What Does it Take to Enforce an Argument? Minimal Change in Abstract Argumentation

Ringo Baumann[1]

Abstract. Argumentation is a dynamic process. The *enforcing problem* in argumentation, i.e. the question whether it is possible to modify a given argumentation framework (AF) in such a way that a *desired* set of arguments becomes an extension or a subset of an extension, was first studied in [3] and positively answered under certain conditions. In this paper, we take up this research and study the more general problem of *minimal change*. That is, in brief, i) is it possible to enforce a desired set of arguments, and if so, ii) what is the minimal number of modifications (additions or removals of attacks) to reach such an enforcement, the so-called *characteristic*. We show for several Dung semantics that this problem can be decided by local criteria encoded by the so-called *value functions*. Furthermore, we introduce the corresponding equivalence notions between two AFs which guarantee equal minimal efforts needed to enforce certain subsets, namely *minimal-E-equivalence* and the more general *minimal change equivalence*. We present characterization theorems for several Dung semantics and finally, we show the relations to standard and the recently proposed strong equivalence [9] for a whole range of semantics.

1 Introduction

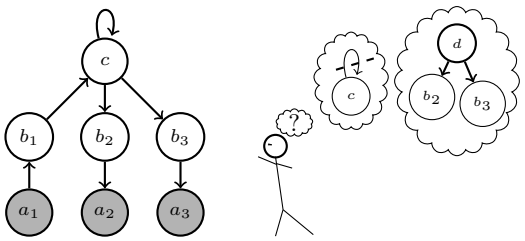
Argumentation theory is a vibrant research area in Artificial Intelligence, covering aspects of knowledge representation, multi-agent systems, and also philosophical questions (for a very good overview see [10]). Dung's abstract argumentation frameworks (AFs) [7] play a dominant role in this area. In AFs arguments and attacks between them are treated as primitives, i.e. the internal structure of arguments is not considered. The major focus is on resolving conflicts. To this end a variety of semantics have been defined, each of them specifying acceptable sets of arguments, so-called *extensions*, in a particular way.

Argumentation is an inherently dynamic process. It is, therefore, rather surprising that the research on the *dynamic behaviour* of abstract AFs has begun only quite recently. In [6] for instance, the impact of additional arguments on the outcome (extensions) of an AF was studied. Another representative work in this context is [9] where the notion of strong equivalence (which holds between two AFs, if they possess the same extensions w.r.t. all expansions) was introduced and characterized. The underlying non-monotonicity makes this field of research a non-trivial task.

In a recent paper, Baumann and Brewka [3] studied the so-called *enforcing problem*. Imagine a multi-agent scenario where, given an AF, an agent A wants another agent B to accept a particular set of arguments E. An enforcement, intuitively, is a modification of the AF such that E is contained in an extension of the modified AF. The paper specified necessary and sufficient conditions under which enforcements are possible.

In this paper we go an important step further. We are not only interested whether enforcements are *possible*, but also in the *effort needed* to enforce a set of arguments. We will use a numerical measure for this effort, roughly corresponding to the number of modifications needed to transform the given AF into an AF in which E is enforced.

Consider the following snapshot of a dialogue among agents A and B. Assume it is A's turn and his desired set of arguments is $E = \{a_1, a_2, a_3\}$. Furthermore, A and B are discussing under preferred semantics, which selects maximal conflict-free and self-defending sets of arguments.

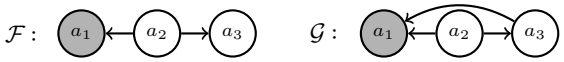


Agent A may come up with new arguments which interact with the old ones (adding attacks) and/or question old arguments or attacks between them, respectively (deleting attacks). The *minimal change problem* which we study in this paper can be formulated as follows: What is the minimal number of modifications (of a specific type) needed to enforce E, the so-called *characteristic* of E. In consideration of the infinite number of possibilities to modify a given argumentation scenario, the tricky thing about characteristics is to provide a local criterion which determines the minimal number. At the end of the paper we will have formally proven that for the example above this number equals 1 if we allow arbitrary modifications (for example through questioning the self-attack of c), 2 if the deletion of former attacks is forbidden (for example through introducing an argument which attacks b_2 and b_3), and ∞ (i.e. it is impossible to enforce E) if A only can come up with weaker arguments. These are fresh arguments which do not attack previous arguments.

The term *minimal change* traditionally concerns belief change, in particular belief revision (cf. the AGM approach [1]). Our investigation shares some common ground with this area, in particular the central role of theories or knowledge bases which are most similar to the initial one. Nevertheless, there are important differences. First of all, completely different formalisms are used, e.g. classical logic in AGM and AFs here. Secondly, while belief revision aims at incorporating new beliefs, we are interested in enforcing already given but not yet accepted sets of arguments through adding new information.

[1] Universität Leipzig, Johannisgasse 26, 04109 Leipzig, Germany email: baumann@informatik.uni-leipzig.de

A further interesting and important question is the mutual replaceability or similarity between AFs in the light of achieving goals (i.e. enforcing desired sets). Without doubt being able to identify such similarities among argumentation scenarios is a big advantage for an agent. For instance, he may safely accept more counterattacks to his arguments knowing that they do not make it more difficult at all to reach his goals. How to decide whether two AFs are mutually replaceable w.r.t. minimal changes? Consider the following AFs which might stem from modeling the same scenario but where the underlying notions of attack are different or, putting it less abstract, the attack (a_3, a_1) is in question.



The frameworks $\mathcal{F}$ and $\mathcal{G}$ are on a par in a static, non-dynamical sense since both possess the unique preferred extension $\{a_2\}$. If we assume that the agents desired set of arguments is $E = \{a_1\}$ the situation becomes different because although, $\mathcal{F}$ and $\mathcal{G}$ are equivalent in the standard sense, the effort needed to getting a_1 accepted is different. To be more precise, using the results presented in this paper it can be verified that the characteristic of E w.r.t. arbitrary expansions equals 1 in $\mathcal{F}$ and 2 in $\mathcal{G}$. This leads us towards novel dynamic notions of equivalence which guarantee equal minimal efforts needed to enforce certain subsets, namely *minimal-E-equivalence* and the more general *minimal change equivalence*.

