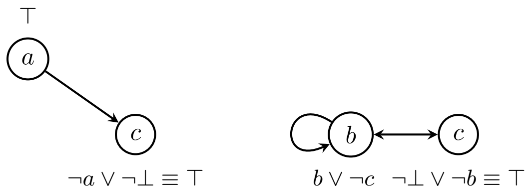


Figure 2: The ADF of Example 2.



***stb*-reduct** D^{v_1} ***stb*-reduct** D^{v_2}

Figure 3: The reduct of ADF D of Example 2.

Intuitively, the grounded interpretation collects all the information that is beyond any doubt, thus, it is said that there is a constructive proof for all arguments presented in the grounded interpretation. Hence, a two-valued model v of D is a stable interpretation (model), if there exists a constructive proof for all arguments assigned to true in v, in case all arguments that are assigned to false in v are actually false. Example 2 clarifies the notion of stable semantics of ADFs.

Example 2. Let $D = (\{a, b, c\}, \{\varphi_a : \neg b, \varphi_b : b \vee \neg c, \varphi_c : \neg a \vee \neg b\})$ be an ADF, depicted in Figure 2. D has two two-valued models, namely $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. We check whether they are stable models. To investigate whether v_1 is a stable model, first we evaluate the *stb*-reduct of D under v_1, namely $D^{v_1} = (A^{v_1}, L^{v_1}, C^{v_1})$. Here $A^{v_1} = \{a, c\}$, $L^{v_1} = \{(a, c)\}$, and $\varphi_a : \neg\bot \equiv \top$ and $\varphi_c : \neg a \vee \neg\bot \equiv \top$. The reduct D^{v_1} is depicted in Figure 3 (on the left). Since the unique grounded interpretation of D^{v_1} is $w = \{a \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} = v_1^{\mathbf{t}}$, we have $v_1 \in stb(D)$.

We show that v_2 is not a stable model of D. To this end, we first evaluate $D^{v_2} = (A^{v_2}, L^{v_2}, C^{v_2})$, where $A^{v_2} = \{b, c\}$, $L^{v_2} = \{(b, b), (b, c), (c, b)\}$, and $\varphi_b : b \vee \neg c$ and $\varphi_c : \neg\bot \vee \neg b \equiv \top$, depicted in Figure 3 (on the right). Since the unique grounded interpretation of D^{v_2} is $w = \{b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\}$, i.e., $w^{\mathbf{t}} \neq v_2^{\mathbf{t}}$, we have $v_2 \notin stb(D)$. Intuitively, it holds that $v_2 \notin stb(D)$, since in v_2 the acceptance of b depends on b itself, that is, there is a cyclic justification. Thus, v_2 violates the main condition of stable semantics that a stable model should have no self-justifying cycles of atoms.

An ADF may have no stable model. Example 3 presents an ADF that has a two-valued model, but no stable model.

Example 3. Let $D = (\{a, b, c\}, \{\varphi_a : c \vee b, \varphi_b : c, \varphi_c : a \leftrightarrow b\})$, depicted in Figure 4. The only two-valued model of D is $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$. However, $w^{\mathbf{t}} = \{\}$ where w is the grounded interpretation of D^v, where $D^v = D$. Thus, $w^{\mathbf{t}} \neq v^{\mathbf{t}}$. Hence, v is not a stable model of D.

Definition 8 presents the associated ADF for a given AF.

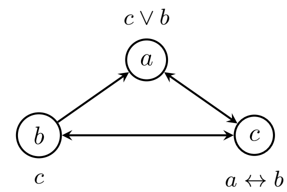


Figure 4: The ADF of Example 3

Definition 8. For an AF $F = (A, R)$, define the ADF associated to F as $D_F = (A, R, C)$ with $C = \{\varphi_a\}_{a \in A}$ such that for each $a \in A$ the acceptance condition is as follows:

$$\varphi_a = \bigwedge_{(b,a) \in R} \neg b$$

The semantics of ADFs generalize those of AFs. Since in AFs there is no direct support link, stable models (i.e., to avoid cyclic support among arguments) and models are equal in the associated ADF D_F of a given AF F.

3 Semi-Stable Semantics for ADFs

Before providing the formal definition of semi-stable semantics for ADFs, we present the intuition why an ADF may have no stable models. An ADF D may not have any stable model due to either of the following two reasons:

1. $mod(D) = \emptyset$, i.e., D does not have any two-valued models from which to pick a stable model; or,
2. $mod(D) \neq \emptyset$, but for any $v \in mod(D)$ it holds that $v \notin stb(D)$; that is, when there is no constructive proof for arguments that are assigned to $\mathbf{t}$ in v where $v \in mod(D)$.

Nonetheless, there are many cases about which one might want to draw a conclusion even when a given ADF does not have any two-valued model or stable model. One option is focusing on other semantics like preferred and grounded semantics that exist for any ADF. However, a unique grounded interpretation presents a piece of information about those arguments about which there is no doubt. That is, it is possible that in a given ADF the grounded interpretation has less information than each of its stable models. In other words, the information of the grounded interpretation is too skeptical. Furthermore, there exists an ADF D such that $stb(D) \neq \emptyset$ but the set of stable semantics of D and the set of preferred semantics of D are not equivalent, i.e., $stb(D) \subsetneq prf(D)$. That is, by preferred semantics some non-stable models may be introduced, even in the case that a stable model exists. Example 4 is an instance of an ADF such that $stb(D) \neq \emptyset$ but $stb(D) \subsetneq prf(D)$.

