

# Doing Argumentation using Theories in Graph Normal Form

Sjur Dyrkolbotn

Department of Informatics  
University of Bergen, Norway  
sjur.dyrkolbotn@ii.uib.no

**Abstract.** We explore some links between abstract argumentation, logic and kernels in digraphs. Viewing argumentation frameworks as propositional theories in graph normal form, we observe that the stable semantics for argumentation can be given equivalently in terms of satisfaction and logical consequence in classical logic. We go on to show that the complete semantics can be formulated using Łukasiewicz three-valued logic.

## 1 Introduction

Abstract argumentation in the style of Dung [11], has gained much popularity in the AI-community, see [3] for an overview. Dung introduced *argumentation frameworks*, networks of abstract arguments that are not assumed to have any particular internal structure. All that arguments can do is attack one another, and extension based semantics for argumentation identify sets of arguments that are in some sense successful. Given a set  $A$  of arguments, some requirements are intuitively natural to stipulate. If, for instance,  $a, b \in A$  and one of  $a$  and  $b$  attacks the other, then it seems problematic to accept  $A$  as successful -  $A$  effectively undermines itself. There are several other more or less natural constraints one might consider, some of which cannot be mutually satisfied, and this has given rise to several different semantics for argumentation frameworks, each able to capture some, but not all, intuitive requirements, see e.g. [2] for an overview and comparison of various approaches. In this paper we introduce *argumentation theories*, a representation of argumentation frameworks as propositional theories in graph normal form, a novel normal form for theories in propositional logic [4], closely connected to directed graphs.

In section 2 we give the necessary definitions from argumentation theory and we observe that stable sets in argumentation frameworks, satisfying assignments to their representations as theories, and kernels in the directed graphs obtained by reversing their attacks, are all one and the same. In section 3 we work with the representation of frameworks as theories, showing that the definition of a *complete extension* can be given equivalently in terms of satisfaction in Łukasiewicz three-valued logic  $\mathbb{L}_3$ .

## 2 Preliminaries

An *argumentation framework*, framework for short, is a finite digraph,  $F = \langle \mathcal{A}, \mathcal{R} \rangle$ , with  $\mathcal{A}$  a set of vertices, called *arguments*, and  $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$  a set of directed edges, called

the *attack relation*. For  $\langle a, a' \rangle \in \mathcal{R}$  we say that the argument  $a$  *attacks* the argument  $a'$ . We use the notation  $\mathcal{R}^+(x) = \{y \mid \langle x, y \rangle \in \mathcal{R}\}$  and  $\mathcal{R}^-(y) = \{x \mid \langle x, y \rangle \in \mathcal{R}\}$ . This notation extends point-wise to sets, e.g.  $\mathcal{R}^-(X) = \bigcup_{x \in X} \mathcal{R}^-(x)$ . We use the convention that  $\mathcal{R}^+(\emptyset) = \mathcal{R}^-(\emptyset) = \emptyset$ .

The most well-known semantics for argumentation, first introduced in [11] and [6] (semi-stable semantics), are given in the following definition.

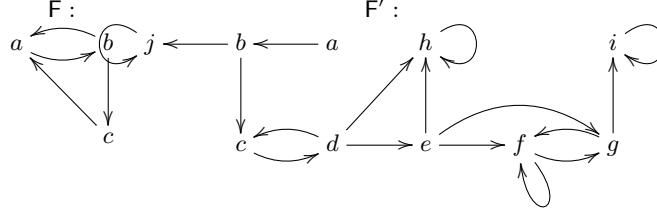
**Definition 21** *Given any argumentation framework  $F = \langle A, \mathcal{R} \rangle$  and a subset  $A \subseteq A$ , we define  $\mathcal{D}(A) = \{x \in A \mid \mathcal{R}^-(x) \subseteq \mathcal{R}^+(A)\}$ , the set of vertices defended by  $A$ . We say that*