Further motivations or theoretical and practical applications of our work are the following:

- *dialogue strategies*: Given an abstract argumentation scenario, assume that one agent reaches his goal if he enforces at least one of several desired sets of arguments. With the help of characteristics he may figure out which set is the closest one to being accepted in the current scenario. Furthermore, he may develop strategies which decrease or at least do not increase the characteristic.
- *monotonic vs. non-monotonic behaviour*: Argumentation is non-monotonic. Since every extension possesses a characteristic of zero, one may analyze and characterize dynamic scenarios where former extensions or at least accepted arguments survive, i.e. the characteristic remains zero. The characterization of monotonic parts contributes to a better understanding of the inherent non-monotonic formalism.
- *implicit vs. explicit information*: As demonstrated by the example above, knowledge about the extensions alone does not make explicit the implicit information w.r.t. minimal efforts needed to enforce certain subsets. In this sense minimal change equivalent AFs are insensitive to dynamics; in other words, they share the same implicit information.

Our main contributions can be summarized as follows:

- Formalizing and characterizing the minimal change problem for weak, strong, normal, arbitrary expansions as well as arbitrary modifications for the stable, preferred, complete and admissible semantics. The most remarkable result is that the characteristic does not change if we switch simultaneously between strong, normal or arbitrary expansions and preferred, complete or admissible semantics.
- For semi-stable semantics we show that its characteristic lies between the stable and preferred characteristic for all mentioned modification types.
- We introduce the notions of minimal-E-equivalence and minimal change equivalence between two AFs and provide characterizations in case of stable, preferred, complete and admissible semantics. Furthermore, we prove its relation to standard and strong equivalence for a whole range of semantics satisfying certain abstract principles. Interestingly, minimal change equivalence w.r.t. stable, semi-stable and preferred semantics implies standard equivalence, whereas this is not the case for complete and admissible semantics.

The paper is organized as follows. The Sect. 2 reviews the necessary definitions at work in abstract argumentation frameworks. We then introduce a pseudometric which lets us describe the minimal change problem formally. Sect. 4, the first main part of this work, contains the characterization theorems w.r.t. different classes of modifications. Sect. 5, the second main part, introduces the novel equivalence notions, provides characterization theorems and draws the relation to standard and strong equivalence. Finally, in Sect. 6 we discuss related results and present our conclusions.

2 Background

An *argumentation framework* $\mathcal{F}$ is a pair (A, R), where A is a non-empty finite set whose elements are called *arguments* and $R \subseteq A \times A$ a binary relation, called the *attack relation*. The set of all AFs is denoted by $\mathscr{A}$. If $(a, b) \in R$ holds we say that a *attacks* b, or b is *defeated* by a in $\mathcal{F}$. An argument $a \in A$ is *defended* by a set $A' \subseteq A$ in $\mathcal{F}$ if for each $b \in A$ with $(b, a) \in R$, b is defeated by some $a' \in A'$ in $\mathcal{F}$. Furthermore, we say that a set $A' \subseteq A$ is *conflict-free* in $\mathcal{F}$ if there are no arguments $a, b \in A'$ such that a attacks b. The set of all conflict-free sets of an AF $\mathcal{F}$ is denoted by $cf(\mathcal{F})$. We call an argument *isolated* if it neither attacks an argument in $\mathcal{F}$ nor is defeated by an argument in $\mathcal{F}$. For an AF $\mathcal{F} = (B, S)$ we use $A(\mathcal{F})$ to refer to B and $R(\mathcal{F})$ to refer to S. Finally, we introduce the union of two AFs as expected, namely $\mathcal{F} \cup \mathcal{G} = (A(\mathcal{F}) \cup A(\mathcal{G}), R(\mathcal{F}) \cup R(\mathcal{G})$.

2.1 Semantics

Semantics determine acceptable sets of arguments for a given AF $\mathcal{F}$, so-called *extensions*. The set of all extensions of $\mathcal{F}$ under semantics σ is denoted by $\mathcal{E}_\sigma(\mathcal{F})$. For two semantics σ, τ we use $\sigma \subseteq \tau$ to indicate that for any $\mathcal{F} \in \mathscr{A}$, $\mathcal{E}_\sigma(\mathcal{F}) \subseteq \mathcal{E}_\tau(\mathcal{F})$. We consider here stable, admissible, preferred, complete and semi-stable semantics [7, 5].

Definition 1 *Given an AF $\mathcal{F} = (A, R)$ and $E \subseteq A$. E is a*

1. *stable extension ($E \in \mathcal{E}_{st}(\mathcal{F})$) iff*
   $E \in cf(\mathcal{F})$ and each $a \in A \backslash E$ is defeated by some $e \in E$,
2. *admissible extension ($E \in \mathcal{E}_{ad}(\mathcal{F})$) iff*
   $E \in cf(\mathcal{F})$ and each $e \in E$ is defended by E in $\mathcal{F}$,
3. *preferred extension ($E \in \mathcal{E}_{pr}(\mathcal{F})$) iff*
   $E \in \mathcal{E}_{ad}(\mathcal{F})$ and for each $E' \in \mathcal{E}_{ad}(\mathcal{F})$, $E \not\subseteq E'$,
4. *complete extension ($E \in \mathcal{E}_{co}(\mathcal{F})$) iff*
   $E \in \mathcal{E}_{ad}(\mathcal{F})$ and for each $a \in A$ defended by E in $\mathcal{A}$, $a \in E$,
5. *semi-stable extension ($E \in \mathcal{E}_{ss}(\mathcal{F})$) iff*
   $E \in \mathcal{E}_{ad}(\mathcal{F})$ and for each $E' \in \mathcal{E}_{ad}(\mathcal{F})$, $R^+_{\mathcal{F}}(E) \not\subseteq R^+_{\mathcal{F}}(E')$ where $R^+_{\mathcal{F}}(E) = E \cup \{b \mid (a, b) \in R, a \in E\}$.