Example 4. Let $D = (\{a, b, c\}, \{\varphi_a : c \vee \neg b, \varphi_b : c \vee \neg a, \varphi_c : b \vee \neg a\})$ be an ADF, depicted in Figure 5. The set of preferred interpretations of D is $prf(D) = \{\{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. Both of the preferred interpretations of D are two-valued models of D. However, $stb(D) = \{\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. That is, $stb(D) \subsetneq prf(D)$.

Furthermore, the unique grounded interpretation of D is the trivial interpretation that has strictly less information than the stable model of D, i.e., $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$.

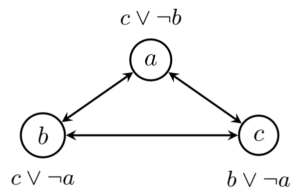


Figure 5: The ADF of Example 4

Still, it is interesting to present a semantics for ADFs that is equal to stable semantics if there exists a stable model. In ADFs, to define the notion of stable semantics as it is in (Brewka et al. 2018), first the notion of two-valued semantics is introduced. Then a two-valued model is called a stable model if it satisfies the conditions of Definition 7, i.e., if it does not contain any support cycle. Since in AFs there is no support cycle, these two notions are equal. That is, for the associated ADF D_F of a given AF F it holds that $mod(D_F) = stb(D_F)$. Due to this distinction between two-valued models and stable models in ADFs, different levels of semi-stable semantics can be considered in ADFs for the notion of semi-stable semantics of AFs. Here we follow a similar method as presented in (Brewka et al. 2018; Brewka et al. 2017b) for stable semantics to present the concept of semi-stable semantics.

The first reason that an ADF D does not have any stable semantics is that D does not have any two-valued model. We focus on this issue to present an alternative semantics for stable semantics of ADFs. In this alternative option, i.e., semi-stable semantics for ADFs, we are looking for a semi-two-valued model, which is a partially two-valued model, presented in Definition 9, that satisfies the condition of Definition 7, that is, it does not contain any support cycles among arguments. These new points of view of acceptance of arguments, which are called *semi-two-valued semantics* and *semi-stable semantics* of ADFs, have to satisfy the requirements presented in Section 1.1. The properties in Section 1.1 are akin to the properties of the notion of semi-stable semantics of AFs, presented in Theorem 1.

Definition 9. Let D be an ADF and let v be an interpretation of D. An interpretation v is a semi-two-valued model (interpretation) of D if the following conditions hold:

1. v is a complete interpretation of D;
2. $v^{\mathbf{u}}$ is $\subseteq$-minimal among all $w^{\mathbf{u}}$ such that w is a complete interpretation of D.

The set of semi-two-valued models of D is denoted by *semi-mod*(D). Note that when an ADF has a two-valued model, then the set of semi-two-valued models and the set of two-valued models coincide, which is shown in Lemma 1. We introduce the concept of semi-stable models over the notion of semi-two-valued models in Definition 10.

Definition 10. Let D be an ADF and let v be a semi-two-valued model of D. An interpretation v is a semi-stable model (interpretation) of D if the following condition holds:

- $v^{\mathbf{t}} = w^{\mathbf{t}}$ s.t w is the grounded interpretation of sub-reduct $D^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$, $L^v = L \cap (A^v \times A^v)$, and $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$.

The set of semi-stable models of D is denoted by *semi-stb*(D). Note that in Definition 10, in sub-reduct D^v we assume that v is a semi-two-valued model (complete interpretation) of D, however, in Definition 7, in sub-reduct D^v it is assumed that a given interpretation v is a two-valued model of D. Since in Definition 10, interpretation v is a semi-two-valued model, it may contain an argument that is assigned to $\mathbf{u}$. Therefore, in sub-reduct D^v in Definition 10, we keep those arguments that are assigned to $\mathbf{u}$ in v as well, i.e., $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$. Arguments that are assigned to $\mathbf{u}$ in v will remain in $\varphi_a[p/\bot : v(p) = \mathbf{f}]$ for each $a \in A^v$. Intuitively, a complete interpretation v is a semi-stable model of D if $v^{\mathbf{u}}$ is $\subseteq$-minimal among complete interpretations of D and there exists a constructive proof for arguments which are assigned to $\mathbf{t}$ in v, in case all arguments which are assigned to false in v are actually false. Corollary 1 is a direct result of Definition 3, which defines the notion of semi-stable model over the set of semi-two-valued models of a given ADF.

Corollary 1. *Let D be an ADF. Each semi-stable model of D is a semi-two-valued model of D.*

Example 5 clarifies the notion of semi-stable semantics of ADFs.