- $A$  is conflict-free if  $\mathcal{R}^+(A) \subseteq A \setminus A$ , i.e. if there are no two arguments in  $A$  that attack each other.
- $A$  is admissible if it is conflict free and  $A \subseteq \mathcal{D}(A)$ . The set of all admissible sets in  $F$  is denoted  $a(F)$ .
- $A$  is complete if it is conflict free and  $A = \mathcal{D}(A)$ . The set of all complete sets in  $F$  is denoted  $c(F)$ .
- $A$  is the grounded set if it is complete and there is no complete set  $B \subseteq A$  such that  $B \subset A$ , it is denoted  $g(F)$ .
- $A$  is preferred if it is admissible and not strictly contained in any admissible set. The set of all preferred sets in  $F$  is denoted  $p(F)$ .
- $A$  is stable if  $\mathcal{R}^+(A) = A \setminus A$ . The set of all stable sets in  $F$  is denoted  $s(F)$ .
- $A$  is semi-stable if it is admissible and there is no admissible set  $B$  such that  $A \cup \mathcal{R}^+(A) \subset B \cup \mathcal{R}^+(B)$ . The set of all semi-stable sets in  $F$  is denoted by  $ss(F)$ .

For any  $\mathcal{S} \in \{a, c, g, p, s, ss\}$ , one also says that  $A \in \mathcal{S}(F)$  is an *extension* (of the type prescribed by  $\mathcal{S}$ ). Given a framework  $F$  and an argument  $a \in A$ , we say that  $a$  is *credulously* accepted with respect to some  $\mathcal{S} \in \{a, c, g, p, s, ss\}$  just in case there is some set  $A \in \mathcal{S}(F)$  such that  $a \in A$ . If  $a \in A$  for *every*  $A \in \mathcal{S}(F)$ , we say that  $A$  is *skeptically* accepted with respect to  $\mathcal{S}$ . If an argument is neither credulously nor skeptically accepted with respect to a semantics, it is *rejected*.<sup>1</sup> Intuitively, if an argument is credulously accepted, then it is involved in some line of argument that is successful; it is potentially useful, and should be considered further. If an argument is skeptically accepted, it is involved in all successful lines of arguments; it is beyond reproach, and arguing against it should be considered useless.

Notice that it follows from Definition 21 that the empty set is admissible and that all stable sets are semi-stable, all semi-stable sets are preferred, all preferred sets are complete and all complete sets are admissible. Also, it is not hard to see that the grounded extension is contained in every complete set of arguments. In Figure 1, we give two argumentation frameworks,  $F$  and  $F'$ , that serve as examples. In the framework  $F$ , every argument is attacked by some argument, and from this it follows that we have  $g(F) = \emptyset$ , i.e., the grounded extension is the empty set. The non-empty conflict-free sets are the singletons  $\{a\}$ ,  $\{b\}$  and  $\{c\}$ , but we observe that  $a$  does not defend himself against the attack he receives from  $c$  (since there is no attack  $(a, c)$ ), and that  $c$  does not defend himself against the attack he receives from  $b$ . So the only possible non-empty admissible set is  $\{b\}$ . It is indeed admissible;  $b$  is attacked only by  $a$  and he defends himself, attacking  $a$  in return. In fact, since  $b$  also attacks  $c$ , the set  $\{b\}$  is the unique stable set of this framework. It follows that  $s(F) = p(F) = ss(F) = \{\{b\}\}$

<sup>1</sup>Arguments that are credulously but not skeptically accepted are typically called *defensible*



**Fig. 1.** Two argumentation frameworks

and  $a(F) = c(F) = \{\emptyset, \{b\}\}$ . For a more subtle example, consider  $F'$ . The first thing to notice here is that we have an unattacked argument  $a$ , so the grounded extension is non-empty. In fact, the framework is such that all semantics from Definition 21 behave differently. It might look a bit unruly, but there are many self-attacking arguments that can be ruled out immediately (since they are not in any conflict-free sets), and it is easy to verify that the extensions of  $F'$  under the different semantics are the following:

$$\begin{aligned} g(F') &= \{a\}, s(F') = \emptyset, ss(F') = \{\{a, d, g\}\} \\ a(F') &= \{\emptyset, \{a\}, \{a, c\}, \{a, c, e\}, \{d\}, \{a, d\}, \{a, d, g\}, \{d, g\}\} \\ p(F') &= \{\{a, d, g\}, \{a, c, e\}\}, c(F') = \{\{a\}, \{a, d, g\}, \{a, c, e\}, \{a, d\}\} \end{aligned}$$

Notice that  $s(F') = \emptyset$  - no stable set exists in the framework. That stable sets are not guaranteed to exist is a major objection against what is otherwise a fairly conclusive notion of success - an internally consistent set of arguments that defeats all others. Other semantics for argumentation can - to quite some extent - be seen as attempts at arriving at a reasonable notion of success that is weak enough so that interesting extensions always exists, while still strong enough so that it is adequate for applications and have interesting theoretical properties. As we will see later, this is intimately related to the problem of *inconsistency handling* in logic, with stable sets not existing in a framework precisely when the corresponding propositional theory is not classically consistent.