2.2 Expansions

We recap the definitions of several expansions, firstly introduced by [3]. These kinds of expansions will be our object of investigation since they represent reasonable types of dynamic argumentation scenarios. For short, normal expansions add new arguments and possibly new attacks which concern at least one of the fresh arguments. Strong (weak) expansions are normal and only add *strong (weak) arguments*,

i.e. the added arguments never are attacked by (attack) former arguments. The study of weak expansions may seem be more of an academic exercise than a task with practical relevance. Being aware of this fact, we want to emphasize that they might be relevant for Value Based AFs where former arguments may possess higher values than the further arguments and consequently cannot be attacked.

Definition 2 *An AF $\mathcal{F}^*$ is an expansion of AF $\mathcal{F} = (A, R)$ (for short, $\mathcal{F} \prec \mathcal{F}^*$) iff $\mathcal{F}^*$ has a representation as $(A \cup A^*, R \cup R^*)$, s.t. at least one of them (A^* or R^*) is not empty and $A^* \cap A = R^* \cap R = \varnothing$ holds. Such an expansion is called*

1. *normal ($\mathcal{F} \prec^N \mathcal{F}^*$) iff*
   $A^* \neq \varnothing \wedge \forall ab\ ((a,b) \in R^* \rightarrow a \in A^* \vee b \in A^*)$,
2. *strong ($\mathcal{F} \prec^N_S \mathcal{F}^*$) iff*
   $\mathcal{F} \prec^N \mathcal{F}^*$ *and* $\forall ab\ ((a,b) \in R^* \rightarrow \neg(a \in A \wedge b \in A^*))$,
3. *weak ($\mathcal{F} \prec^N_W \mathcal{F}^*$) iff*
   $\mathcal{F} \prec^N \mathcal{F}^*$ *and* $\forall ab\ ((a,b) \in R^* \rightarrow \neg(a \in A^* \wedge b \in A))$.

As usual we use $\mathcal{F} \preceq \mathcal{F}^*$ to indicate that the equality case is included, i.e. $\mathcal{F} \prec \mathcal{F}^* \vee \mathcal{F} = \mathcal{F}^*$ holds. The same applies to the specific kinds of expansions.

2.3 Strong and Standard Equivalence

The simplest kind of equivalence between two AFs is to have the same extensions. This term of equivalence is in correspondence with a non-dynamical, static argumentation scenario. All queries w.r.t. credulous or skeptical accepted arguments are answered identically. In this sense both are mutually replaceable.

Definition 3 *Two AFs $\mathcal{F}$ and $\mathcal{G}$ are equivalent to each other w.r.t. a semantics σ, in symbols $\mathcal{F} \equiv^\sigma \mathcal{G}$, iff they posses the same extensions under σ, i.e. $\mathcal{E}_\sigma(\mathcal{F}) = \mathcal{E}_\sigma(\mathcal{G})$ holds.*

Standard equivalence of two AFs is not sufficient for their mutual replaceability in dynamic argumentation scenarios. That means, possessing the same extensions does not guarantee to share the same acceptable sets of arguments w.r.t. all expansions. This kind of equivalence, so-called *strong equivalence*, was recently introduced and characterized by Oikarinen and Woltran [9] and is defined as follows.

Definition 4 *Two AFs $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent to each other w.r.t. a semantics σ, in symbols $\mathcal{F} \equiv^\sigma_\preceq \mathcal{G}$, iff for each AF $\mathcal{H}$, $\mathcal{F} \cup \mathcal{H} \equiv^\sigma \mathcal{G} \cup \mathcal{H}$ holds.*

3 The (σ, Φ)-characteristic

In [3] (Theorem 4) it is shown that whenever a set C is conflict-free in an AF $\mathcal{F}$ we may add one additional argument a and $n = |A(\mathcal{F}) \backslash C|$ attacks so that the union of C and $\{a\}$ is an extension of the constructed AF. Furthermore it is emphasized there, that in special cases the enforcement of a desired set C may be reached with less additional attack relations. This is exactly the question we have studied in this paper and which we call the *minimal change problem*. This means we are not only interested in whether enforcements are possible, but also in a minimal distance between the former argumentation scenario and the resulting one.

3.1 A pseudometric d on $\mathscr{A}$

To formalize the minimal change problem we have to introduce a numerical measure which indicates how far apart two argumentation scenarios are. We decided to count only added and removed attacks for a simple reason: the very nature of argumentation is the treatment of conflicting arguments. Adding or removing an isolated argument does not contribute at all to solving or increasing a given conflict, i.e. the conflicting information remains the same. This means, the decrease or increase of a conflict is directly linked to upcoming or disappearing attacks. The following definition takes this idea into account and can be formalized by the well-known symmetric difference $A \Delta B =_{def} (A \backslash B) \cup (B \backslash A)$.

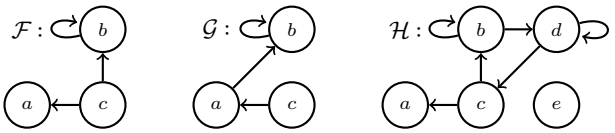
Definition 5 *The distance between two AFs $\mathcal{F}$ and $\mathcal{G}$ is a natural number defined by the following function*

$$d : \mathscr{A} \times \mathscr{A} \rightarrow \mathbb{N} \quad (\mathcal{F}, \mathcal{G}) \mapsto |R(\mathcal{F}) \,\Delta\, R(\mathcal{G})|\,.$$

The following proposition states that the class of all AFs $\mathscr{A}$ together with the above defined distance d constitute a pseudometric space. Remember that in pseudometric spaces two distinct elements may have distance zero.

