Complexity Properties of
Critical Sets of Arguments

Richard BOOTH [a] Martin CAMINADA [b] Paul E. DUNNE [c]
Mikołaj PODLASZEWSKI [a] Iyad RAHWAN [d]

[a] *University of Luxembourg*
[b] *Department of Computing Science, University of Aberdeen, UK*
[c] *Department of Computer Science, University of Liverpool, UK*
[d] *Masdar Institute of Science and Technology, UAE*

Abstract. In an abstract argumentation framework, there are often multiple plausible ways to evaluate (or *label*) the status of each argument as accepted, rejected, or undecided. But often there exists a *critical set* of arguments whose status is sufficient to determine uniquely the status of every other argument. Once an agent has decided its position on a critical set of arguments, then essentially the entire framework has been evaluated. Likewise, once a group, e.g. a jury, agree on the status of a critical set of arguments, all of their different views over all other arguments are resolved. Thus, critical sets of arguments are important both for efficient evaluation by individual agents and collective agreement by groups of such. To exploit this idea in practice, however, a number of computational questions must be considered. In particular, how much computational effort is needed to verify that a set is, indeed, a critical set or a *minimal* critical set. In this paper we determine exact bounds on the computational complexity of these and related questions. In addition we provide similar analyses of *issues*: a concept closely related to critical set and derived in terms of (equivalence) classes of argument related through "common" labelling behaviours.

Keywords. argumentation frameworks; labelling schemes; computational complexity.

Introduction

Labelling schemes have received increasing attention as a basis for analyzing semantic properties of Dung's seminal abstract model of argumentation [18] and its developments, e.g. [31,11,19]. Informally, the basic structures used in this approach are: a set of argument *labels*; criteria for determining whether an argument can (or must) be assigned a particular label and for distinguishing "valid" from "improper" labellings; and, in the context of labelling-based argumentation algorithms, criteria for determining whether a particular labelling is "terminal" or allows for the evolvement of further labellings.

Among the benefits offered by this approach is the potential to develop algorithmic schemes for standard decision and enumeration problems. Empirical studies have indicated that a number of the algorithms exploiting label-based techniques perform reasonably. Labelling-based algorithms and proof procedures have been described in the work of Verheij [32], Cayrol *et al.* [16], Caminada [6,9], Thang *et al.* [30], Nofal *et al.* [27]. In

contrast to the proven worst-case complexity classifications that have been demonstrated for the associated problems, see e.g. the survey of Dunne and Wooldridge [22], experimental studies indicate that these, often, deliver results within a feasible time. A further advantage of labelling formalisms, is that they allow for the specification of simple and straightforward discussion games to determine the status of a particular argument (acceptable or not w.r.t. a particular semantics[1]) [26,15,8,14]

In recent work, Booth *et al.* [3] raised questions concerning appropriate mechanisms by which to compare distinct labellings belonging to the same general class, i.e. the class of labellings coinciding with a particular argumentation semantics [1]. In discussing this question, Booth *et al.* [3] apply the concept of a *critical set* [24]. Informally a critical set of arguments with respect to a labelling-based semantics is one for which any valid labelling under this semantics uniquely determines which labels can be assigned to every other argument. Thus, in principle, by identifying "small" critical sets one has a method of finding all valid labellings and thereby determining labelling-based semantics quickly: find a small critical set, S say, and, having verified that S *is*, indeed, critical, one need only consider labellings of S to determine labellings of arguments not in S. Hence, identifying critical sets of arguments can be useful for the efficient evaluation of the entire argument graph.

Another application of the critical sets approach can be found in argument-based judgement aggregation. Suppose a group of evaluators (e.g. a jury) wishes to collectively label a given set of arguments presented to all of them (e.g. all evidence and arguments from the defense team and prosecution) [29,13]. If the group members agree on the labelling of a critical set of arguments, they will have resolved all of their different views over all other arguments. Thus, critical sets of arguments can facilitate efficient collective agreement by a group of agents.

Of course, there is one obvious obstacle facing such methods: the question of how much computational effort one needs to invest in order to identify a "minimal" critical set. It is this question which is the central topic of discussion in the current article.

We present formal background and definitions in the next section. In Section 2 we formulate precisely the decision problems relating to questions of interest. Section 3 presents our main technical results with conclusions and discussion offered in Section 4.

1. Notation and Definitions

The following concepts were introduced in Dung [18].

Definition 1 *An argumentation framework (*AF*) is a pair $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$, in which $\mathcal{X}$ is a finite set of arguments and $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{X}$ is the attack relationship for $\mathcal{H}$. A pair $\langle x, y \rangle \in \mathcal{A}$ is referred to as 'y is attacked by x' or 'x attacks y'. For R, S subsets of arguments in the* AF $\mathcal{H}(\mathcal{X}, \mathcal{A})$, *we say that $s \in S$ is attacked by R – written $attacks(R, s)$ – if there is some $r \in R$ such that $\langle r, s \rangle \in \mathcal{A}$. For subsets R and S of $\mathcal{X}$ we write $attacks(R, S)$ if there is some $s \in S$ for which $attacks(R, s)$ holds; $x \in \mathcal{X}$ is* acceptable with respect to S *if for every $y \in \mathcal{X}$ that attacks x there is some $z \in S$ that attacks y; S is* conflict-free *if no argument in S is attacked by any other argument in S. For $S \subseteq \mathcal{X}$, S^- (resp. S^+) denote*

the sets $\{ p \ : \ \exists\, q \in S \ \text{such that } \langle p,q \rangle \in \mathcal{A} \}$ (resp. $\{ p \ : \ \exists\, q \in S \ \text{such that } \langle q,p \rangle \in \mathcal{A} \}$). The characteristic function, $\mathcal{F} \ : \ 2^{\mathcal{X}} \to 2^{\mathcal{X}}$, *is defined as*

$$\mathcal{F}(S) \ = \ \{\, x \in \mathcal{X} \ : \ x \text{ is acceptable with respect to } S \,\}$$

A conflict-free set S is admissible *iff $S \subseteq F(S)$.*

1.1. Extension-based semantics for AFs

Based on the approach of a complete extension, one can then proceed to define various extension-based argumentation semantics. The idea is to define properties that a subset of $\mathcal{X}$ must satisfy in order to be considered justifiable. Thus if $\sigma \ : \ 2^{\mathcal{X}} \to \langle \top, \bot \rangle$ then the extensions of an AF, $\langle \mathcal{X}, \mathcal{A} \rangle$, with respect to σ (more concisely the σ-*extensions*) are denoted $\mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle)$ and formed by

$$\mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle) \ = \ \{\, S \subseteq \mathcal{X} \ : \ \sigma(S)\}$$

The next definition presents some widely studied choices for σ.