Argumentation has close links to established concepts in both graph theory and logic. We start by briefly accounting for the link with what is known as kernels in the theory of directed graphs. Given a directed graph (digraph)  $D = \langle D, N \rangle$  with  $N \subseteq D \times D$ , a set  $K \subseteq D$  is a *kernel* in  $D$  if

$$N^-(K) = D \setminus K$$

Kernels were introduced by Von Neumann and Morgenstern in [18] in the context of cooperative games, but have since attracted quite a bit of interest from graph-theorists, see [5] for a recent overview. The connection to argumentation should be apparent. If we let  $\overleftarrow{D}$  denote the digraph obtained by reversing all edges in  $D$ , then it is not hard to see that a kernel in  $D$  is a stable set in  $\overleftarrow{D}$  and vice versa. In kernel theory, one also considers *local kernels* [17], which are sets  $L \subseteq D$  such that

$$N^+(L) \subseteq N^-(L) \subseteq D \setminus L$$

It is easy to verify that a local kernel in  $D$  is an admissible set in  $\overleftarrow{D}$  and vice versa. These two observations seem valuable to argumentation, since in graph theory, several interesting results and techniques have been found, especially concerning the question

of finding structural conditions that ensure the *existence* of kernels, see e.g. [9, 10, 15]. Still, as far as we are aware, the connection to argumentation has never been systematically explored.<sup>2</sup>

In this paper, we will describe some connections between argumentation and logic, and we will not explore the link with kernels any further. What is important for us, is that digraphs inspire a normal form for propositional theories where an assignment is satisfying iff it gives rise to a kernel in the corresponding digraph, see [4, 19] for two recent papers that explore this link. We remark that the observation we make here is in some sense implicit already in the original paper by Dung [11], who connects argumentation to logic programming. Still, the direct representation of argumentation frameworks as propositional theories is more straightforward and also seems useful, as suggested by the results we obtain in Section 3, and the fact that they can be established so easily.

A propositional formula,  $\phi$ , is said to be in *graph normal form* iff  $\phi = x \leftrightarrow \bigwedge_{y \in X} \neg y$  for propositional letters  $\{x\} \cup X$ . In [4] it is shown that this is indeed a normal form for propositional logic, every propositional theory has an equisatisfiable one containing only formulas of this form.<sup>3</sup> The correspondence with digraphs is detailed in [4, 19]. Instead of reversing edges in order to connect it to argumentation, we give our own formulation that can be applied directly. Given a framework  $F$ , we define the *argumentation theory*  $\text{TF}$  as follows:

$$\text{TF} = \{x \leftrightarrow \bigwedge_{y \in \mathcal{R}^-(x)} \neg y \mid x \in \mathcal{A}\} \quad (1)$$

For instance, the argumentation theory for the framework  $F$  depicted in Figure 1, consists of the following equivalences:

$$a \leftrightarrow \neg b \wedge \neg c, \quad b \leftrightarrow \neg a, \quad c \leftrightarrow \neg b$$

We observe that the only satisfying assignment in classical logic is  $\delta : \mathcal{A} \rightarrow \{\mathbf{0}, \mathbf{1}\}$  such that  $\delta(a) = \delta(c) = \mathbf{0}$  and  $\delta(b) = \mathbf{1}$ , corresponding to the fact that the only successful line of argument under stable semantics is the set containing the argument  $b$ , defeating  $a$  and  $c$ .

Argumentation theories are written using graph normal form, and we can also move in the opposite direction. The construction seems to be of independent interest, so we will present it here. It provides, in particular, a simple way in which results and techniques from argumentation can be applied to logic. Solving SAT, for instance, becomes the search for a stable set, while the other semantics for argumentation could provide new and useful ways in which to extract information from inconsistent theories.