Example 5. Let $D = (\{a, b, c\}, \{\varphi_a : \neg a, \varphi_b : c \wedge (\neg a \vee c), \varphi_c : b \wedge (a \vee b)\})$ be an ADF, depicted in Figure 6. The set of preferred interpretations of D is *prf*$(D) = \{\{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}\}$. None of the preferred interpretations is a two-valued model. Thus, D does not have any stable model. Both $v_1 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ and $v_2 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}\}$ are complete interpretations of D. Furthermore, both v_1 and v_2 are semi-two-valued models of D, since $v_1^{\mathbf{u}} = v_2^{\mathbf{u}} = \{a\}$. However, we show that only v_2 is a semi-stable model of D. To this end, we first evaluate sub-reduct D^{v_2}. Since no argument is assigned to $\mathbf{t}$ in v_2 and only a is assigned to $\mathbf{u}$ in v, $A^{v_2} = \{a\}$. Thus, $D^{v_2} = (\{a\}, \{\varphi_a : \neg a\})$, depicted in Figure 7. It is clear that the unique grounded interpretation D^{v_2} is $w = \{a \mapsto \mathbf{u}\}$. Since $w^{\mathbf{t}} = v_2^{\mathbf{t}} = \emptyset$, it holds that v_2 is a semi-stable model of D.

On the other hand, v_1 is not a semi-stable model of D. In v_1, both b and c are assigned to $\mathbf{t}$ and a is assigned to $\mathbf{u}$, therefore, $A^{v_1} = A$. Since no argument is assigned to $\mathbf{f}$ in v_1, we have $D^{v_1} = D$. The grounded interpretation of D/D^{v_1} is $w = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}$. That is, $w^{\mathbf{t}} = \emptyset$. However, $v_1^{\mathbf{t}} = \{b, c\}$, i.e., $w^{\mathbf{t}} \neq v_1^{\mathbf{t}}$. Thus, v_1 is not a semi-stable model of D.

Proposition 1 shows the first property of semi-two-valued model presented in Section 1.1.

Proposition 1. *Let D be an ADF, and let v be a semi-two-valued model of D. It holds that v maximizes the union of the sets of the accepted and of the denied among all complete interpretations of D, i.e., $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is maximal with respect to subset inclusion.*

Proof. Let $D = (A, L, C)$ be a given ADF. Assume that v is a semi-two-valued model of D. Toward a contradiction,

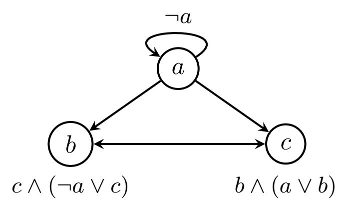


Figure 6: The ADF of Example 5

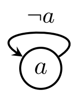


Figure 7: reduct D^{v_2} of ADF of Example 5

assume that $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is not $\subseteq$-maximal among all complete interpretations of D. Thus, there exists a complete interpretation w such that $w^{\mathbf{t}} \cup w^{\mathbf{f}}$ is $\subseteq$-maximal. Thus, it holds that $v^{\mathbf{t}} \cup v^{\mathbf{f}} \subsetneq w^{\mathbf{t}} \cup w^{\mathbf{f}}$. Hence, it holds that $w^{\mathbf{u}} \subseteq v^{\mathbf{u}}$, i.e., $v^{\mathbf{u}}$ is not minimal among complete interpretations of D. That is, by Definition 9, v is not a semi-two-valued model of D. This contradicts the assumption that v is a semi-two-valued model of D. Thus, the assumption that $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is not $\subseteq$-maximal among complete interpretations is wrong. $\square$

Proposition 1 clarifies the distinction between preferred semantics and semi-two-valued models of ADFs. While interpretation v is a preferred interpretation of D if it is $\leq_i$-maximal in *com*(D), interpretation v is a semi-two-valued model of D if $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal in *com*(D). Corollary 2 is a direct result of Proposition 1 and the fact that each semi-stable model is a semi-two-valued model.

Corollary 2. *Let D be an ADF, and let v be a semi-stable model of D. It holds that v maximizes the union of the sets of the accepted and of the denied among all complete interpretations of D, i.e., $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal in com(D).*

Theorem 2 presents the second and the third required properties for semi-stable/semi-two valued semantics for ADFs, presented in Section 1.1.

Theorem 2. Let D be an ADF.

1. Each semi-two-valued model of D is a preferred interpretation of D;
2. Each semi-stable model of D is a preferred interpretation of D;
3. Each stable model of D is a semi-two-valued model of D;
4. Each stable model of D is a semi-stable model of D.

Proof. Let D be an ADF.

- Proof of item 1: assume that v is a semi-two-valued model of D. We show that v is a preferred interpretation of D. Toward a contradiction, assume that $v \notin$ *prf*(D). By Definition 9, v is a complete interpretation of D. That is, if v is not a preferred interpretation, then there exists a preferred interpretation v' such that $v <_i v'$. Thus, $v'^{\mathbf{u}} \subsetneq v^{\mathbf{u}}$. Hence, by Definition 9, v is not a semi-two-valued model of D. This contradicts the assumption that v is a semi-two-valued model of D. Therefore, the assumption that v is not a preferred interpretation of D is wrong.
- Proof of item 2: assume that v is a semi-stable model of D. We show that v is a preferred interpretation of D. By Corollary 1, each semi-stable model of D is a semi-two-valued model of D. Thus, v is a semi-two-valued model of D. By the first item of this theorem, v is a preferred interpretation of D.
- Proof of item 3: Assume that v is a stable model. First, each stable model is a complete interpretation. Thus, the first item of Definition 9 is satisfied. Second, each stable model is a two-valued model, i.e., $v^{\mathbf{u}} = \emptyset$. Thus, $v^{\mathbf{u}}$ is $\subseteq$-minimal among all $w^{\mathbf{u}}$, where w is a complete interpretation of D. Hence, the second item of Definition 9 is satisfied. Thus, v is a semi-two-valued model of D.
- Proof of item 4: assume that v is a stable model. By the previous item, v is a semi-two-valued model of D. We show that v satisfies the condition of Definition 10. Since v is a stable model, by Definition 7, $v^{\mathbf{t}} = w^{\mathbf{t}}$ such that w is the grounded interpretation of sub-reduct $D^v = (A^v, L^v, C^v)$. Since $v^{\mathbf{u}} = \emptyset$, in Definition 10, $A^v = v^{\mathbf{t}}$. That is, Definition 10 (semi-stable model) and Definition 7 (stable-model) coincide for v. Thus, if v is a stable model of D, then v is a semi-stable model of D.