Proposition 1 *$(\mathscr{A}, d)$ is a pseudometric space.*

The distances between the following three AFs are: $d(\mathcal{F}, \mathcal{G}) = 2$, $d(\mathcal{G}, \mathcal{H}) = 5$ and $d(\mathcal{F}, \mathcal{H}) = 3$.



3.2 The minimal change problem (formally)

Now we are prepared to describe the minimal change problem formally. We define therefore a function, the so-called (σ, Φ)-*characteristic*[2]. The σ indicates the considered semantics and Φ describes the possible modifications.

Definition 6 *Given a semantics σ, a binary relation $\Phi \subseteq \mathscr{A} \times \mathscr{A}$ and an AF $\mathcal{F}$. The (σ, Φ)-characteristic of a set $C \subseteq A(\mathcal{F})$ is a natural number or infinity defined by the following function*

$$N^{\mathcal{F}}_{\sigma,\Phi} : \wp(A(\mathcal{F})) \rightarrow \mathbb{N}_\infty$$

$$C \mapsto \begin{cases} 0, & \exists C' : C \subseteq C' \text{ and } C' \in \mathcal{E}_\sigma(\mathcal{F}) \\ k, & k = \min\{d(\mathcal{F}, \mathcal{G}) \mid (\mathcal{F}, \mathcal{G}) \in \Phi, N^{\mathcal{G}}_{\sigma,\Phi}(C) = 0\} \\ \infty, & \text{otherwise.} \end{cases}$$

Since d is a pseudometric there is an unlimited number of AFs $\mathcal{G}$ which possess distance k to a given AF $\mathcal{F}$. This means there is no brute-force-method to determine the characteristic of a given set $C \subseteq A(\mathcal{F})$.

The following proposition constitutes a first relation between different characteristics if certain subset properties between the considered semantics and/or the kinds of modifications are fulfilled.

Proposition 2 *Let σ, τ be semantics and Φ, Ψ binary relations over $\mathscr{A}$, s.t. $\sigma \subseteq \tau$ and $\Phi \subseteq \Psi$. For any AF $\mathcal{F}$, $N^{\mathcal{F}}_{\sigma,\Phi} \geq N^{\mathcal{F}}_{\tau,\Psi}$.*

It is well-known that the considered semantics satisfy the following subset relation, $st \subseteq ss \subseteq pr \subseteq co \subseteq ad$. Hence, we obtain the following corollary.

Corollary 3 *For any AF $\mathcal{F}$ and any binary relation Φ over $\mathscr{A}$,*

$$N^{\mathcal{F}}_{st,\Phi} \geq N^{\mathcal{F}}_{ss,\Phi} \geq N^{\mathcal{F}}_{pr,\Phi} \geq N^{\mathcal{F}}_{co,\Phi} \geq N^{\mathcal{F}}_{ad,\Phi}.$$

[2] If the semantics σ and the modification type Φ are clear from the context (or unimportant) we only refer to characteristic. The same holds for the later defined value functions and novel terms of equivalence.

A similar result may be obtained for different kinds of modifications (compare Def. 2). Note that strong and weak expansions are incomparable, i.e. there is no subset relation between them. $\mathcal{U}$ denotes the universal relation over $\mathscr{A}$, i.e. $\mathcal{U} = \mathscr{A} \times \mathscr{A}$.

Corollary 4 *For any AF $\mathcal{F}$ and any semantics σ,*

$$N^{\mathcal{F}}_{\sigma,\leq^N_W}, N^{\mathcal{F}}_{\sigma,\leq^N_S} \geq N^{\mathcal{F}}_{\sigma,\leq^N} \geq N^{\mathcal{F}}_{\sigma,\leq} \geq N^{\mathcal{F}}_{\sigma,\mathcal{U}}.$$

The following proposition strengthens the result of Corollary 4. It shows that the preferred, complete and admissible characteristics coincide.

Proposition 5 *For any $\mathcal{F}$ and any binary relation Φ over $\mathscr{A}$,*

$$N^{\mathcal{F}}_{st,\Phi} \geq N^{\mathcal{F}}_{ss,\Phi} \geq N^{\mathcal{F}}_{pr,\Phi} = N^{\mathcal{F}}_{co,\Phi} = N^{\mathcal{F}}_{ad,\Phi}.$$

4 Determining the characteristic: value functions

The characteristic is a precisely defined numerical value, but as yet we have not provided any means to actually compute this value. This is the question we address in the following three subsections. The significant challenge here is to define a function whose values are equal to the considered characteristics and, furthermore, whose values can be established in a **finite** number of steps based on rather simple properties of the underlying AF.

4.1 The $(\sigma, \leq^N_W)$-value

The first characteristics we are interested in are characteristics w.r.t. weak expansions. This means, what is the minimal number w.r.t. the distance d if only the addition of weaker arguments, i.e. arguments which do not attack previous arguments, is allowed. It turns out that there are only two possibilities, namely either a desired set C is already contained in an extension, i.e. the characteristic equals zero, or C is not enforceable, i.e. the characteristic equals infinity.

Definition 7 *Given an AF $\mathcal{F}$. The $(\sigma, \leq^N_W)$-value ($\sigma \in \{st, ad\}$) of a set $C \subseteq A(\mathcal{F})$ is zero or infinity defined by the following function*

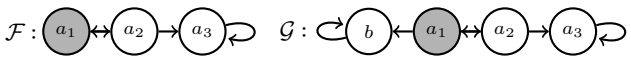
$$V^{\mathcal{F}}_{\sigma,\leq^N_W} : \wp(A(\mathcal{F})) \to \{0, \infty\}$$

$$C \mapsto \begin{cases} 0, & \exists C' : C \subseteq C' \text{ and } C' \in \mathcal{E}_\sigma(\mathcal{F}) \\ \infty, & \text{otherwise.} \end{cases}$$

Theorem 6 *For any AF $\mathcal{F}$ and any semantics $\sigma \in \{pr, co, ad\}$,*

$$N^{\mathcal{F}}_{st,\leq^N_W} = V^{\mathcal{F}}_{st,\leq^N_W} \text{ and } N^{\mathcal{F}}_{ad,\leq^N_W} = V^{\mathcal{F}}_{\sigma,\leq^N_W}.$$

Interestingly, semi-stable semantics does possess values between zero and infinity and is therefore not adequately characterized by the $(st, \leq^N_W)$ or $(ad, \leq^N_W)$ values. Consider the following AFs $\mathcal{F}$ and $\mathcal{G}$.