For a positive natural number  $n$ , let  $[n] = \{1, \dots, n\}$ . Then, for any theory,  $\text{T} = \{x_1 \leftrightarrow \bigwedge_{x \in X_1} \neg x, \dots, x_n \leftrightarrow \bigwedge_{x \in X_n} \neg x\}$ , we define an argumentation framework  $F_T = \langle F_T, \mathcal{R}_T \rangle$  as follows

$$\begin{aligned} \mathcal{A}_T &= \bigcup_{i \in [n]} (\{x_i\} \cup X_i \cup \{\bar{x} \mid x \in X_i \wedge \forall i \in [n] : x \neq x_i\}) \\ \mathcal{R}_T &= (\bigcup_{i \in [n]} \{(x, x_i) \mid x \in X_i\}) \cup \{(x, \bar{x}), (\bar{x}, x) \mid x \in \mathcal{A}_T\} \end{aligned} \quad (2)$$

<sup>2</sup>The connection has been noted, for instance in [8], where the authors remark that some of their results concerning symmetric argumentation frameworks (all attacks are mutual) follow from basic results in kernel theory.

<sup>3</sup>Equisatisfiable means that for every satisfying assignment to one there is a satisfying assignment to the other, i.e., the assignments are not necessarily the same (the addition of new propositional letters must be permitted)

We introduce a fresh argument  $\bar{x}$  for every propositional letter  $x$  that does not occur to the left of any equivalence. The reason is that we do not want to force acceptance of  $x$  when viewed as an argument. Rather,  $x$  should be open for both acceptance and rejection, depending on the rest of the theory. This is achieved by adding the symmetric attack  $\{(x, \bar{x}), (\bar{x}, x)\}$ .

Let  $I = \{\bar{x} \mid x \in \mathcal{A}\}$  denote all the arguments from  $F_T$  that do not correspond to propositional letters used in  $T$ . Given a function  $\alpha : X \rightarrow Y$ , let  $\alpha|_Z$  denote its restriction to domain  $Z \subseteq X$ . Also, given  $\delta : X \rightarrow \{\mathbf{0}, \mathbf{1}\}$ , we let  $\bar{\delta}$  denote the boolean evaluation of formulas induced by  $\delta$  according to classical logic.

Given a stable set  $E \subseteq F_T$ , consider  $\delta_E : \mathcal{A} \rightarrow \{\mathbf{0}, \mathbf{1}\}$  defined by  $\delta_E(x) = \mathbf{1}$  iff  $x \in E$  (meaning that  $\delta_E(x) = \mathbf{0}$  for all  $x \in \mathcal{A}_T \setminus E$ ). It is not hard to show that  $\delta_E|_{\mathcal{A}_T \setminus I}$  is a satisfying assignment for  $T$  in classical logic, i.e., that  $\bar{\delta}_E(\phi) = \mathbf{1}$  for all  $\phi \in T$ . Similarly, if we are given a satisfying assignment  $\delta : \mathcal{A}_T \setminus I \rightarrow \{\mathbf{0}, \mathbf{1}\}$ , we obtain the stable set  $E_\delta = \{x \in \mathcal{A}_T \mid \delta(x) = \mathbf{1}\} \cup \{\bar{x} \in I \mid \delta(x) = \mathbf{0}\}$ .

The constructions given in this section are completely analogous to the ones given in [4] in order to establish equivalence between theories in graph normal form and kernels in digraphs. Therefore, we will simply summarize, without a formal proof, one obvious consequence for argumentation. We let  $\models$  denote the satisfaction relation in classical propositional logic, and we let  $\perp$  denote contradiction in classical logic.

**Theorem 24** *Given an argumentation framework  $F$ , and an argument  $a \in \mathcal{A}$*

- (1)  *$F$  admits a stable extension iff  $TF \not\models \perp$  (i.e. iff  $TF$  is satisfiable)*
- (2)  *$a \in \mathcal{A}$  is credulously accepted with respect to stable semantics iff  $\{a\} \cup TF \not\models \perp$*
- (3)  *$a \in \mathcal{A}$  is skeptically accepted with respect to stable semantics iff  $TF \models a$*