Definition 2 *Given $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$*

> a. $\mathcal{E}_{co}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \text{ is conflict-free} \wedge F(S) = S\}$
> b. $\mathcal{E}_{gr}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge \forall\, T \in \mathcal{E}_{co}(\mathcal{H}) \ \neg(T \subset S)\}$
> c. $\mathcal{E}_{pr}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge \forall\, T \in \mathcal{E}_{co}(\mathcal{H}) \ \neg(S \subset T)\}$
> d. $\mathcal{E}_{sst}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge \forall\, T \in \mathcal{E}_{co}(\mathcal{H}) \ \neg(S \cup S^+ \subset T \cup T^+)\}$
> e. $\mathcal{E}_{st}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge S \cup S^+ = \mathcal{X}\}$
> f. $\mathcal{E}_{id}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge S \subseteq \cap\mathcal{E}_{pr} \wedge \forall T \in \mathcal{E}_{co} \ \neg(T \subseteq \cap\mathcal{E}_{pr} \wedge S \subset T)\}$
> g. $\mathcal{E}_{eag}(\mathcal{H}) \ = \ \{S \subseteq \mathcal{X} \ : \ S \in \mathcal{E}_{co} \wedge S \subseteq \cap\mathcal{E}_{sst} \wedge \forall T \in \mathcal{E}_{co} \ \neg(T \subseteq \cap\mathcal{E}_{sst} \wedge S \subset T)\}$

These correspond in turn to: complete extensions (a); the grounded extension (b); preferred extensions (c); semi-stable extensions [31,11](d); stable extensions (e); the ideal extension [19](f); the eager extension [7] (g).

Please notice that Definition 2 specifies the argumentation semantics in a slightly different way than in [18,19,7] but equivalence is shown in [12,10]. The advantage of Definition 2 is that it emphasizes that the most common argumentation semantics are based on complete semantics. That is, they select among the complete extensions.

For a given semantics, σ, and its associated σ-extensions, $\mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle)$ a number of natural decision questions can be formulated. Thus, given $x \in \mathcal{X}$ or $S \subseteq \mathcal{X}$ we might ask:

> 1. Is $x \in S$ for *at least one* $S \in \mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle)$? (*Credulous Acceptance*, CA$_\sigma$)
> 2. Is $x \in S$ for *every* $S \in \mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle)$? (*Sceptical Acceptance*, SA$_\sigma$)
> 3. Is $S \in \mathcal{E}_\sigma(\langle \mathcal{X}, \mathcal{A} \rangle)$? (*Verification*, VER$_\sigma$)

1.2. Labelling semantics for AFs.

The other widely used approach for defining AF semantics, is that of applying argument *labellings*, as was pioneered by Pollock [28], Jakobovic and Vermeir [25] and Verheij [31]. The starting point for such schemes is a set of *labels*: for the purposes of our sub-

sequent presentation we use $\{I, O, U\}$ (corresponding to *I*n, *O*ut, *U*ndecided).[2] The key concept of interest in this paper is that of a *complete* labelling [5,12].

Definition 3 *A* labelling *of an* AF $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$ *is a function* $\lambda : \mathcal{X} \to \{I, O, U\}$. *A labelling is said to be a* complete labelling *iff for each* $x \in \mathcal{X}$ *it holds that:*

$$\lambda(x) = I \iff (\forall y : \langle y, x \rangle \in \mathcal{A} \Rightarrow \lambda(y) = O)$$
$$\lambda(x) = O \iff (\exists y : \langle y, x \rangle \in \mathcal{A} \wedge \lambda(y) = I)$$

If λ is a labelling then we write λ^I for $\{x \in \mathcal{X} : \lambda(x) = I\}$, λ^O for $\{x \in \mathcal{X} : \lambda(x) = O\}$ and λ^U for $\{x \in \mathcal{X} : \lambda(x) = U\}$. Furthermore, if λ_1 and λ_2 are labellings, we say that $\lambda_1 \sqsubseteq \lambda_2$ iff $\lambda_1^I \subseteq \lambda_2^I$ and $\lambda_1^O \subseteq \lambda_2^O$, and $\lambda_1 \sqsubset \lambda_2$ iff $\lambda_1 \sqsubseteq \lambda_2$ and $\lambda_1 \neq \lambda_2$. Also, if L is a set of labellings, we define $\sqcap L$ as $\{(x, I) : x \in \mathcal{X} \text{ and } \forall \lambda \in L, \lambda(x) = I\} \cup \{(x, O) : x \in \mathcal{X} \text{ and } \forall \lambda \in L, \lambda(x) = O\} \cup \{(x, U) : x \in \mathcal{X} \text{ and } \neg\forall \lambda \in L, \lambda(x) = I \text{ and } \neg\forall \lambda \in L, \lambda(x) = O\}$.

The next definition presents some common labelling-based argumentation semantics.

