Defense semantics of argumentation: encoding reasons for accepting arguments

Beishui Liao[1] and Leendert van der Torre[2]

[1] Zhejiang University, China
[2] University of Luxembourg, Luxembourg

Abstract. In this paper we show how the defense relation among abstract arguments can be used to encode the reasons for accepting arguments. After introducing a novel notion of defenses and defense graphs, we propose a defense semantics together with a new notion of defense equivalence of argument graphs, and compare defense equivalence with standard equivalence and strong equivalence, respectively. Then, based on defense semantics, we define two kinds of reasons for accepting arguments, i.e., direct reasons and root reasons, and a notion of root equivalence of argument graphs. Finally, we show how the notion of root equivalence can be used in argumentation summarization.

Keywords: abstract argumentation, defense graph, defense semantics, argumentation equivalence, argumentation summarization

1 Introduction

In the field of AI, abstract argumentation is mainly about evaluating the status of arguments in an argument graph [1–3], which is composed of a set of abstract arguments and a set of attacks between them [4]. In many topics such as equivalence [5–7], summarization [8], and dynamics in argumentation [9, 10], the notion of extensions plays a central role. Since in classical argumentation semantics, an extension is a set of arguments that are collectively accepted, the existing theories and approaches based on this notion are mainly focused on exploiting the status of individual arguments. However, besides the status of individual arguments, in many situations, we need to know the reasons for accepting arguments in terms of a defense relation. The following are two simple examples.

$$\mathcal{F}_1: \quad a \rightrightarrows c_1 \longrightarrow c_2 \rightrightarrows b \qquad \mathcal{F}_2: \quad a \leftrightarrow b$$
$$c_4 \longleftarrow c_3$$
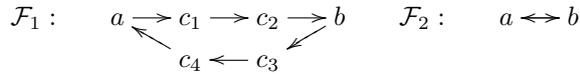
First, consider arguments a, b in $\mathcal{F}_1$ and $\mathcal{F}_2$. In $\mathcal{F}_1$, accepting a is a reason to accept c_2, accepting c_2 is a reason to accept c_3, and accepting c_3 is a reason to accept a. If we allow this relation to be transitive, we find that accepting a is a reason to accept a. Similarly, accepting b is a reason to accept b. Meanwhile, in $\mathcal{F}_2$, we have: accepting a is a reason to accept a, and accepting b is a reason to accept b. So, from the perspective of the reasons for accepting a and b, $\mathcal{F}_2$ is equivalent to $\mathcal{F}_1$, or $\mathcal{F}_2$ is a summarization of $\mathcal{F}_1$.

Second, consider the question when two argument graphs are equivalent in a dynamic setting. For $\mathcal{F}_3$ and $\mathcal{F}_4$ below, both of them have a complete extension $\{a, c\}$. However, the reasons of accepting c in $\mathcal{F}_3$ and $\mathcal{F}_4$ are different. For the former, c is defended by a, while for the latter, c is unattacked and has no defender. In this sense, $\mathcal{F}_3$ and $\mathcal{F}_4$ are not equivalent. For example, in order to change the status of argument c from "accepted" to "rejected", in $\mathcal{F}_3$, one may produce a new argument to attack the defender a, or to directly attack c. However, in $\mathcal{F}_4$ using an argument to attack a cannot change the status of c, since a is not a defender of c.

$$\mathcal{F}_3: \quad a \longrightarrow b \longrightarrow c \qquad\qquad \mathcal{F}_4: \quad a \longrightarrow b \qquad c$$

From the above two examples, one question arises: under what conditions, can two argument graphs be viewed as equivalent? The existing notions of argumentation equivalence, including standard equivalence and strong equivalence, are not sufficient to capture the equivalence of the argument graphs in the situations mentioned above. More specifically, $\mathcal{F}_1$ and $\mathcal{F}_2$ are not equivalent in terms of the notion of standard equivalence or that of strong equivalence, but they are equivalent in the sense that the reasons for accepting arguments a and b in these two graphs are the same. $\mathcal{F}_3$ and $\mathcal{F}_4$ are equivalent in terms of standard equivalence, but they are not equivalent in the sense that the reasons for accepting c in these two graphs are different. Although the notion of strong equivalence can be used to identify the difference between $\mathcal{F}_3$ and $\mathcal{F}_4$, conceptually it is not defined from the perspective of reasons for accepting arguments.

Note that the reasons for accepting arguments in the above two examples are depicted in terms of a defense relation, which plays a central role in Dung's concept of admissibility and thus in admissibility based semantics. So, it is natural to define a new semantics in this paper based on a defense relation such that the reasons for accepting arguments can be encoded.