When we succeed in giving semantics for argumentation in terms of logical consequence and consistency, we no longer need to consider just atomic arguments and statements about them specified in some informal or semi-formal way. Rather, one can now use logic, and form a propositional formula which can then be evaluated against the theory corresponding to the framework. For instance, given a framework  $F$  with  $a \in \mathcal{A}$ , we have that  $a$  is not credulously accepted with respect to stable semantics iff  $TF$  is a model of  $\neg a$ , i.e., iff we have  $TF \models \neg a$ . For another example, consider how one may write simply  $a \rightarrow b$  to indicate that every successful line of argument which involves accepting  $a$  must also involve accepting  $b$ . More generally, when we succeed in providing a formulation in terms of a known logic, we can use reasoning systems developed for this logic both to address standard notions from argumentation such as credulous and skeptical acceptance, and also to check how arbitrary propositional formulas fare with respect to an argumentation theory; formulas that can now be seen as statements about interaction among arguments in the underlying argumentation framework. Conversely, we may use techniques developed in argumentation theory to analyze logical theories in graph normal form - something that might prove particularly useful for algorithmic problems, since the digraph structure of argumentation frameworks should prove particularly useful in this regard. We remark that interesting results have already been obtained which exploits digraph-properties for tackling algorithmic problems concerning the semantic properties of argumentation frameworks, see e.g., [12, 13].

It seems, in light of this, that the search for nice logical accounts of argumentation is a highly worthwhile direction of research. In fact, it has recently been taken up by logicians coming at argumentation from a more theoretical, less application-oriented,

angle, see e.g., [7, 16]. Conceptually, we differ from these approaches in that we rely on the link with known concepts in digraph theory and the representation of frameworks as propositional theories, rather than on the use of modal logic.

### 3 Argumentation and Łukasiewicz logic L3

In this section we will characterize the complete extensions logically. This necessitates a move away from classical logic. In Caminada and Gabbay [7], a characterization is provided using modal logic and a complicated modal version of Equation 1. In this section we show that complete extensions can be characterized much more simply using Łukasiewicz three valued logic L3. This observation is also made in [14], but there it is formalized only with respect to local kernels and a new logic for reasoning about paradoxes. The details and consequences are not worked out in the context of argumentation.

Here, we directly link L3 to argumentation by showing that any satisfying assignment to TF gives rise to a complete extension in F and vice versa. The argument we give proceeds by showing that an assignment is satisfying iff it is what is known in argumentation theory as a *complete Caminada labeling*. While not technically challenging, this result seems nice, since it implies that skeptical and credulous acceptance of arguments with respect to the complete semantics reduces to logical consequence and satisfiability in L3. We remark, in particular, that Łukasiewicz three-valued logic admits a nice proof theory, see e.g., [1]. We also remark that since the grounded extension is complete and also contained in every complete extension, characterizing skeptical acceptance with respect to complete semantics amounts to completely characterizing the grounded extension.

Given a propositional language over connectives  $\{\neg, \wedge, \rightarrow\}$  and propositional alphabet  $\mathcal{A}$ , the semantics of logic L3 is defined as follows.

**Definition 31** *Given an assignment  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  its extension,  $\bar{\delta}$ , is defined inductively on complexity of formulas.*

- $\bar{\delta}(a) = \delta(a)$  for all  $a \in \mathcal{A}$
- $\bar{\delta}(\neg\phi) = 1 - \bar{\delta}(\phi)$
- $\bar{\delta}(\phi \rightarrow \psi) = \min\{1, (1 - \bar{\alpha}(\phi)) + \bar{\alpha}(\psi)\}$
- $\bar{\delta}(\phi \wedge \psi) = \min\{\bar{\alpha}(\phi), \bar{\alpha}(\psi)\}$

*The consequence relation of Łukasiewicz logic is  $\models_{\mathbf{L}} \subseteq 2^{\mathcal{L}} \times \mathcal{L}$ , defined such that  $\Phi \models_{\mathbf{L}} \phi$  iff for all  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$ , we have that  $\bar{\delta}(\phi) = \mathbf{1}$  whenever  $\bar{\delta}(\psi) = \mathbf{1}$  for all  $\psi \in \Phi$ ,*

Notice that  $\phi \rightarrow \psi$  obtains semantic value 1 ("true") under some assignment just in case  $\psi$  does not receive a lower semantic value than  $\phi$ . Defining  $\phi \leftrightarrow \psi = (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$  as usual, one also notices that  $\phi \leftrightarrow \psi$  obtains the value 1 just in case  $\phi$  and  $\psi$  obtain the same semantic value.