Definition 4 *Given* $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$

a. $L_{co}(\mathcal{H}) = \{\lambda : \lambda \text{ is a complete labelling of } \mathcal{H}\}$
b. $L_{gr}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \forall \lambda' \in L_{co}(\mathcal{H}) \neg(\lambda'^I \subset \lambda^I)\}$
c. $L_{pr}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \forall \lambda' \in L_{co}(\mathcal{H}) \neg(\lambda^I \subset \lambda'^I)\}$
d. $L_{sst}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \forall \lambda' \in L_{co}(\mathcal{H}) \neg(\lambda'^U \subset \lambda^U)\}$
e. $L_{st}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \lambda^U = \emptyset\}$
f. $L_{id}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \lambda \sqsubseteq \sqcap L_{pr} \wedge \forall \lambda' \in L_{co} \neg(\lambda' \sqsubseteq \sqcap L_{pr} \wedge \lambda \sqsubset \lambda')\}$
g. $L_{eag}(\mathcal{H}) = \{\lambda : \lambda \in L_{co} \wedge \lambda \sqsubseteq \sqcap L_{sst} \wedge \forall \lambda' \in L_{co} \neg(\lambda' \sqsubseteq \sqcap L_{sst} \wedge \lambda \sqsubset \lambda')\}$

These correspond in turn to: complete labellings (a); the grounded labelling (b); preferred labellings (c); semi-stable labellings (d); stable labellings (e); the ideal labelling (f); the eager labelling (g).

We first recall some well-known properties of argument labellings.

Fact 1 *Let* $\mathcal{H} = \langle \mathcal{X}, \mathcal{A} \rangle$ *be an* AF.

a. *If* λ *is a complete (resp. grounded, preferred, semi-stable, stable, ideal or eager) labelling of* $\mathcal{H}$ *then* λ^I *is a complete (resp. grounded, preferred, semi-stable, stable, ideal or eager) extension of* $\mathcal{H}$.
b. *If* S *is a complete (resp. grounded, preferred, semi-stable, stable, ideal or eager) extension of* $\mathcal{H}$ *then* λ *with* $\lambda^I = S$, $\lambda^O = S^+$ *and* $\lambda^U = \mathcal{X} \setminus (S \cup S^+)$ *is a complete (resp. grounded, preferred, semi-stable, stable, ideal or eager) labelling of* $\mathcal{H}$.

It has been observed in [1] that labellings and extensions are one-to-one related. In essence, an argument extension is simply the I-labelled part of an argument labelling.

In the remaining part of this paper, we will focus on complete labellings. We do so not only because these turn out to be the basis of the mainstream argumentation semantics (see Definition 4) but also because we aim to follow the approach of [3].

---

[2]In recent work on algorithmic techniques, e.g. Nofal *et al.* [27] additional labels have been proposed.

2. Critical argument sets and decision problems

A subset S of $\mathcal{X}$ is a *critical set* of $\mathcal{H}$ if,

$$\forall \langle \lambda_1, \lambda_2 \rangle \in L_{co}(\mathcal{H}) \times L_{co}(\mathcal{H}) \ (\lambda_1(S) = \lambda_2(S) \ \Rightarrow \ \lambda_1 = \lambda_2)$$

That is, S is a critical set of $\mathcal{H}$ if its labelling within any complete labelling uniquely determines the labelling of every argument in $\mathcal{X}$. Treating the concept of "S is a critical set in $\langle \mathcal{X}, \mathcal{A} \rangle$" as defining a collection of σ–extensions, we use $\mathcal{E}_{cs}$ to denote

$$\mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle) \ = \ \{ \ S \subseteq \mathcal{X} \ : \ S \text{ is a critical set of } \langle \mathcal{X}, \mathcal{A} \rangle \}$$

Recall that the *standard translation* of a CNF $\varphi(Z)$ with clauses $\{C_1, \ldots, C_m\}$ is the AF, $\mathcal{H}_\varphi(\mathcal{X}_\varphi, \mathcal{A}_\varphi)$ with arguments

$$\{ \ z_i, \neg z_i \ : \ z_i \in Z \} \ \cup \ \{C_1, \ldots, C_m\} \ \cup \ \{\varphi\}$$

and attack relation,

$$\begin{aligned}
&\{ \ \langle z_i, \neg z_i \rangle, \ \langle \neg z_i, z_i \rangle \ : \ z_i \in Z \} \cup \\
&\{ \ \langle z_i, C_j \rangle \ : \ z_i \text{ is a literal in clause } C_j \} \ \cup \\
&\{ \ \langle \neg z_i, C_j \rangle \ : \ \neg z_i \text{ is a literal in clause } C_j \} \ \cup \\
&\{ \ \langle C_j, \varphi \rangle \ : \ 1 \leq j \leq m \}
\end{aligned}$$

This AF is (with some very minor modifications) originally presented in work of Dimopolous and Torres [17] wherein the credulous acceptance problem with respect to admissibility semantics (CA_{adm}) was shown to be NP–complete. Specifically, we have

Fact 2 *For $\varphi(Z)$ a CNF formula and $\mathcal{H}_\varphi(\langle \mathcal{X}_\varphi, \mathcal{A}_\varphi \rangle)$ the AF defined by the standard translation of $\varphi(Z)$.*

$$\begin{aligned}
\text{CA}_{adm}(\mathcal{H}_\varphi, \varphi) \ &\Leftrightarrow \ \varphi(Z) \text{ is satisfiable} \\
&\Leftrightarrow \ \exists \, \lambda \in L_{co}(\mathcal{H}_\varphi) \ : \ \lambda(\varphi) = I
\end{aligned}$$

From $\{\mathcal{X}\} \in \mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle)$, the credulous acceptance problem, $\text{CA}_{cs}(\mathcal{H}, x)$, is trivial.

3. Complexity in critical set computations

We now consider the computational complexity of the *verification* problem VER_{cs} and related questions. We first observe that this, in common with similar questions concerning preferred and semi-stable extensions is unlikely to be computationally feasible.

Lemma 1 VER_{cs} *is coNP–complete.*

Proof: For membership in coNP, consider the complementary problem ($\neg\text{VER}_{cs}$) that accepts instances $\langle \mathcal{H}, S \rangle$ for which $S \notin \mathcal{E}_{cs}(\mathcal{H})$. That this is in NP follows by the algorithm which guesses labellings, $\langle \lambda_1, \lambda_2 \rangle$, verifies that these are both in $L_{co}(\mathcal{H})$, have $\lambda_1(S) = \lambda_2(S)$, but are distinct (i.e $\lambda_1 \neq \lambda_2$). All these verification steps being polynomial time decidable, it follows that $\neg\text{VER}_{cs} \in$ NP, thus VER_{cs} is in coNP.