$\square$

The first two items of Theorem 2 imply that the set of semi-stable/semi-two-valued models of an ADF D is a subset of the set of preferred interpretations of D, i.e., *semi-stb*$(D) \subseteq$ *prf*(D) and *semi-mod*$(D) \subseteq$ *prf*(D). However, Proposition 2 indicates that the notion of preferred semantics coincides neither with the notion of semi-stable semantics nor with the notion of semi-two-valued semantics. That is, there exists an ADF D such that *prf*$(D) \nsubseteq$ *semi-stb*(D) and *prf*$(D) \nsubseteq$ *semi-mod*(D).

Proposition 2. *There is an ADF D such that the set of preferred interpretations of D does not coincide with the set of semi-stable models, nor with the set of semi-two-valued models of D.*

Proof. We show that there exists an ADF with a preferred interpretation which is not a semi-two-valued model. To this end, we use the ADF presented in ((Diller et al. 2020, Theorem 6)). Consider ADF $D = (\{a, b, c, d, e\}, \{\varphi_a : \neg c \wedge (\neg d \vee \neg b), \varphi_b : \neg a \wedge (\neg d \vee \neg c), \varphi_c : \neg b \wedge (\neg d \vee \neg a), \varphi_d : \neg e \wedge (\neg a \vee \neg b \vee \neg c), \varphi_e : \neg d\})$, depicted in Figure 8. D has four preferred interpretations, namely $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f}, c \mapsto \mathbf{t}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$, $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$, $v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$, and $v_4 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, d \mapsto \mathbf{f}, e \mapsto \mathbf{t}\}$. It holds that v_1, v_2, v_3 are semi-two-valued models/two-valued models of D, since $v_1^{\mathbf{u}} = v_2^{\mathbf{u}} = v_3^{\mathbf{u}} = \emptyset$. However, v_4 is not a semi-two-valued/semi-stable model, since $v_4^{\mathbf{u}} = \{a, b, c\}$, that is, $v_4^{\mathbf{u}}$ is not $\subseteq$-minimal among $v_i^{\mathbf{u}}$, for $1 \leq i \leq 4$. Thus, in ADFs, the notion of preferred semantics is not equal to the notion of semi-stable/semi-two-valued semantics.

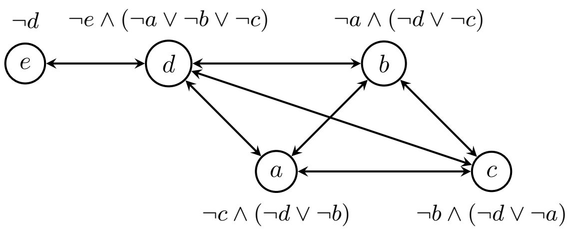


Figure 8: An ADF with a preferred interpretation that is not a semi-stable/semi-two-valued model

□

Proposition 3 presents the fourth property required for semi-two-valued semantics, presented in Section 1.1.

Proposition 3. *Each ADF has at least one semi-two-valued model.*

Proof. Let D be an ADF. Each ADF has a unique grounded interpretation. By the facts that the grounded interpretation is the least fixed point of Γ_D and the grounded interpretation is a least complete interpretation with respect to the $\leq_i$-ordering, we conclude that each ADF has at least one complete interpretation. By Definition 9, each semi-two-valued model v is a complete interpretation where $v^{\mathbf{u}}$ is $\subseteq$-minimal among other complete interpretations of D. Since the number of arguments is finite, the set of complete interpretations is finite. That is, there exists a complete interpretation v where $v^{\mathbf{u}}$ is $\subseteq$-minimal among all compete interpretations of D. Thus, the set of semi-two-valued models of D is non-empty. □

In Theorem 3, we show the fifth property of semi-stable semantics, presented in Section 1.1: If an ADF D has a stable model, then the set of stable models and the set of semi-stable models of D coincide. To show this theorem, we need some auxiliary results that are shown in Lemmas 1–3.