Next we define the complete Caminada labellings, not using the original definition, but the equivalence stated and proven in [7, Proposition 1, p. 6-7]

**Definition 32** *A function  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  is a complete Caminada labeling iff for all  $x \in \mathcal{A}$  we have:*

- If  $\delta(x) = 1$  then for all  $y \in \mathcal{R}^-(x) : \delta(y) = 0$

- If  $\delta(x) = 0$  then there is  $y \in \mathcal{R}^-(x) : \delta(y) = 1$
- If  $\delta(x) = 1/2$  then there is  $y \in \mathcal{R}^-(x)$  such that  $\delta(y) = 1/2$  and there is no  $z \in \mathcal{R}^-(x)$  such that  $\delta(z) = 1$

In [7, Theorem 2], the authors prove that if  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  is a complete Caminada labeling then  $\delta^1 = \{x \in \mathcal{A} \mid \delta(x) = 1\}$  is a complete extension for  $\mathbf{F}$ . They also show that if  $E \subseteq \mathcal{A}$  is a complete extension, then there is a corresponding complete Caminada labeling  $\delta_E : \mathcal{A} \rightarrow \{0, 1/2, 1\}$ , defined as follows:

- $\delta_E(x) = 1$  for all  $x \in E$
- $\delta_E(x) = 0$  for all  $x \in \mathcal{R}^+(E)$
- $\delta_E(x) = 1/2$  for all  $x \in \mathcal{A} \setminus (E \cup \mathcal{R}^+(E))$

In the following, we prove that a complete Caminada labeling can be equivalently defined as a satisfying assignment for  $\mathbf{TF}$  in the logic  $\mathbf{L3}$ . This shows that the results from [7] have direct logical content, and that we do not need to introduce modal logic to characterize complete extensions logically.

**Theorem 33** *Given an argumentation framework  $\mathbf{F}$ , we have that  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  is a complete Caminada labeling iff  $\bar{\delta}(\phi) = 1$  for all  $\phi \in \mathbf{TF}$ .*

PROOF.  $\Rightarrow$ ) Assume that  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  is a complete Caminada labeling and consider an arbitrary  $\phi = x \leftrightarrow \bigwedge_{y \in \mathcal{R}^-(x)} \neg y \in \mathbf{TF}$ . We need to show that  $\bar{\delta}(x \leftrightarrow \bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 1$ . There are three cases:

- $\delta(x) = 1$ . Since  $\delta$  is a complete Caminada labeling we have, for all  $y \in \mathcal{R}^-(x)$ ,  $\delta(y) = 0$ . It follows that  $\bar{\delta}(\neg y) = 1$  for all  $y \in \mathcal{R}^-(x)$ . Since all conjuncts are true, we conclude that  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 1 = \delta(x) = \bar{\delta}(x)$ . So  $\bar{\delta}(\phi) = 1$  as desired.
- $\delta(x) = 0$ . Since  $\delta$  is a complete Caminada labeling it follows that there is some  $y \in \mathcal{R}^-(x)$  such that  $\delta(y) = 1$ . Then  $\bar{\delta}(\neg y) = 0$ , so it follows that  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 0 = \delta(x) = \bar{\delta}(x)$ . So  $\bar{\delta}(\phi) = 1$  as desired.
- $\delta(x) = 1/2$ . Since  $\delta$  is a complete Caminada labeling, there is some  $y \in \mathcal{R}^-(x)$  such that  $\delta(y) = 1/2$ . It also follows that there is no  $z \in \mathcal{R}^-(x)$  such that  $\delta(z) = 1$ . From this we conclude that  $1/2 = \min\{\bar{\delta}(y) \mid y \in \mathcal{R}^-(x)\}$  which means  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 1/2 = \delta(x) = \bar{\delta}(x)$ . So  $\bar{\delta}(\phi) = 1$  as desired.