For hardness we again use the complementary problem, showing this to be NP–hard by a reduction from CNF-SAT. Let $\varphi(Z)$ be an instance of CNF-SAT with $\mathcal{H}_\varphi = \langle \mathcal{X}_\varphi, \mathcal{A}_\varphi \rangle$ the AF given by the standard translation. The instance of $\neg\text{VER}_{cs}$ uses an AF, $\mathcal{F}_\varphi$, formed by adding three arguments – $\{y, \neg y, \psi\}$ to $\mathcal{X}_\varphi$, together with attacks

$$\{\langle \varphi, \psi \rangle, \ \langle \psi, y \rangle. \ \langle \psi, \neg y \rangle, \ \langle y, \neg y \rangle, \langle \neg y, y \rangle\}$$

The resulting AF is shown in Fig 1.
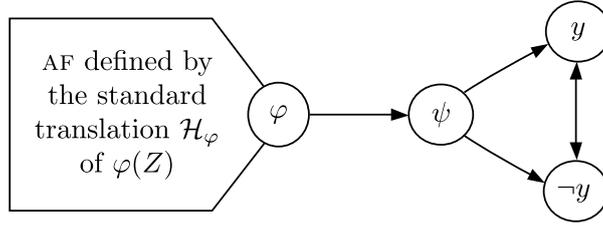
Figure 1. The AF, $\mathcal{F}_\varphi$ formed from the standard translation of $\varphi(Z)$.

Finally, S the candidate critical set is formed by $\mathcal{X}_\varphi \cup \{\psi\}$.

We claim that S is not a critical set (for this AF) if and only if $\varphi(Z)$ is satisfiable.

Suppose first that α_Z is an assignment of propositional values to Z for which $\varphi(\alpha_Z) = \top$, i.e that $\varphi(Z)$ is satisfiable. Consider the labelling of S in which $\lambda(x) = I$ if $x = z_i$ and $\alpha_i = \top$, or $x = \neg z_i$ and $\alpha_i = \bot$ or $x = \varphi$. For all other $x \in S$, $\lambda(x) = O$. This can be extended to a labelling λ_1 in which $\lambda_1(y) = I$, $\lambda_1(\neg y) = O$; and a labelling λ_2, with $\lambda_2(y) = O, \lambda_2(\neg y) = I$. Now both λ_1 and λ_2 are in L_{co}: from the properties of the standard translation described in Fact 2 and the fact that $\lambda(\psi) = O$. Hence we have found $\langle \lambda_1, \lambda_2 \rangle$ with $\lambda_1(S) = \lambda_2(S)$ but $\lambda_1 \neq \lambda_2$. Thus S is not a critical set.

Conversely, suppose that S is *not* a critical set. We show that, in this case $\varphi(Z)$ is satisfiable. Let $\langle \lambda_1, \lambda_2 \rangle$ be complete labellings witnessing that S is not critical and denote by λ the restriction of these to the arguments in S (noting that λ is well-defined since $\lambda_1(S) = \lambda_2(S)$ from the premise). Then we cannot have $\lambda(\psi) = I$, for then both y and $\neg y$ can only be labelled O. Similarly, if $\lambda(\psi) = U$ then both y and $\neg y$ must be labelled U (neither can be labelled I, since $\lambda(\psi) \neq O$). It follows that, since $\langle \lambda_1, \lambda_2 \rangle$ is a witness to S not being a critical set, we must have $\lambda(\psi) = O$, whence it follows that $\lambda(\varphi) = I$. It is immediate (from Fact 2 and the construction of $\mathcal{H}_\varphi$) that $\varphi(Z)$ is satisfiable. In total, S is a not a critical set if and only if $\varphi(Z)$ is satisfiable, from which it follows VER_{cs} is coNP–hard. □

The notion of critical set imposes quite strong conditions on the relationship between $S \in \mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle)$ and arguments in $\mathcal{X} \setminus S$: no matter how we label S within $\lambda \in L_{co}(\langle \mathcal{X}, \mathcal{A} \rangle)$ only one labelling is possible for $\mathcal{X} \setminus S$.

Rather than this "global" condition governing the relationship between S and *all* arguments outside S, suppose we refine this property by considering an equivalence relation between *individual* arguments. That is to say, the equivalence relation, $\equiv$ over $\mathcal{X} \times \mathcal{X}$ defined for $\mathcal{H}(\langle \mathcal{X}, \mathcal{A} \rangle)$ via $x \equiv y$ if and only if

1. $\forall \lambda \in L_{co}(\mathcal{H})\ \lambda(x) = \lambda(y)\ \lor$
2. $\forall \lambda \in L_{co}(\mathcal{H})\ (\lambda(x) = I \Leftrightarrow \lambda(y) = O) \land (\lambda(x) = O \Leftrightarrow \lambda(y) = I)$

That $\equiv$ *is* an equivalence relation is proved in Booth *et al.* [3, Propn. 6]. Now consider the (sets of arguments in) the equivalence classes of $\equiv$ which we refer to as the *issues* of $\langle \mathcal{X}, \mathcal{A} \rangle$ and let $\mathcal{E}_{issue}(\mathcal{H})$ denote

$$\{\ S \subseteq \mathcal{X}\ :\ \forall\ \langle x, y \rangle \in S \times S\ x \equiv y,\ \text{and}\ \forall T \supset S \exists \langle u, v \rangle \in T \times T\ \neg(u \equiv v)\}$$

Thus $S \in \mathcal{E}_{issue}(\mathcal{H})$ if and and only if S describes an equivalence class of $\mathcal{H}$ under the relation $\equiv$.