The $(st, \leq^N_W)$ - value of $\{a_1\}$ equals infinity since there are no supersets which are stable extensions in $\mathcal{F}$. In case of admissible semantics we deduce $V^{\mathcal{F}}_{ad,\leq^N_W}(\{a_1\}) = 0$ since $\{a_1\}$ is admissible in $\mathcal{F}$. Furthermore, $\{a_1\}$ and all its proper supersets are not semi-stable. This means, $N^{\mathcal{F}}_{ss,\leq^N_W}(\{a_1\}) \neq 0$. The weak expansion $\mathcal{G}$ of $\mathcal{F}$ shows that $N^{\mathcal{F}}_{ss,\leq^N_W}(\{a_1\}) \leq 2$ since $\{a_1\}$ is semi-stable in $\mathcal{G}$.

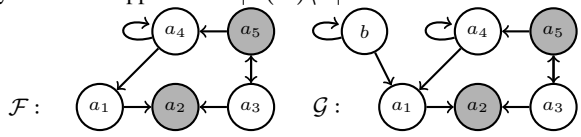
4.2 The $(\sigma, \leq^N_S)$-value

How does the situation change if we consider strong expansions? Since the class of strong expansions provides the possibility to attack or defend former arguments there should be a greater range of the characteristics than in case of weak expansions. We will see that this is indeed the case. Furthermore, we will show that the strong expansion characteristics are a lower bound for their corresponding weak expansion terms. Quite surprisingly, we were able to show that the characteristics of arbitrary expansions as well as normal expansions coincide with the strong expansion characteristic. This means, the additional feature of arbitrary extensions in contrast to strong expansions, namely bringing into play new attacks between existing arguments or new attacks from old argument to new arguments, is *useless* in the sense of achieving a lower value for the minimal change problem.

Now let's start with a first approximation for the characteristics w.r.t. strong expansions. As a by-product of Theorem 4 in [3] we state the following upper bound.

Corollary 7 *Given an AF $\mathcal{F}$ and a set $C \in cf(\mathcal{F})$. For any semantics $\sigma \in \{st, ss, pr, co, ad\}$, $\infty > |A(\mathcal{F})\backslash C| \geq N^{\mathcal{F}}_{\sigma,\leq^N_S}(C)$.*

This means, whenever a set C is conflict-free we may enforce C in finitely many steps. The following AFs $\mathcal{F}$ and $\mathcal{G}$ show that we may stay below the upper bound $|A(\mathcal{F})\backslash C|$.



Consider the set $C = \{a_2, a_4\}$. The upper bound $|A(\mathcal{F})\backslash C|$ equals 3. One may check that $C \cup \{b\} \in \mathcal{E}_{st}(\mathcal{G})$. Since $\mathcal{G}$ is a strong expansion of $\mathcal{F}$ and $d(\mathcal{F}, \mathcal{G}) = 1$ it follows that $N^{\mathcal{F}}_{st,\leq^N_S}(C) \leq 1$.

Inspired by the example above we give the following strong expansion values. We use $R^+_{\mathcal{F}}(C)$ for $C \cup \{b \mid \exists c \in C, (c, b) \in R(\mathcal{F})\}$ and analogously, $R^-_{\mathcal{F}}(C)$ for $C \cup \{b \mid \exists c \in C, (b, c) \in R(\mathcal{F})\}$.

Definition 8 *Given an AF $\mathcal{F}$. The $(\sigma, \leq^N_S)$-value ($\sigma \in \{st, ad\}$) of a set $C \subseteq A(\mathcal{F})$ is a natural number or infinity defined by the following function*

$$V^{\mathcal{F}}_{\sigma,\leq^N_S} : \wp(A(\mathcal{F})) \to \mathbb{N}_\infty$$

$$C \mapsto \min\left(\{|\sigma(\mathcal{F}, C')| \mid C \subseteq C' \text{ and } C' \in cf(\mathcal{F})\} \cup \{\infty\}\right),$$

where $ad(\mathcal{F}, C') = R^-_{\mathcal{F}}(C')\backslash R^+_{\mathcal{F}}(C')$ and
$st(\mathcal{F}, C') = A(\mathcal{F})\backslash R^+_{\mathcal{F}}(C')$.

As a direct consequence of the definition above we may state that the weak expansion value is greater than or equal to the strong expansion value.

Corollary 8 *For any AF $\mathcal{F}$ and any semantics $\sigma \in \{st, ad\}$, $V^{\mathcal{F}}_{\sigma,\leq^N_W} \geq V^{\mathcal{F}}_{\sigma,\leq^N_S}$.*

The following theorem shows that the stable value and admissible value adequately determine the strong expansion characteristics for stable or preferred, complete and admissible semantics, respectively. Furthermore, this characteristic does not vary if we shift from strong expansions to normal or strong expansions and vice versa.

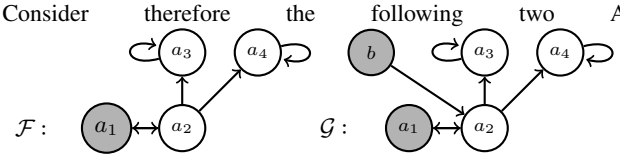
Theorem 9 *For any AF $\mathcal{F}$, any semantics $\sigma \in \{pr, co, ad\}$ and any $\Phi \in \{\leq, \leq^N, \leq^N_S\}$, $N^{\mathcal{F}}_{st,\Phi} = V^{\mathcal{F}}_{st,\leq^N_S}$ and $N^{\mathcal{F}}_{\sigma,\Phi} = V^{\mathcal{F}}_{ad,\leq^N_S}$.*

A valuable side-effect of the theorem above is that we have now clarified the relation between the classes of weak and strong expansions in case of stable, preferred, complete and admissible semantics. Remember that there is no subset relation between them since they are

completely different concepts. Nevertheless, the following proposition states that the weak expansion characteristic is greater than or equal to the strong expansion characteristic for the considered semantics. Here is the whole picture.