Since the new semantics is defined at the level of abstract argumentation, it can be applied to various structured argumentations systems. In particular, in the field of legal reasoning [11], argumentation can be used to model legal interpretation, dialogue, and deontic reasoning, etc. In all these applications, it is useful to make clear the reasons for accepting arguments in terms of a defense relation. In this paper, we will formulate a defense semantics for abstract argumentation, while its application to various structured argumentation systems is left to future work. The structure of this paper is as follows. In Section 2, we introduce some basic notions of argumentation semantics. In Section 3, we propose the notions of defenses and defense graphs, which lay a foundation of this paper. In Section 4, we formulate defense semantics by applying classical argumentation semantics to defense graphs, and study some properties of this new semantics. In Section 5, we introduce two kinds of reasons for accepting arguments in terms of defense semantics. We conclude in Section 6.

2   Argumentation semantics

An argument graph or argumentation framework (AF) is defined as $\mathcal{F} = (\mathcal{A}, \rightarrow)$, where $\mathcal{A}$ is a finite set of arguments and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a set of attacks between arguments [4].

Let $\mathcal{F} = (\mathcal{A}, \rightarrow)$ be an argument graph. Given a set $B \subseteq \mathcal{A}$ and an argument $\alpha \in \mathcal{A}$, B attacks α, denoted $B \rightarrow \alpha$, iff there exists $\beta \in B$ such that $\beta \rightarrow \alpha$. Given an argument $\alpha \in \mathcal{A}$, let $\alpha^{\leftarrow} = \{\beta \in \mathcal{A} \mid \beta \rightarrow \alpha\}$ be the set of arguments attacking α, and $\alpha^{\rightarrow} = \{\beta \in \mathcal{A} \mid \alpha \rightarrow \beta\}$ be the set of arguments attacked by α. When $\alpha^{\leftarrow} = \emptyset$, we say that α is unattacked, or α is an initial argument.

Given $\mathcal{F} = (\mathcal{A}, \rightarrow)$ and $E \subseteq \mathcal{A}$, we say: E is *conflict-free* iff $\nexists \alpha, \beta \in E$ such that $\alpha \rightarrow \beta$; $\alpha \in \mathcal{A}$ is *defended* by E iff $\forall \beta \in \alpha^{\leftarrow}$, it holds that $E \rightarrow \beta$; E is *admissible* iff E is conflict-free, and each argument in E is defended by E; E is a *complete extension* iff E is admissible, and each argument in $\mathcal{A}$ that is defended by E is in E; E is a *grounded extension* iff E is the minimal (w.r.t. set-inclusion) complete extension; E is a *preferred extension* iff E is a maximal (w.r.t. set-inclusion) complete extension; E is a *stable extension* iff E is conflict-free and E attacks each argument that is not in E. We use $\sigma(\mathcal{F})$ to denote the set of argument extensions of $\mathcal{F}$ under semantics σ, where σ is a function mapping each argument graph to a set of argument extensions. We use $\mathrm{co}, \mathrm{gr}, \mathrm{pr}$ and st to denote complete, grounded, preferred and stable semantics respectively. There are some other argumentation semantics (cf. [2] for an overview).

For argument graphs $\mathcal{F}_1 = (\mathcal{A}_1, \rightarrow_1)$ and $\mathcal{F}_2 = (\mathcal{A}_2, \rightarrow_2)$, we use $\mathcal{F}_1 \cup \mathcal{F}_2$ to denote $(\mathcal{A}_1 \cup \mathcal{A}_2, \rightarrow_1 \cup \rightarrow_2)$. The standard equivalence and strong equivalence of argument graphs are defined as follows. For simplicity, in this paper, when we talk about equivalence of AFs, we mainly consider the cases under complete semantics, while the full-fledged study of equivalence will be presented in an extended version of the present paper.