Let EQUIV denote the decision problem whose instances – $(\langle \mathcal{X}, \mathcal{A} \rangle, x, y)$ are accepted if and only if $x \equiv y$ with respect to complete labellings of $\langle \mathcal{X}, \mathcal{A} \rangle$ (similarly we use INEQUIV to denote the complementary problem).

Lemma 2 EQUIV *is co*NP*–complete.*

Proof: As with the previous lemma, the argument is couched in terms of the complementary problem so that we show INEQUIV to be NP–complete. That INEQUIV$\in$NP follows by noting, given an instance $(\langle \mathcal{X}, \mathcal{A} \rangle, x, y)$ that $\neg(x \equiv y)$ if and only if there are complete labellings – $\langle \lambda_1, \lambda_2 \rangle$ – for which $\lambda_1(x) \neq \lambda_1(y)$ under λ_1 (so that λ_1 witnesses $\langle x, y \rangle$ failing to satisfy condition 1); and λ_2, similarly, witnesses that $\langle x, y \rangle$ do not satisfy condition (2). We note that any labelling in which $\lambda(x) = \lambda(y) = I$ or $\lambda(x) = \lambda(y) = O$ suffices for the latter (although not one for which $\lambda(x) = \lambda(y) = U$).

Guessing two labellings $\langle \lambda_1, \lambda_2 \rangle$ and validating their properties can be accomplished by an NP algorithm.

To show that INEQUIV is NP–hard, we use a reduction from CNF-SAT. Given an instance, $\varphi(Z)$ of this, form exactly the same AF, $\mathcal{F}_\varphi$, described in the proof of Lemma 1. Within this AF, we consider the arguments, φ and y. We claim that $\neg(\varphi \equiv y)$ if and only if $\varphi(Z)$ is satisfiable.

Suppose that α_Z is an assignment for which $\varphi(\alpha_Z) = \top$. We, therefore find a labelling under which $\lambda(\varphi) = I$ forcing $\lambda(\psi) = O$. This labelling, however, is consistent with labellings λ_1 with $\lambda_1(\varphi) = I$ and $\lambda_1(y) = O$ (by using $\lambda_1(\neg y) = I$) and λ_2 with $\lambda_2(\varphi) = I$ and $\lambda_2(y) = I$ (by using $\lambda_2(\neg y) = O$). We deduce that λ_1 violates condition (1) and λ_2 condition (2) so that the satisfiability of φ implies $\neg(\varphi \equiv y)$.

Conversely suppose $\neg(\varphi \equiv y)$. Observing that $\neg(\varphi \equiv y)$ is witnessed by *two* complete labellings, $\langle \lambda_1, \lambda_2 \rangle$ it suffices to show that one of these allows the satisfiability of $\varphi(Z)$ to be inferred. Let λ_1 be a labelling under which $\lambda_1(\varphi) \neq \lambda_1(y)$. If $\lambda_1(\varphi) = U$ then, contradicting the premise, this forces $\lambda_1(y) = U$. If $\lambda_1(\varphi) = O$ then, again in contradiction, we get $\lambda_1(y) = O$ (since $\lambda_1(\psi) = I$). Hence $\lambda_1(\varphi) = I$ and we can choose $\lambda_1(y) = O$ (via $\lambda_1(\neg y) = I$). From the fact that $\lambda_1(\varphi) = I$ it is immediate that φ is satisfiable. □

The main structures we are interested in are *minimal* (wrt $\subseteq$) critical sets and *maximal* sets of equivalent arguments, i.e. equivalence classes (issues) under $\equiv$.

We now address the complexity of related verification questions, i.e.

MIN-CS
Instance: $(\langle \mathcal{X}, \mathcal{A} \rangle, S)$ with $S \subseteq \mathcal{X}$.
Question: Is $S \in \mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle)$ *but* no *strict* subset, T of S is in $\mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle)$?

ISSUE
Instance: $(\langle \mathcal{X}, \mathcal{A} \rangle, S)$ with $S \subseteq \mathcal{X}$.
Question: Is S an issue for $\langle \mathcal{X}, \mathcal{A} \rangle$, i.e. an equivalence class of $\equiv$ wrt complete labellings of $\langle \mathcal{X}, \mathcal{A} \rangle$?

We first establish upper bounds on the complexity of these. Recall that the complexity class D^p consists of languages, $\mathcal{L}$, that may be expressed in the form $\mathcal{L} = \mathcal{L}_1 \cap \mathcal{L}_2$, where $\mathcal{L}_1 \in \text{NP}$ and $\mathcal{L}_2 \in \text{CONP}$.

Lemma 3

    a. MIN-CS $\in \text{D}^p$.
    b. ISSUE $\in \text{D}^p$.

Proof: For part (a), consider the following two languages,

$$\mathcal{L}_1 = \{ (\langle \mathcal{X}, \mathcal{A} \rangle, S) : S \in \mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle) \}$$
$$\mathcal{L}_2 = \{ (\langle \mathcal{X}, \mathcal{A} \rangle, S) : \forall x \in S, \ S \setminus \{x\} \notin \mathcal{E}_{cs}(\langle \mathcal{X}, \mathcal{A} \rangle) \}$$

From Lemma 1, it is immediate that $\mathcal{L}_2 \in \text{NP}$ and $\mathcal{L}_1 \in \text{CONP}$.[3] We now have MIN-CS $= \mathcal{L}_1 \cap \mathcal{L}_2$ and, hence, in D^p.