Lemma 1. *Let D be an ADF. Assume that D has a two-valued model. Then, the set of semi-two-valued models of D and the set of two-valued models of D coincide.*

Proof. Assume that D has a two-valued model v. Since v is a two-valued model, it holds that $v^{\mathbf{u}} = \emptyset$. Thus, by Definition 9, v is a semi-two-valued model, i.e., *mod*$(D) \subseteq$ *semi-mod*(D). It remains to show that every semi-two-valued model of D is also a two-valued model. Toward a contradiction, assume that D has a semi-two-valued model w which is not a two-valued model. Since w is a semi-two-valued but not a two-valued model, it holds that w is a complete interpretation and $w^{\mathbf{u}} \neq \emptyset$. However, since D has a two-valued model, $w^{\mathbf{u}}$ is not $\subseteq$-minimal among all complete interpretations of D. That is, by Definition 9, w is not a semi-two-valued model. This contradicts the assumption that w is a semi-two-valued model. That is, if D has a two-valued model, then *semi-mod*$(D) \subseteq$ *mod*(D). Hence, if ADF D has a two-valued model, then *semi-mod*$(D) =$ *mod*(D). □

Lemma 2. *Let D be an ADF. Assume that D has a stable model. Then, the set of semi-two-valued models of D and the set of two-valued models of D coincide.*

Proof. Let D be an ADF that has a stable model v. By the fact that each stable model of a given ADF is a two-valued model, it holds that v is a two-valued model. By Lemma 1, if there exists a two-valued model, then the set of two-valued models and the set of semi-two-valued models coincide. So if an ADF has a stable model, then *semi-mod*$(D) =$ *mod*(D). □

Lemma 3. *Let D be an ADF. Assume that D has a stable model. Then each semi-stable model of D is a two-valued model of D.*

Proof. Let D be an ADF that has at least one stable model v. By Lemma 2, the set of semi-two-valued models of D coincides with the set of two-valued models of D, i.e., *semi-mod*$(D) =$ *mod*(D). Moreover, by Corollary 1, each semi-stable model of D is a semi-two-valued model of D, i.e., *semi-stb*$(D) \subseteq$ *semi-mod*(D). Thus, if D has a stable model, then each semi-stable model of D is a two-valued model of D, i.e., *semi-stb*$(D) \subseteq$ *mod*(D). □

Theorem 3. If ADF D has a stable model, then the sets of stable models and semi-stable models of D coincide.

Proof. Let D be an ADF. By the forth item of Theorem 2, each stable model of D is a semi-stable model of D, i.e., *stb*$(D) \subseteq$ *semi-stb*(D).

Assume that D has a stable model v and a semi-stable model v'. We show that v' is a stable model of D. Toward a contradiction, assume that v' is not a stable model of D. By Lemma 3, v' is a two-valued model of D, i.e., $v'^{\mathbf{u}} = \emptyset$. If v' is not a stable model of D, by Definition 7, it has to be held that $v'^{\mathbf{t}} \neq w^{\mathbf{t}}$ where w is the grounded interpretation of the *stb*-reduct $D^{v'} = (A^{v'}, L^{v'}, C^{v'})$, where $A^{v'} = v'^{\mathbf{t}}$, $L^{v'} = L \cap (A^{v'} \times A^{v'})$, and $\varphi_a[p/\bot : v'(p) = \mathbf{f}]$ for each $a \in A^{v'}$. This implies that the condition of Definition 10 does not hold for v', since $v'^{\mathbf{u}} = \emptyset$. Thus, v' is not a semi-stable model of D. This is a contradiction by the assumption that v' is a semi-stable model of D. Hence, the assumption that D has a semi-stable model which is not a stable-model is wrong. Thus, if D has a stable model, then *semi-stb*$(D) \subseteq$ *stb*(D). Hence, if ADF D has a stable model, then *stb*$(D) =$ *semi-stb*(D).

□

Proposition 3 says that each ADF has at least one semi-two-valued model. In contrast, Proposition 4 shows that an ADF may have no semi-stable model. As we presented in the beginning of this section the notions of semi-two-valued semantics and semi-stable semantics of ADFs together fulfil the properties required for the concept of semi-stable semantics, presented in Section 1.1.

Proposition 4. *There exists an ADF that does not have any semi-stable model.*

Proof. Let D be the ADF presented in Example 3, i.e., $D = (\{a, b, c\}, \{\varphi_a : c \vee b, \varphi_b : c, \varphi_c : a \leftrightarrow b\})$. We showed, in Example 3, that $v = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}\}$ is a two-valued model of D, however, it is not a stable model of D. Thus, by Lemma 1, v is a semi-two-valued model of D. As we know $grd(D^v) = \{\emptyset\}$, however, $v^{\mathbf{t}} = \{a, b, c\}$. Thus, v is not a semi-stable model. □

Corollary 3. *Let D be an ADF that has a two-valued model. If none of the two-valued models D is a stable model of D, then D does not have any semi-stable model.*

Proof. Let D be an ADF that has a two-valued model. Thus, by Lemma 1 the set of two-valued models of D coincides with the set of semi-two-valued models of D. That is, for each two-valued model/semi-two-valued model v of D it holds that the condition of semi-stable model in Definition 10 coincides with the definition of stable model in Definition 7. Thus, if for each $v \in mod(D)$, v is not a stable model, then v is not a semi-stable model of D, as well. □

As Corollary 3 says, if an ADF has a two-valued model but no stable model, then it will not have any semi-stable model either. As we presented in the beginning of Section 3, the semi-stable semantics presented in this work deal with the first issue, namely, that an ADF may not have a stable model. That is, semi-stable semantics is a new point of view on the acceptance of arguments if an ADF does not have any two-valued model.

4 Generalization of the Semi-Stable Semantics of AFs

In this section, we show that the notions of semi-stable and semi-two-valued semantics for ADFs satisfy the last property presented in Section 1.1, required for these semantics. To this end, we show that the concept of semi-stable/semi-two-valued semantics for ADFs is a proper generalization of the concept of semi-stable semantics for AFs (Verheij 1996; Caminada 2006), in Theorems 4 and 5. Furthermore, we show that the concepts of semi-stable models and semi-two-valued models coincide for the associated ADF of a given AF, in Proposition 5.