$\Leftarrow$ ) Assume that  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$  is a satisfying assignment for  $\mathbf{TF}$ , i.e. that  $\bar{\delta}(\phi) = 1$  for all  $\phi \in \mathbf{TF}$ . Consider arbitrary  $\phi = x \leftrightarrow \bigwedge_{y \in \mathcal{R}^-(x)} \neg y \in \mathbf{TF}$ . Again there are three cases:

- $\delta(x) = 1$ . Since  $\bar{\delta}(\phi) = 1$ , we have  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 1$ . It follows that  $\bar{\delta}(\neg y) = 1$  and therefore  $\delta(y) = 0$  for all  $y \in \mathcal{R}^-(x)$ . So the criterion of Definition 32 is met in this case.
- $\delta(x) = 0$ . Since  $\bar{\delta}(\phi) = 1$ , we have  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 0$ . It follows that  $\bar{\delta}(\neg y) = 0$ , and therefore  $\delta(y) = 1$  for some  $y \in \mathcal{R}^-(x)$ . So the criterion of Definition 32 is met.
- $\delta(x) = 1/2$ . Since  $\bar{\delta}(\phi) = 1$ , we have  $\bar{\delta}(\bigwedge_{y \in \mathcal{R}^-(x)} \neg y) = 1/2$ . This means that  $1/2 = \min\{\bar{\delta}(\neg y) \mid y \in \mathcal{R}^-(x)\}$ . So there must be some  $y \in \mathcal{R}^-(x)$  such that  $\bar{\delta}(\neg y) = 1/2$ , which means  $\delta(y) = 1/2$ . Also, it follows that there is no  $z \in \mathcal{R}^-(x)$  such that  $\bar{\delta}(\neg z) = 0$ . So there is no  $z \in \mathcal{R}^-(x)$  such that  $\delta(z) = 1$ , meaning that the criterion of Definition 32 is met in this case as well.

□

We conclude by stating the following corollary, which sums up the immediate consequences for argumentation. We let  $\perp$  denote contradiction in L3.

**Corollary 34** *Given an argumentation framework  $F$ . An argument  $a \in \mathcal{A}$  is skeptically accepted with respect to complete semantics iff  $\text{TF} \models_{\text{L}} a$  and credulously accepted iff  $\text{TF} \cup \{a\} \not\models_{\text{L}} \perp$ .*

PROOF. For the first claim, remember that  $a \in \mathcal{A}$  is said to be skeptically accepted with respect to complete semantics iff for all complete extensions  $E \subseteq \mathcal{A}$  we have  $a \in E$  iff  $\delta(a) = 1$  for all complete Caminada labellings  $\delta : \mathcal{A} \rightarrow \{0, 1/2, 1\}$ . By Theorem 33 this is the same as saying that  $\delta(a) = 1$  for all  $\delta$  such that  $\bar{\delta}(\phi) = 1$  for all  $\phi \in \text{TF}$ . By Definition 31, this is the same as  $\text{TF} \models_{\text{L}} a$ . The second claim follows similarly. □

As already noted, the logical approach means that we can form complex statements to express various claims about arguments and their interaction in the framework. For the case of complete semantics, we may write, for instance,  $a \leftrightarrow \neg a$  to indicate that the argument  $a$  can be neither defeated nor accepted. It is not hard to see that this formula is true in a model  $\text{TF}$  iff neither  $a$  nor any of its attackers is credulously accepted with respect to the complete semantics. So it does capture the intended meaning, and we believe that the formula is a beautiful representation of an argument having malfunctioned, becoming instead a paradox: an argument such that accepting it is logically equivalent with defeating it!

Using a logical language to talk about argumentation provides clarity, but also sheds light on subtleties that we might not otherwise come to fully appreciate. As an example, consider again the framework  $F'$  from Figure 1. With some thought, we see that in order for both  $h$  and  $i$  to be defeated, it is necessary to use  $d$  to defeat  $h$  (since using  $e$  will defeat  $g$  - the only attacker of  $i$  except  $i$  itself). But if we try the formula  $(\neg h \wedge \neg i) \rightarrow d$  and check if it follows logically from  $\text{TF}'$  in L3, we find that it does not. The explanation for this is that implication in L3 treats an argument that is neither accepted nor defeated as closer to truth than an argument that is defeated. Since there is a complete set  $(\{a, c, e\})$  such that  $d$  is defeated while  $i$  is neither accepted nor defeated, it follows from this that a countermodel to the implication can be found. Still, it *is* possible to express a claim that correctly describes the state of affairs that obtains. In fact, after some thought, it is seen that the claim we stated informally - that you must accept  $d$  to defeat both  $h$  and  $i$  - while true, actually serves to misrepresent the situation at hand. For what we actually have is something stronger, namely that in order for  $h$  and  $i$  to obtain *any* of the two classical values (true/false or, if you like, defeated/accepted), we have to accept  $d$ . This we can express by the implication  $(\neg(h \leftrightarrow \neg h) \wedge \neg(i \leftrightarrow \neg i)) \rightarrow d$ , and now we observe that this implication does indeed follow logically from  $\text{TF}'$ . Also, we obtain the further logical consequence  $(\neg(h \leftrightarrow \neg h) \wedge \neg(i \leftrightarrow \neg i)) \rightarrow (\neg h \wedge \neg i)$  which is also *stronger* than the original intuition that we had. Thus, what seemed at first sight a shortcoming of the logical approach really suggested a more precise analysis of the situation, one that we might not as easily have arrived at without a logical formulation.