For part (b), let $\mathcal{L}_1$ and $\mathcal{L}_2$ be given by,

$$\mathcal{L}_1 = \{ (\langle \mathcal{X}, \mathcal{A} \rangle, S) : \forall \langle x, y \rangle \in S \times S, \ x \equiv y \}$$
$$\mathcal{L}_2 = \{ (\langle \mathcal{X}, \mathcal{A} \rangle, S) : \forall \langle x, y \rangle \in S \times \mathcal{X} \setminus S, \ \neg(x \equiv y) \}$$

Again it is easily seen that ISSUE $= \mathcal{L}_1 \cap \mathcal{L}_2$ (noting, again, $(\langle \mathcal{X}, \mathcal{A} \rangle, S) \in \mathcal{L}_2$ does not indicate that $x \equiv y$ for every $\langle x, y \rangle \in S \times S$). To complete the proof, it suffices to show $\mathcal{L}_1 \in \text{CONP}$ and $\mathcal{L}_2 \in \text{NP}$. Since the complementary language to $\mathcal{L}_1$ is

$$\{ (\langle \mathcal{X}, \mathcal{A} \rangle, S) : \exists \langle x, y \rangle \in S \times S, \ \neg(x \equiv y) \}$$

a language in NP using witnesses of the form $(x, y, \langle \lambda_1, \lambda_2 \rangle)$ and the results of Lemma 2, it follows that $\mathcal{L}_1 \in \text{CONP}$. For $\mathcal{L}_2$, denoting $S = \{p_1, \ldots, p_r\}$ and $\mathcal{X} \setminus S = \{q_1, \ldots, q_t\}$ we need only guess a witness $\underline{w}$ of the form

$$\langle \lambda_1^{1,1}, \lambda_2^{1,1} \rangle \# \ \cdots \ \# \langle \lambda_1^{i,j}, \lambda_2^{i,j} \rangle \# \cdots \# \langle \lambda_1^{r,s}, \lambda_2^{r,s} \rangle$$

where $\langle \lambda_1^{i,j}, \lambda_2^{i,j} \rangle \in L_{co}(\langle \mathcal{X}, \mathcal{A} \rangle) \times L_{co}(\langle \mathcal{X}, \mathcal{A} \rangle)$ are complete labellings witnessing that $\neg(p_i \equiv q_j)$. The correctness of these labellings (of which there will at most $|S| \times |\mathcal{X} \setminus S| \leq |\mathcal{X}|^2/4$) can be validated in polynomial time. Hence, $\mathcal{L}_2 \in \text{NP}$, so completing the proof that ISSUE $\in \text{D}^p$.   □

We can now proceed to the main result of this section.

Theorem 1

    a. MIN-CS *is* D^p*–complete.*
    b. ISSUE *is* D^p*–complete.*

---

[3]Notice that $\mathcal{L}_2$ does *not* require S itself to be critical, simply that every subset obtained by removing a single argument of S fails to be critical.

Proof: Lemma 3 has already shown that both problems are in D^p so it remains only to show both are D^p–hard. For MIN-CS we proceed via a reduction from the canonical D^p–hard problem SAT-UNSAT, instances of which are a pair of CNF-formulae $\langle \varphi_1, \varphi_2 \rangle$ (without loss of generality over disjoint sets of propositional variables), such instances being accepted if and only if φ_1 is satisfiable *and* φ_2 is unsatisfiable.

Given an instance $\langle \varphi_1(Y), \varphi_2(Z) \rangle$ of SAT-UNSAT, form the AF, $\mathcal{H}_{\varphi_1, \varphi_2}$ consisting of AFs, $\mathcal{F}_1$ and $\mathcal{F}_2$ resulting from the translation presented in Lemma 1 applied respectively to $\varphi_1(Y)$ and $\varphi_2(Z)$. We use $\{\psi_1, p, \neg p\}$ for the arguments added (to the standard translation) in $\mathcal{F}_1$ and $\{\psi_2, q, \neg q\}$ for those added in $\mathcal{F}_2$. To complete the instance of MIN-CS, we set $S = \{y_1, \ldots, y_n, z_1, \ldots, z_n, p\}$.

We claim that S is a minimal critical set of $\mathcal{H}_{\varphi_1, \varphi_2}$ if and only if $\varphi_1(Y)$ is satisfiable *and* $\varphi_2(Z)$ is not satisfiable.

Suppose that $\varphi_1(Y)$ is satisfiable and $\varphi_2(Z)$ has no satisfying assignment. First observe that S is, *indeed* a critical set: given any labelling of S, this uniquely determines the labellings of $\{\neg y_i, \neg z_i : 1 \le i \le n\}$, and thence the labellings of each of the "clause" arguments in $\mathcal{F}_1$ and $\mathcal{F}_2$. In consequence the labellings of $\{\varphi_1, \varphi_2\}$ are fixed as well as the arguments $\{\psi_1, \psi_2, \neg p\}$. Finally from the premise that φ_2 is unsatisfiable the labellings of $\{q, \neg q\}$ are determined via similar arguments to those from Lemma 1. In addition to being critical, however, S is also a *minimal* such set: for $x \in \{y_1, \ldots, y_n, z_1, \ldots, z_n\}$, the set $S \setminus \{x\}$ fails to be critical since, we cannot uniquely determine the labellings of $\{x, \neg x\}$ from $S \setminus \{x\}$. Finally, since φ_1 is satisfiable, again via Lemma 1, we deduce that $S \setminus \{p\}$, cannot be a critical set. It follows that if $\langle \varphi_1, \varphi_2 \rangle$ is accepted as an instance of SAT-UNSAT then $\langle \mathcal{H}_{\varphi_1, \varphi_2}, S \rangle$ is accepted as an instance of MIN-CS.

Conversely, suppose that S is a minimal critical set of $\mathcal{H}_{\varphi_1, \varphi_2}$. If $\varphi_1(Y)$ is unsatisfiable, then this contradicts minimality since $S \setminus \{p\}$ remains critical. Similarly, if $\varphi_2(Z)$ is satisfiable, this contradicts S being critical, since there is a labelling of S that does not uniquely determine the labelling of $\{q, \neg q\}$. It follows that $\langle \mathcal{H}_{\varphi_1, \varphi_2}, S \rangle$ being a positive instance of MIN-CS implies that $\langle \varphi_1, \varphi_2 \rangle$ is accepted as an instance of SAT-UNSAT. This completes the proof that MIN-CS is D^p–complete.

For part (b), we also use a reduction from SAT-UNSAT, however, since arguments $\{x, y\}$ in separate frameworks are not, in general equivalent[4] some modification to the reduction is needed.