Given an AF $F = (A, R)$ and its corresponding ADF $D_F = (A, R, C)$ (see Definition 8), the set of all possible conflict-free extensions of $F$is denoted by $\mathcal{E}$ and the set of all possible conflict-free interpretations of D_F is denoted by $\mathcal{V}$. The functions *Ext2Int*$_F$ and *Int2Ext*$_{D_F}$ in Definitions 11–12 are modifications of the labelling functions as given in (Baroni, Caminada, and Giacomin 2018). Function *Ext2Int*$_F(e)$ represents the interpretation associated to a given extension S in F, and function *Int2Ext*$_{D_F}(v)$ indicates the extension associated to a given interpretation v of D_F.

Definition 11. Let $F = (A, R)$ be an AF, and let S be an extension of F. The truth value assigned to each argument $a \in A$ by the three-valued interpretation v_S associated to S is given by *Ext2Int*$_F : \mathcal{E} \to \mathcal{V}$ as follows.

$$\mathit{Ext2Int}_F(S)(a) = \begin{cases} \mathbf{t} & \text{if } a \in S, \\ \mathbf{f} & \text{if } a \in S^+, \\ \mathbf{u} & \text{otherwise.} \end{cases}$$

It is shown in (Keshavarzi Zafarghandi, Verbrugge, and Verheij 2021, Proposition 20) that if S is a conflict-free extension of F, then *Ext2Int*$_F(S)$ is well-defined. Moreover, the basic condition that S has to be a conflict-free extension is a necessary condition for *Ext2Int*$_F(S)$ being well-defined. By Definition 3, every semi-stable extension of an AF is a complete extension and a conflict-free extension. Thus, if S is a semi-stable extension of AF F, then *Ext2Int*$_F(S)$ is well-defined. An interpretation of D_F can be represented as an extension via the function *Int2Ext*$_{D_F}$, presented in Definition 12.

Definition 12. Let $D_F = (A, R, C)$ be the ADF associated with AF $F = (A, R)$, and let v be an interpretation of D_F, that is, $v \in \mathcal{V}$. The associated extension S_v of v is obtained via application of *Int2Ext*$_{D_F} : \mathcal{V} \to \mathcal{E}$ on v, as follows:

$$\mathit{Int2Ext}_{D_F}(v) = \{s \in S \mid s \mapsto \mathbf{t} \in v\}$$

Theorem 4 presents that the notion of semi-two-valued model semantics for ADFs is a generalization of the concept of semi-stable semantics for AFs.

Theorem 4. For any AF $F = (A, R)$ and its associated ADF $D_F = (A, R, C)$, the following properties hold:

- if S is a semi-stable extension of F, then *Ext2Int*$_F(S)$ is a semi-two-valued model of D_F;
- if v is a semi-two-valued model of D_F, then *Int2Ext*$_{D_F}(v)$ is a semi-stable extension of F.

Proof. Let F be an AF and let D_F be its associated ADF, as in Definition 8.

- We assume that $\{S_0, S_1, \ldots, S_k\}$ is the set of all complete extensions of F. Since F is a finite AF, the set of complete extensions of F is finite. Assume that $\{v_0, v_1, \ldots, v_k\}$ is the set of corresponding complete interpretations of D_F, i.e., $v_i = \mathit{Ext2Int}_F(S_i)$ for i with $0 \leq i \leq k$.
  Without loss of generality, assume that S_0 is a semi-stable extension of F. By Definition 3, S_0 is a complete extension of F such that $S_0 \cup S_0^+$ is maximal. We show that $v_0 = \mathit{Ext2Int}_F(S_0)$ is a semi-two-valued model of D_F. Since v_0 is a complete interpretation of D_F, to show that v_0 is a semi-two-valued interpretation of D_F, it remains to show that $v_0^{\mathbf{u}}$ is $\subseteq$-minimal among all ${v_i}^{\mathbf{u}}$ for i with $0 < i \leq k$. Toward a contradiction, assume that ${v_0}^{\mathbf{u}}$ is not $\subseteq$-minimal among all ${v_i}^{\mathbf{u}}$ for i with $0 < i \leq k$. Thus, there exists a j for $0 < j \leq k$ such that ${v_j}^{\mathbf{u}} \subsetneq {v_0}^{\mathbf{u}}$. Thus, there exists an a such that $a \notin {v_j}^{\mathbf{u}}$ and $a \in {v_0}^{\mathbf{u}}$. Thus, by Definition 11, it holds that, for each such an a, $a \in S_j \cup {S_j}^+$ but $a \notin S_0 \cup {S_0}^+$. Thus, $S_0 \cup {S_0}^+$ is not maximal. This contradicts the assumption that S_0 is a semi-stable extension of F. Hence, v_0 is a semi-two-valued model of D_F.
- Assume that v is a semi-two-valued model of D_F; we show that $S = \mathit{Int2Ext}_{D_F}(v)$ is a semi-stable extension of F. To show that S is a semi-stable extension of F, we show that $S \cup S^+$ is maximal. Toward a contradiction, assume that $S \cup S^+$ is not maximal. Thus, there exists a complete extension of F, namely S', with $S' \cup S'^+$ is maximal, i.e., $S \cup S^+ \subsetneq S' \cup S'^+$. Thus, by Definition

11, it holds that $v \leq_i v'$, where $v' = \textit{Ext2Int}(S')$. Thus, v' is a complete interpretation of D_F such that $v'^{\mathbf{u}} \subsetneq v^{\mathbf{u}}$. Hence, v is not a semi-two-valued model of D_F. This contradicts the assumption that v is a semi-two-valued model of D_F. Thus, the assumption that $S \cup S^+$ is not maximal among all complete extensions of F is wrong. Hence, S is a semi-stable extension of F.