Proposition 10 *For any AF $\mathcal{F}$ and any semantics $\sigma \in \{st, pr, co, ad\}$, $N^{\mathcal{F}}_{\sigma,\leq^N_W} \geq N^{\mathcal{F}}_{\sigma,\leq^N_S} \geq N^{\mathcal{F}}_{\sigma,\leq^N} \geq N^{\mathcal{F}}_{\sigma,\leq} \geq N^{\mathcal{F}}_{\sigma,\mathcal{U}}$.*

Let's have a look at semi-stable semantics. Consider therefore the following two AFs.



The $(st, \leq^N_S)$ - value of $C = \{a_1\}$ in $\mathcal{F}$ equals 2 since first, a_3 and a_4 are unattacked by C and second, there is no conflict-free superset of C. In case of admissible semantics we have $R^+_{\mathcal{F}}(C) \backslash R^-_{\mathcal{F}}(C) = \varnothing$ and hence $V^{\mathcal{F}}_{ad,\leq^N_S}(\{a_1\}) = 0$. Since C is not semi-stable and does not have proper and conflict-free supersets in $\mathcal{F}$ we conclude $N^{\mathcal{F}}_{ss,\leq^N_W}(\{a_1\}) \neq 0$. Consider now the strong expansion $\mathcal{G}$ of $\mathcal{F}$. One may easily check that $C \cup \{b\}$ is semi-stable in $\mathcal{G}$. Hence, $N^{\mathcal{F}}_{ss,\leq^N_S}(C) = 1$. This means, neither the stable value nor the admissible value adequately determine the semi-stable characteristic w.r.t. strong expansions.

4.3 The $(\sigma,\mathcal{U})$-value

Now let's turn to arbitrary modifications, i.e. in contrast to arbitrary expansion we even allow deleting attacks between previous arguments. What consequences does this have for the minimal change problem? As a first and important difference we will show that any desired set of arguments may be enforced by a finite manipulation i.e. the $(\sigma,\mathcal{U})$-characteristic has to be finite for all considered semantics. The following proposition gives us an upper bound.

Proposition 11 *Given an AF $\mathcal{F}$ and a set $C \in A(\mathcal{F})$. For any semantics $\sigma \in \{st, ss, pr, co, ad\}$, $\infty > |R(\mathcal{F}) \cap (C \times C)| + |A(\mathcal{F}) \backslash C| \geq N^{\mathcal{F}}_{\sigma,\mathcal{U}}(C)$.*

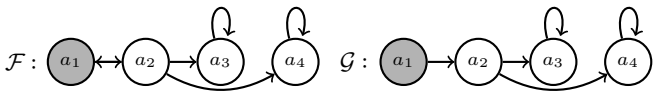
Inspired by the proposition above we present the following definitions which adequately determine the characteristics w.r.t. arbitrary modifications

Definition 9 *Given an AF $\mathcal{F}$. The $(\sigma,\mathcal{U})$-value ($\sigma \in \{st, ad\}$) of a set $C \subseteq A(\mathcal{F})$ is a natural number defined by the following function*

$$V^{\mathcal{F}}_{\sigma,\mathcal{U}} : \wp(A(\mathcal{F})) \to \mathbb{N}$$
$$C \mapsto \min\left(\{|R(\mathcal{F})_{\downarrow C'}| + |\sigma(\mathcal{F}, C')| \mid C \subseteq C' \subseteq A(\mathcal{F})\}\right),$$
where $R(\mathcal{F})_{\downarrow C'} = R(\mathcal{F}) \cap (C' \times C')$.

Theorem 12 *For any AF $\mathcal{F}$ and any semantics $\sigma \in \{pr, co, ad\}$, $N^{\mathcal{F}}_{st,\mathcal{U}} = V^{\mathcal{F}}_{st,\mathcal{U}}$ and $N^{\mathcal{F}}_{\sigma,\mathcal{U}} = V^{\mathcal{F}}_{ad,\mathcal{U}}$.*

The following AFs prove that the $(ss,\mathcal{U})$-characteristic is not adequately characterized by the $(st,\mathcal{U})$- or $(ad,\mathcal{U})$-values.



The $(st,\mathcal{U})$-value of $\{a_1\}$ equals 2 because first, $|R(\mathcal{F})_{\downarrow\{a_1,a_2\}}|+|st(\mathcal{F},\{a_1,a_2\})| = 2$ and second, there is no other superset C of $\{a_1\}$, s.t. $|R(\mathcal{F})_{\downarrow C}|+|st(\mathcal{F},C)| < 2$. In case of admissible semantics we deduce $V^{\mathcal{F}}_{ad,\leq^N_W}(\{a_1\}) = 0$ since $\{a_1\}$ is admissible in $\mathcal{F}$. On the other hand $N^{\mathcal{F}}_{ss,\mathcal{U}}(\{a_1\}) \neq 0$ since $\{a_1\}$ and all its proper supersets are not semi-stable. The AF $\mathcal{G}$ shows that $N^{\mathcal{F}}_{ss,\leq^N_W}(\{a_1\}) = 1$ since $\{a_1\}$ is semi-stable in $\mathcal{G}$ and $d(\mathcal{F},\mathcal{G}) = 1$.

4.4 Summary of Results

The following table gives a comprehensive overview over results presented in this section. Note that the values may shrink from above to below and from left to right.