## 4 Conclusion

In this paper we have observed how argumentation frameworks can be viewed as propositional theories in graph normal form. We have shown that this makes it possible to capture the stable and complete semantics using classical and three-valued Łukasiewicz logic respectively. Moreover, we have argued that argumentation theories provide a nice way in which to talk about argumentation in a logical language. It is much more straightforward than the modal approach from [7], and exploring it further seems worthwhile. For a possible first step in future work, we remark that both the preferred and semi-stable sets seem to involve notions of maximal consistency that it should be possible, and interesting, to express in terms of logic.

## References

1. Arnon Avron. Natural 3Valued Logics - Characterization and Proof Theory. *Journal of Symbolic Logic*, 56:276–294, 1991.
2. Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10–15):675–700, 2007.
3. T.J.M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10–15):619–641, 2007.
4. Marc Bezem, Clemens Grabmayer, and Michał Walicki. Expressive power of digraph solvability. *Annals of Pure and Applied Logic*, 2011. [to appear].
5. Endre Boros and Vladimir Gurvich. Perfect graphs, kernels and cooperative games. *Discrete Mathematics*, 306:2336–2354, 2006.
6. Martin Caminada. Semi-stable semantics. In *Proceedings of the 2006 conference on Computational Models of Argument: Proceedings of COMMA 2006*, pages 121–130. IOS Press, 2006.
7. Martin W. A. Caminada and Dov M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.
8. Sylvie Coste-marquis, Caroline Devred, and Pierre Marquis. Symmetric argumentation frameworks. In *Proc. 8th European Conf. on Symbolic and Quantitative Approaches to Reasoning With Uncertainty (ECSQARU)*, volume 3571 of *LNAI*, pages 317–328. Springer-Verlag, 2005.
9. Pierre Duchet. Graphes noyau-parfaits, II. *Annals of Discrete Mathematics*, 9:93–101, 1980.
10. Pierre Duchet and Henry Meyniel. Une généralisation du théorème de Richardson sur l’existence de noyaux dans les graphes orientés. *Discrete Mathematics*, 43(1):21–27, 1983.
11. Phan Minh 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.
12. Paul E. Dunne. Computational properties of argument systems satisfying graph-theoretic constraints. *Artif. Intell.*, 171(10-15):701–729, 2007.
13. Wolfgang Dvorák, Reinhard Pichler, and Stefan Woltran. Towards fixed-parameter tractable algorithms for abstract argumentation. *Artif. Intell.*, 186:1–37, 2012.
14. Sjur Dyrkolbotn and Michał Walicki. Propositional discourse logic. (*submitted*), 2011. [www.ii.uib.no/~michal/graph-paradox.pdf](http://www.ii.uib.no/~michal/graph-paradox.pdf).
15. Hortensia Galeana-Sánchez and Victor Neumann-Lara. On kernels and semikernels of digraphs. *Discrete Mathematics*, 48(1):67–76, 1984.

16. Davide Grossi. On the logic of argumentation theory. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1*, AAMAS '10, pages 409–416, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.
17. Victor Neumann-Lara. Seminúcleos de una digráfica. Technical report, Anales del Instituto de Matemáticas II, Universidad Nacional Autónoma México, 1971.
18. John von Neumann and Oscar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944 (1947).
19. Michał Walicki and Sjur Dyrkolbotn. Finding kernels or solving SAT. *Journal of Discrete Algorithms*, 10:146–164, 2012.