□

Proposition 5. *Let $F = (A, R)$ be an AF and let D_F be its associated ADF. The semi-two-valued semantics of D_F coincide with the semi-stable models of D_F.*

Proof. Let $F = (A, R)$ be an AF and let D_F be its associated ADF. By Corollary 1, $\textit{semi-stb}(D_F) \subseteq \textit{semi-mod}(D_F)$. Thus it remains to show that $\textit{semi-mod}(D_F) \subseteq \textit{semi-stb}(D_F)$.

Assume that v is a semi-two-valued model of D_F. To show that v is a semi-stable model of D_F, we show that $v^{\mathbf{t}} = w^{\mathbf{t}}$, where w is the grounded interpretation of sub-reduct $D_F^v = (A^v, L^v, C^v)$, where $A^v = v^{\mathbf{t}} \cup v^{\mathbf{u}}$. We show that $v^{\mathbf{t}} \subseteq w^{\mathbf{t}}$. Assume that $a \mapsto \mathbf{t} \in v$. Since D_F is an associated ADF to AF F, $\varphi_a : \bigwedge_{b \in par(a)} \neg b$. Thus, if $a \in v^{\mathbf{t}}$, then either a is an initial argument of D_F or for each $b \in par(a)$ it holds that $b \in v^{\mathbf{f}}$. In both cases, it is clear that $\varphi_a[p/\bot : v(p) = \mathbf{f}] \equiv \top$. Therefore, $a \in w^{\mathbf{t}}$. Thus, $v^{\mathbf{t}} = w^{\mathbf{t}}$. Hence, v is a semi-stable model of D_F. □

Theorem 5. For any AF $F = (A, R)$ and its associated ADF D_F, the following properties hold:

- if S is a semi-stable extension of F, then $\textit{Ext2Int}_F(S)$ is a semi-stable model of D_F;
- if v is a semi-stable model of D_F, then $\textit{Int2Ext}_{D_F}(v)$ is a semi-stable extension of F.

Proof. [Sketch] The theorem is a direct result of combining Theorem 4, which says that semi-two-valued semantics of ADFs are a generalization of semi-stable semantics of AFs, and Proposition 5, which says that in the associated ADF D_F of a given AF F, the notions of semi-stable semantics and semi-two-valued semantics coincide. □

5 Conclusion

In this work, we have defined the semi-stable and semi-two-valued semantics for finite ADFs. From a theoretical perspective, in Sections 3 and 4, we observe that the notions of semi-stable and semi-two-valued semantics for ADFs fulfil the requirements for these two notions presented in Section 1.1.

An ADF may have no stable model, for one of two reasons: 1. D does not have any two-valued model; or 2. each two-valued model contains a support cycle. The condition presented in Definition 7 characterizes the stable semantics for ADFs. The condition says that a two-valued model is stable if it does not contain any support cycle, i.e., if there exists a constructive proof for the arguments that are assigned to $\mathbf{t}$. Thus, to present an alternative definition for stable semantics we focus on the first reason that an ADF does not have a stable model, and we present a partial two-valued semantics in Section 3, called semi-two-valued semantics in Definition 9. Then we define the notion of semi-stable semantics over semi-two-valued semantics in Definition 10.

In Section 3, we show that the notions of semi-two-valued/semi-stable semantics of ADFs presented in this work satisfy the main requirements presented in Section 1.1. Specifically:

1. Proposition 1 and Corollary 2 say that if v is a semi-two-valued/semi-stable model of D, then $v^{\mathbf{t}} \cup v^{\mathbf{f}}$ is $\subseteq$-maximal among all complete interpretations of D.
2. Theorem 2 says that each semi-stable/semi-two-valued model is a preferred interpretation and each stable model of an ADF is a semi-stable/semi-two-valued model of that ADF.
3. Proposition 3 says that each ADF has at least one semi-two-valued model.
4. Theorem 3 says that if an ADF has a stable model, then the sets of stable models and semi-stable models coincide.

In Section 4, we show that the notions of semi-stable/semi-two-valued semantics of ADFs are proper generalizations of the notion of semi-stable semantics of AFs. In Proposition 5, we show that the concepts of semi-stable and semi-two-valued semantics coincide in the associated ADF of a given AF, intuitively, since in AFs there cannot be a support cycle.

Alcântara and Sá (2018) have also considered the semi-stable semantics for ADFs. To prevent confusion with the notion of semi-stable semantics presented in the current work, we call their notion semi-stable2 semantics, abbreviated SSS2. A key difference between our notion and SSS2 is that ours is compatible with the standard ADF definitions. In particular, in their discussion, the characteristic operator Γ_D and in addition, the semantics of ADFs, and specifically the complete semantics, have not been presented in the way as introduced by Brewka and Woltran (2018; 2010). For instance, by their deviating definition of complete labelling (Alcântara and Sá 2018), only $\{\neg a, \neg b\}$ is a complete labelling/grounded model of $D = (\{a, b\}, \{\varphi_a : b, \varphi_b : a\})$. Hence—unlike the standard definitions—the set of preferred labellings of D is in their approach not a subset of the set of complete labellings of D, and the unique grounded labelling $\{\}$ is not a complete labelling.