Given $\langle \varphi_1, \varphi_2 \rangle$, an instance of SAT-UNSAT the instance of ISSUE uses $\mathcal{H}_{\varphi_1, \varphi_2}$ as described in (a), but with an additional attack $\{\langle \varphi_2, \psi_1 \rangle\}$. This AF is shown in Fig. 2

The candidate issue, S, is $\{\varphi_2, \psi_2, q, \neg q\}$.

We claim that S is an issue of (the modified) $\mathcal{H}_{\varphi_1, \varphi_2}$, if and only if $\langle \varphi_1, \varphi_2 \rangle$ is accepted as an instance of SAT-UNSAT.

To begin, suppose that φ_1 is satisfiable but φ_2 is not. It is certainly the case that $S \subseteq T$ for *some* issue T, since as argued in the proof of Lemma 2, $\varphi_2 \equiv q$ and, trivially, $\varphi_2 \equiv \psi_2$. Noting, from the premise that for any complete labelling, λ of $\mathcal{H}_{\varphi_1, \varphi_2}$, the only

---

[4]With the trivial exception of arguments which can only be labelled U, e.g. the AF $\langle \{x, y\}, \{\langle x, x \rangle, \langle y, y \rangle\} \rangle$ whose only complete labelling is $\lambda(x) = \lambda(y) = U$ or when $\{x\}^- = \{y\}^- = \emptyset$.
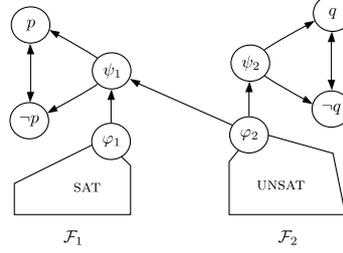
Figure 2. The AF, $\mathcal{H}_{\varphi,\varphi_2}$: ISSUE($\{\varphi_2, \psi_2, q, \neg q\}$) $\Leftrightarrow$ SAT-UNSAT($\langle \varphi_1, \varphi_2 \rangle$)

possibilities for $\lambda(\langle \varphi_2, \psi_2, q, \neg q \rangle)$ are $\{UUUU, OIOO\}$. This set, however, must also be maximal, i.e. $S = T$: we cannot add any clause argument to S, since we can always identify a labelling of Z (or Y) under which such arguments can be labelled either I or O. Similarly, we cannot add any literal $x \in \{y_i, \neg y_i, z_i, \neg z_i\}$ to S since, again we can always construct complete labellings having $\lambda(x) = O$ and $\lambda(x) = I$. Hence, if S is not an issue the only possibilities are from $\{\varphi_2 \equiv \varphi_1, \ \varphi_2 \equiv \psi_1, \ \varphi_2 \equiv p, \ \varphi_2 \equiv \neg p\}$. From the premise that $\varphi_1(Y)$ is satisfiable, it follows that there is a labelling with $\lambda(\varphi_1) = I$; trivially, however, there is also a labelling with $\lambda(\varphi_1) = O$ (since we can always arrange that some clause argument of $\mathcal{F}_1$ is labelled I). It now follows from $\varphi_1 \equiv \psi_1$ (from the fact that $\lambda(\varphi_2) \in \{U, O\}$) we cannot have $\varphi_2 \equiv \psi_1$. Finally since, as noted in the proof of Lemma 2, from any labelling under which $\lambda(\varphi_1) = I$, we can find labellings allowing $\lambda(p)$ to be either I or O (similarly, $\lambda(\neg p)$ to be either O or I) we deduce (via the satisfiability of φ_1) that S is an issue.

Conversely suppose that S is an issue. We wish to show that in this case, $\langle \varphi_1, \varphi_2 \rangle$ is accepted as an instance of SAT-UNSAT.

From the fact that S is an issue, it must be the case that $\neg(\varphi_1 \equiv \varphi_2)$ *and* $\varphi_2 \equiv q$. The second of these holds if and only if φ_2 is unsatisfiable as argued in Lemma 2. Consider the possible labellings for $\langle \varphi_1, \psi_1 \rangle$ (given that we have shown $\lambda(\varphi_2) \in \{U, O\}$). Since $\neg(\varphi_2 \equiv p)$ it cannot be the case that every complete labelling leads to $\lambda(\psi_1) \in \{I, U\}$, hence there must be some labelling under which $\lambda(\psi_1) = O$, i.e. either $\lambda(\varphi_1) = I$ or $\lambda(\varphi_2) = I$. The latter, as we have seen from the premise that S is an issue cannot occur, therefore such a labelling results in $\lambda(\varphi_1) = I$, hence φ_1 is satisfiable as required.

We deduce that $\langle \varphi_1, \varphi_2 \rangle$ is accepted as an instance of SAT-UNSAT if and only if $\{\varphi_2, \psi_2, q, \neg q\}$ is an issue of $\mathcal{H}_{\varphi_1,\varphi_2}$ and that ISSUE is D^p–complete. $\qquad\square$

For the argument that MIN-CS is D^p–complete we chose as the candidate minimal critical set structure to be verified a set $\{y_1, y_2, \ldots, y_n\} \cup \{z_1, z_2, \ldots, z_n\} \cup \{p\}$.

There is no need within the proof structure, however, to use arguments corresponding only to *positive* literals: that is to say exactly the same proof holds were S to be formed by, $\{\neg y_1, \neg y_2, \ldots, \neg y_n\} \cup \{\neg z_1, \neg z_2, \ldots, \neg z_n\} \cup \{p\}$. Using this observation the following consequence is immediate.

Corollary 1 *There are* AF*s,* $\langle \mathcal{X}, \mathcal{A} \rangle$ *with* $|\mathcal{X}| = n$ *and*

$$|\{S \subseteq \mathcal{X} \ : \ S \text{ is a minimal critical set in } \langle \mathcal{X}, \mathcal{A} \rangle\}| \ \geq \ 2^{n/3}$$

Proof: Consider any CNF, φ over, say, m variables, Z, and having exactly m clauses. The standard translation of φ to an AF has exactly $3m + 1$ arguments and any set S with exactly one argument from each $\{z_i, \neg z_i\}$ is a minimal critical set. □

4. Conclusions and discussion

We have studied the computational complexity of different decision problems centered around critical sets of arguments: sub-sets of arguments that, once labelled, uniquely determine the labels of all the other arguments in the argumentation framework. Also, we have examined the complexity of different decision problems related to the different issues [3] that can be identified.