$N^{\mathcal{F}}_{\sigma,\Phi}$	$\leq^N_W$	$\leq^N_S, \leq^N, \leq$	$\mathcal{U}$
st	$V^{\mathcal{F}}_{st,\leq^N_W}$	$V^{\mathcal{F}}_{st,\leq^N_S}$	$V^{\mathcal{F}}_{st,\mathcal{U}}$
pr, co, ad	$V^{\mathcal{F}}_{ad,\leq^N_W}$	$V^{\mathcal{F}}_{ad,\leq^N_S}$	$V^{\mathcal{F}}_{ad,\mathcal{U}}$

5 Minimal Change Equivalence

Strong equivalence between two AFs guarantees their mutual replaceability for *any* dynamic scenario without loss of information. In contrast to taking into account any dynamic scenario we are interested in an equivalence notion which corresponds to the very nature of a dispute where (counter)arguments are put forward with the objective to convince the participants of a certain opinion E. Two AFs are called *minimal-E-equivalent* if the minimal effort needed to enforce E is the same for both. If minimal-E-equivalence holds for every subset E of the AFs in question we call them *minimal change equivalent*. Here are the formal definitions.

Definition 10 *Two AFs $\mathcal{F}$ and $\mathcal{G}$ are*

1. minimal-E-equivalent *(in symbols: $\mathcal{F} \equiv^{\sigma,E}_{\Phi} \mathcal{G}$) or*
2. minimal change equivalent *(in symbols: $\mathcal{F} \equiv^{\sigma,MC}_{\Phi} \mathcal{G}$),*

w.r.t. a semantics σ and a binary relation $\Phi \subseteq \mathscr{A} \times \mathscr{A}$ iff

1. *$E \subseteq A(\mathcal{F})$ and $N^{\mathcal{F}}_{\sigma,\Phi}(E) = N^{\mathcal{G}}_{\sigma,\Phi}(E)$ and*
2. *for any E, s.t. $E \subseteq A(\mathcal{F})$ or $E \subseteq A(\mathcal{G})$, $N^{\mathcal{F}}_{\sigma,\Phi}(E) = N^{\mathcal{G}}_{\sigma,\Phi}(E)$.*

A few properties are clear by definition: First, minimal change equivalence guarantees sharing the same arguments (in contrast to minimal-E-equivalence) and second, minimal change equivalence implies minimal-E-equivalence for all subsets E.

In the following subsections we are interested in characterization theorems for the introduced equivalence notions as well as their relation to standard and strong equivalence.

5.1 Characterizing Minimal Change Equivalence

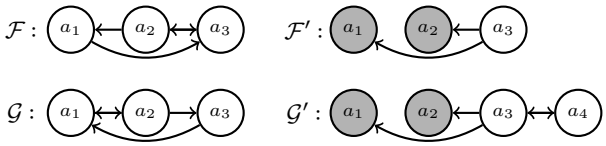
In Sect. 4 we proved that the introduced value-functions adequately determine the corresponding characteristics and, therefore, the minimal change problem. Now we take advantage of this work. The value-functions provide us with a procedure for deciding whether two AFs are minimal change equivalent or not since they are based on local criteria of the underlying AF and thus can be established in a finite number of steps.

Due to the limited space we only present a characterization for stable and admissible semantics w.r.t. weak and strong expansions as well as arbitrary modifications. Note that the characterizations w.r.t. the other semantics and modifications discussed in section 4 are implicitly given.

Theorem 13 *Let $\sigma \in \{st, ad\}$ and $\Phi \in \{\preceq_W^N, \preceq_S^N, \mathcal{U}\}$. For any AFs $\mathcal{F}$ and $\mathcal{G}$,*

1. *$\mathcal{F} \equiv_\Phi^{\sigma,E} \mathcal{G}$ iff $E \subseteq A(\mathcal{F})$ and $V_{\sigma,\Phi}^{\mathcal{F}}(E) = V_{\sigma,\Phi}^{\mathcal{G}}(E)$,*
2. *$\mathcal{F} \equiv_\Phi^{\sigma,MC} \mathcal{G}$ iff for any E, s.t. $E \subseteq A(\mathcal{F})$ or $E \subseteq A(\mathcal{G})$, $V_{\sigma,\Phi}^{\mathcal{F}}(E) = V_{\sigma,\Phi}^{\mathcal{G}}(E)$.*

The following AFs exemplify the novel equivalence notions.



The AFs $\mathcal{F}$ and $\mathcal{G}$ are minimal change equivalent w.r.t. preferred semantics and strong expansions, i.e. $\mathcal{F} \equiv_{\preceq_S^N}^{pr,MC} \mathcal{G}$. This can be seen as follows: First, $V_{pr,\preceq_S^N}^{\mathcal{F}}(\{a_1\}) = V_{pr,\preceq_S^N}^{\mathcal{G}}(\{a_1\}) = V_{pr,\preceq_S^N}^{\mathcal{F}}(\{a_3\}) = V_{pr,\preceq_S^N}^{\mathcal{G}}(\{a_3\}) = 1$, second, $V_{pr,\preceq_S^N}^{\mathcal{F}}(\{a_2\}) = V_{pr,\preceq_S^N}^{\mathcal{G}}(\{a_2\}) = V_{pr,\preceq_S^N}^{\mathcal{F}}(\emptyset) = V_{pr,\preceq_S^N}^{\mathcal{G}}(\emptyset) = 0$ and finally, all other sets $E \subseteq \{a_1, a_2, a_3\}$ do not possess conflict-free supersets in $\mathcal{F}$ or $\mathcal{G}$ and thus take the value infinity. Similarly, one can show that $\mathcal{F}' \equiv_{\preceq_S^N}^{pr,\{a_1,a_2\}} \mathcal{G}'$.

5.2 Relation to Strong and Standard Equivalence

In this subsection we study the relation between minimal change equivalence and strong or standard equivalence in general, i.e. rather than considering specific semantics we provide our results for a whole range of semantics satisfying certain abstract principles.