The computational complexity of semantics of AFs and ADFs is presented in (Dvořák and Dunne 2017). Computational complexity of semi-stable semantics of AFs is studied in (Dunne and Caminada 2008). As a future work, it would be interesting to clarify the computational complexity of investigating: 1. whether a given interpretation is a semi-stable/semi-two-valued model, 2. whether a given argument is credulously/skeptically acceptable/deniable under semi-stable/semi-two-valued semantics of a given ADFs.

Acknowledgments

This research is supported by the Center of Data Science & Systems Complexity (DSSC) Doctoral Programme, at Bernoulli Institute of the University of Groningen.

References

Alcântara, J. F. L., and Sá, S. 2018. On three-valued acceptance conditions of abstract dialectical frameworks. In *LSFA*, volume 344 of *Electronic Notes in Theoretical Computer Science*, 3–23. Elsevier.

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. *Handbook of Formal Argumentation* 1:157–234.

Brewka, G., and Woltran, S. 2010. Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)*, 102–111.

Brewka, G.; Strass, H.; Ellmauthaler, S.; Wallner, J. P.; and Woltran, S. 2013. Abstract dialectical frameworks revisited. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 2013)*, 803–809.

Brewka, G.; Diller, M.; Heissenberger, G.; Linsbichler, T.; and Woltran, S. 2017a. Solving advanced argumentation problems with answer-set programming. In *Conference on Artificial Intelligence AAAI*, 1077–1083. AAAI Press.

Brewka, G.; Ellmauthaler, S.; Strass, H.; Wallner, J. P.; and Woltran, S. 2017b. Abstract dialectical frameworks. An overview. *IFCoLog Journal of Logics and their Applications (FLAP)* 4(8):2263–2317.

Brewka, G.; Ellmauthaler, S.; Strass, H.; Wallner, J. P.; and Woltran, S. 2018. Abstract dialectical frameworks: An overview. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*. College Publications. chapter 5, 237—-285.

Brewka, G.; Polberg, S.; and Woltran, S. 2014. Generalizations of Dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems* 29(1):30–38.

Caminada, M., and Verheij, B. 2010. On the existence of semi-stable extensions. In *Proceedings of the 22nd Benelux Conference on Artificial Intelligence (BNAIC 2010)*. University of Luxemburg.

Caminada, M. 2006. Semi-stable semantics. In *COMMA*, volume 144 of *Frontiers in Artificial Intelligence and Applications*, 121–130. IOS Press.

Caminada, M. 2007a. An algorithm for computing semi-stable semantics. In *ECSQARU*, volume 4724 of *Lecture Notes in Computer Science*, 222–234. Springer.

Caminada, M. 2007b. Comparing two unique extension semantics for formal argumentation: Ideal and eager. In *Proceedings of the 19th Belgian-Dutch conference on artificial intelligence (BNAIC 2007)*, 81–87. Utrecht University Press.

Diller, M.; Keshavarzi Zafarghandi, A.; Linsbichler, T.; and Woltran, S. 2020. Investigating subclasses of abstract dialectical frameworks. *Argument & Computation* 11(1-2):191–219.

Dung, P. M.; Mancarella, P.; and Toni, F. 2007. Computing ideal sceptical argumentation. *Artificial Intelligence* 171(10-15):642–674.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77:321–357.

Dunne, P. E., and Caminada, M. 2008. Computational complexity of semi-stable semantics in abstract argumentation frameworks. In *JELIA*, volume 5293 of *Lecture Notes in Computer Science*, 153–165. Springer.

Dvořák, W., and Dunne, P. 2017. Computational problems in formal argumentation and their complexity. *Journal of Logics and their Applications* 4(8):2557–2622.

Dvořák, W.; Keshavarzi Zafarghandi, A.; and Woltran, S. 2020. Expressiveness of SETAFs and support-free ADFs under 3-valued semantics. In *COMMA*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, 191–202. IOS Press.

Gaggl, S. A.; Rudolph, S.; and Straß, H. 2021. On the decomposition of abstract dialectical frameworks and the complexity of naive-based semantics. *J. Artif. Intell. Res.* 70:1–64.

Gelfond, M., and Lifschitz, V. 1988. The stable model semantics for logic programming. In *ICLP/SLP*, 1070–1080. MIT Press.

Gelfond, M., and Lifschitz, V. 1991. Classical negation in logic programs and disjunctive databases. *New Generation of Computing* 9(3/4):365–386.

Keshavarzi Zafarghandi, A.; Verbrugge, R.; and Verheij, B. 2021. Strong admissibility for abstract dialectical frameworks. *Argument & Computation* (Under revision).

Polberg, S. 2017. *Developing the abstract dialectical framework*. Ph.D. Dissertation, TU Wien, Institute of Information Systems.

Verheij, B. 1996. Two approaches to dialectical argumentation: Admissible sets and argumentation stages. *Proc. NAIC* 96:357–368.

Verheij, B. 2003. DefLog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation* 13(3):319–346.