The complexity classifications obtained are at a level typically viewed as intractable under the standard assumptions, namely coNP-complete and D^p-complete. It is noted, however, that this is at a similar level as a number of decision questions that have previously been studied in extension-based semantics of argumentation. For example, the questions of verifying a given subset as a preferred or semi-stable extension are both coNP-complete [17,11], as is the question of verifying if a set is an ideal *set* (that is a, not necessarily maximal, admissible subset contained in all preferred extensions) [20]. Indeed a number of common decision questions are well-known to involve rather higher levels of complexity: e.g. sceptical acceptability under both preferred and semi-stable semantics [21,23]; the verification problem for ideal extensions (that is, maximal ideal sets) [20]. From such perspectives, just as efforts to identify both tractable fragments and reasonable heuristics continue with regard to Dung-style extension based models, so too, similar investigation of techniques for identifying minimal (or "near" minimal) critical sets are well motivated. This is especially the case, given the gains (with respect to, among others, enumeration of labellings in a given class) that the formal structure of critical sets offers.

As a final point we mention that the notions of (minimal) critical sets and issues are related to specific argumentation semantics. So far, these have only been defined in the context of complete semantics [3]. It would, however, be equally possible to define them in terms of preferred, semi-stable or stable semantics. As an example how critical sets and issues change when the semantics is changed, consider the argumentation framework of Figure 3. Here, $\{E\}$ is a critical set under semi-stable and stable semantics, but not a critical set under complete and preferred semantics. One of the open research challenges is to broaden the notions of critical sets and issues of [3] also to work under different semantics than complete, and to examine how this affects the complexity of the associated decision problems.
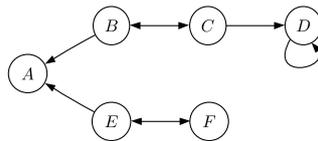

Figure 3. Different semantics yield different critical sets and issues.

References

[1] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.

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

[3] R. Booth, M. Caminada, M. Podlaszewski, and I. Rahwan. Quantifying disagreement in argument-based reasoning. In *Proc. 11th AAMAS*, pages 493–500, 2012.

[4] M. Caminada. Semi-stable semantics. In *Proc. 1st COMMA*, pages 121–128, 2006.

[5] M. Caminada. On the issue of reinstatement in argumentation. *Proc. 10th JELIA*, pages 111–123, 2006.

[6] M. Caminada. An algorithm for computing semi-stable semantics. *Proc. ECSQARU*, 222–234, 2007.

[7] M. Caminada. Comparing two unique extension semantics for formal argumentation: ideal and eager. In *Proc. 19th BNAIC*, pages 81–87, 2007.

[8] M. Caminada. Preferred semantics as socratic discussion. *Proc. 11th AI*IA*, pages 209–216, 2010.

[9] M. Caminada. An algorithm for stage semantics. In *Proc. 3rd COMMA*, pages 147–158, 2010.

[10] M. Caminada. A labelling approach for ideal and stage semantics. *Argument and Comp.*, 2:1–21, 2011.

[11] M. Caminada, W.A. Carnielli, and P.E. Dunne. Semi-stable semantics. *Jnl. of Logic and Computation*, 22(5):1207–1254, 2012.

[12] M. Caminada and D.M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.

[13] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Proc. 10th AAMAS*, 22(1):64–102, 2011.

[14] M. Caminada and M. Podlaszewski. Grounded semantics as persuasion dialogue. In *Proc. 4th COMMA*, pages 478–485, 2012.

[15] M. Caminada and Y. Wu. An argument game of stable semantics. *Logic Jnl. of IGPL*, 17:77–90, 2009.

[16] C. Cayrol, S. Doutre, and J. Mengin. On decision problems related to the preferred semantics for argumentation frameworks. *Jnl. Logic and Computation*, 13(3):377–403, 2003.

[17] Y. Dimopoulos and A. Torres. Graph theoretical structures in logic programs and default theories. *Theor. Comput. Sci.*, 170(1-2):209–244, 1996.

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

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

[20] P. E. Dunne. The Computational Complexity of Ideal Semantics. *Artificial Intelligence*, 173(18):1559–1591, 2009.

[21] P. E. Dunne and T. J. M. Bench-Capon. Coherence in finite argument systems. *Artificial Intelligence*, 141:187–203, 2002.

[22] P. E. Dunne and M. Wooldridge. Complexity of abstract argumentation. In *Argumentation in AI*, Springer, Berlin, pages 85–104, 2009.

[23] W. Dvořák and S. Woltran. Complexity of semi-stable and stage semantics in argumentation frameworks. *Inf. Process. Lett.*, 110:425–430, 2010.

[24] D. M. Gabbay. Fibring argumentation frames. *Studia Logica*, 93(2-3):231–295, 2009.

[25] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Jnl. of Logic and Computation*, 9(2):215–261, 1999.

[26] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in AI*, pages 105–129. Springer, 2009.

[27] S. Nofal, K. Atkinson and P. E. Dunne Algorithms for Decision Problems in Argument Systems under Preferred Semantics. *Artificial Intelligence*, 207:23-51, 2014.

[28] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, 1995.

[29] I. Rahwan, and F. Tohmé. Collective argument evaluation as judgement aggregation. In *Proc. 9th AAMAS*, pages 417–424, 2010.

[30] P.M. Thang, P.M Dung, and N.D. Hung. Towards a common framework for dialectical proof procedures in abstract argumentation. *Jnl. Logic and Computation*, 1071–1109, 2009.

[31] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proc. NAIC'96*, pages 357–368, 1996.

[32] B. Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In *Proc. 20th IJCAI*, pages 623–628, 2007.