To get the first insight into the relations consider again the minimal change equivalent AFs $\mathcal{F}$ and $\mathcal{G}$. One main result in [9] is that strong equivalence is immediately linked to the existence of self-loops, in particular it collapses to syntactic equivalence in case of self-loop free AFs. This means, $\mathcal{F}$ and $\mathcal{G}$ are not strong equivalent which indicates that minimal change equivalence is possibly weaker than strong equivalence. The following theorem shows that the assertion holds for any semantics σ which satisfy that strong equivalent AFs have to share the same arguments. Let us call this property *regularity*.[3]

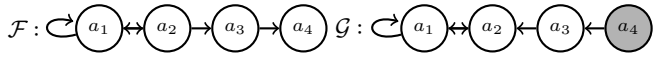
Theorem 14 *For any AFs $\mathcal{F}$, $\mathcal{G}$, any regular semantics σ and any binary $\Phi \in \{\preceq, \preceq^N, \preceq_S^N, \preceq_W^N\}$, $\mathcal{F} \equiv_\preceq^\sigma \mathcal{G} \Rightarrow \mathcal{F} \equiv_\Phi^{\sigma,MC} \mathcal{G}$.*

Since extensions possess a characteristic of zero, one might expect minimal change equivalence to imply standard equivalence. However, it turns out that this implication does not hold in general but can be shown for semantics satisfying *I-maximality* [2]. For short, I-maximality is fulfilled, if no extension can be a proper subset of another one.

Theorem 15 *For any AFs $\mathcal{F}$, $\mathcal{G}$, any semantics σ satisfying I-maximality and any binary $\Phi \in \{\mathcal{U}, \preceq, \preceq^N, \preceq_S^N, \preceq_W^N\}$, $\mathcal{F} \equiv_\Phi^{\sigma,MC} \mathcal{G} \Rightarrow \mathcal{F} \equiv^\sigma \mathcal{G}$.*

Remember that stable, semi-stable and preferred semantics satisfy I-maximality whereas complete and admissible does not. We want to conclude our study with two AFs showing that we cannot drop the I-maximality criterion in Theorem 15. It can be checked that $\mathcal{F} \equiv_\preceq^{co,MC} \mathcal{G}$. Furthermore, $\mathcal{F} \not\equiv^{co} \mathcal{G}$ since $\{a_4\} \in \mathcal{E}_{co}(\mathcal{G})$ and $\{a_4\} \notin \mathcal{E}_{co}(\mathcal{F})$.

[3] Note that regularity is a very weak criterion. One can imagine that for reasonable defined semantics it is always possible to find an AF $\mathcal{H}$, s.t. $\mathcal{E}_\sigma(\mathcal{F} \cup \mathcal{H}) \neq \mathcal{E}_\sigma(\mathcal{G} \cup \mathcal{H})$ if $A(\mathcal{F}) \neq A(\mathcal{G})$. All considered semantics in [9] satisfy regularity.



6 Related Work and Conclusions

We studied the minimal change problem in the context of Dung's abstract AFs. We defined the so-called characteristics which represent the problem formally. We then provided value functions which adequately determine the corresponding characteristics. Furthermore, we introduced and characterized minimal-E-equivalence and minimal change equivalence for certain classical semantics. Finally, we clarified the relation between minimal change and strong or standard equivalence w.r.t. semantics satisfying regularity and/or I-maximality.

There are a number of natural directions in which this research can be further pursued. For instance, a detailed classification of different kinds of changes w.r.t. the set of extensions if new information is added. A preliminary study (one argument, one interaction) is to be found in [6]. Since accepted arguments and therefore extensions possess a characteristic of zero a change of the outcome corresponds to a change of characteristics and thus can be analyzed by using the introduced value-functions. For the same reason, another important problem in dynamic argumentation is closely related to our study, namely the determination of the status of an argument. Since computing the justification state of an argument from scratch each time new information is added is very inefficient the possibility to *reuse* some previous computations would be a great advantage. Some results in this line of research (compare [3, 8]) show that reusing is possible if the considered semantics satisfy the *directionality criterion*, introduced in [2].

On a final note, our results and/or introduced techniques can contribute to solve open problems addressed in several works dealing with dynamics. In [4] for instance the authors say: "Moreover, ..., we are interested in questions about the change needed to change an argument from being accepted to rejected, or vice versa." or slightly different versions like in [6], "How to make the minimal change to a given argumentation framework so that it has a unique non-empty extension?".

REFERENCES

[1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson, 'On the logic of theory change: Partial meet contraction and revision functions', *J. Symb. Log.*, 510–530, (1985).

[2] P. Baroni and M. Giacomin, 'On principle-based evaluation of extension-based argumentation semantics.', *Artif. Intell.*, 675–700, (2007).

[3] R. Baumann and G. Brewka, 'Expanding argumentation frameworks: Enforcing and monotonicity results', in *COMMA 2010*, pp. 75–86, (2010).

[4] G. Boella, S. Kaci, and L. van der Torre, 'Dynamics in argumentation with single extensions: Abstraction principles and the grounded extension', in *ECSQARU 2009*, pp. 107–118, (2009).

[5] M. W. Caminada, 'Semi-stable semantics', in *COMMA 2006*, pp. 121–130, (2006).

[6] C. Cayrol, F. D. de Saint-Cyr, and M.-C. Lagasquie-Schiex, 'Revision of an argumentation system', in *KR 2008*, pp. 124–134, (2008).

[7] P. M. Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artifi. Intell.*, 321 – 357, (1995).

[8] B. Liao, L. Jin, and R.C. Koons, 'Dynamics of argumentation systems: A division-based method', *Artif. Intell.*, 1790–1814, (2011).

[9] E. Oikarinen and S. Woltran, 'Characterizing strong equivalence for argumentation frameworks', *Artif. Intell.*, 1985–2009, (2011).

[10] *Argumentation in Artificial Intelligence*, eds., I. Rahwan and G. R. Simari, Springer, 2009.