Definition 1 (Standard and strong equivalence of AFs). *[5] Let $\mathcal{F}$ and $\mathcal{G}$ be two argument graphs, and σ be a semantics.*

- *$\mathcal{F}$ and $\mathcal{G}$ are of standard equivalence w.r.t. a semantics σ, in symbols $\mathcal{F} \equiv^{\sigma} \mathcal{G}$, iff $\sigma(\mathcal{F}) = \sigma(\mathcal{G})$.*
- *$\mathcal{F}$ and $\mathcal{G}$ are of strong equivalence w.r.t. a semantics σ, in symbols $\mathcal{F} \equiv_s^{\sigma} \mathcal{G}$, iff for every argument graph $\mathcal{H}$, it holds that $\sigma(\mathcal{F} \cup \mathcal{H}) = \sigma(\mathcal{G} \cup \mathcal{H})$.*

Example 1. Consider $\mathcal{F}_1 - \mathcal{F}_4$ in Section 1. In terms of Definition 1, under complete semantics, since $\mathrm{co}(\mathcal{F}_1) \neq \mathrm{co}(\mathcal{F}_2)$, $\mathcal{F}_1 \not\equiv^{\mathrm{co}} \mathcal{F}_2$, which implies that $\mathcal{F}_1 \not\equiv_s^{\mathrm{co}} \mathcal{F}_2$. And, since $\mathrm{co}(\mathcal{F}_3) = \mathrm{co}(\mathcal{F}_4)$, $\mathcal{F}_3 \equiv^{\mathrm{co}} \mathcal{F}_4$. Let $\mathcal{H} = (\{d\}, \{d \rightarrow a\})$. Since $\mathrm{co}(\mathcal{F}_3 \cup \mathcal{H}) \neq \mathrm{co}(\mathcal{F}_4 \cup \mathcal{H})$, $\mathcal{F}_3 \not\equiv_s^{\mathrm{co}} \mathcal{F}_4$.

Given an argument graph $\mathcal{F} = (\mathcal{A}, \rightarrow)$, the kernel of $\mathcal{F}$ under complete semantics, call *c-kernel*, is defined as follows.

Definition 2 (c-kernel of an AF). *[5] For an argument graph $\mathcal{F} = (\mathcal{A}, \rightarrow)$, the c-kernel of $\mathcal{F}$ is defined as $\mathcal{F}^{\mathrm{ck}} = (\mathcal{A}, \rightarrow^{\mathrm{ck}})$, where*

$$\rightarrow^{\mathrm{ck}} = \rightarrow \setminus \{\alpha \rightarrow \beta \mid \alpha \neq \beta, \alpha \rightarrow \alpha, \beta \rightarrow \beta\} \tag{1}$$

According to [5], it holds that $\mathrm{co}(\mathcal{F}) = \mathrm{co}(\mathcal{F}^{\mathrm{ck}})$, and for any AFs $\mathcal{F}$ and $\mathcal{G}$, $\mathcal{F}^{\mathrm{ck}} = \mathcal{G}^{\mathrm{ck}}$ iff $\mathcal{F} \equiv_s^{\mathrm{co}} \mathcal{G}$.

3   Defenses and defense graph

According to classical argumentation semantics, with respect to an extension E, an argument $\alpha \in E$ is accepted because it is initial or for all $\gamma \in \alpha^{\leftarrow}$, γ is attacked by an argument in E. So, for all $\alpha, \beta \in E$, if there exists $\gamma \in \mathcal{A} \setminus E$ such that $\alpha \to \gamma$ and $\gamma \to \beta$, we say that accepting α is a (partial) reason to accept β, denoted as $\langle \alpha, \beta \rangle$. And, for all $\beta \in E$ if $\beta^{\leftarrow} = \emptyset$ (i.e., β is an initial argument), we say that β is accepted without a reason, denoted as $\langle \emptyset, \beta \rangle$ where $\emptyset$ is a symbol denoting an empty position. In this paper, $\langle \alpha, \beta \rangle$ or $\langle \emptyset, \beta \rangle$ is called a *defense*.

Without referring to any specific extension, a defense $\langle \alpha, \beta \rangle$ can be viewed as a relation between α and β satisfying some constraints. Intuitively, there are the following two minimal constraints. First, $\{\alpha, \beta\}$ is conflict-free. Otherwise, they can not be both accepted. Second, there exists $\gamma \in \mathcal{A} \setminus \{\alpha, \beta\}$ such that $\alpha \to \gamma$ and $\gamma \to \beta$, in the sense that α defends β by attacking β's attacker γ. Regarding the defense $\langle \emptyset, \beta \rangle$, the only constraint is that β is initial.

Example 2. Consider $\mathcal{F}_5$ below. $\langle \emptyset, a \rangle$, $\langle a, c \rangle$ and $\langle b, d \rangle$ are defenses. Note that the three defenses do not refer to a specific extension.

$$\mathcal{F}_5: \qquad a \longrightarrow b \longrightarrow c \longrightarrow d$$

Based on the above analysis, we have the following definition.

Definition 3 (Defense). *Let $\mathcal{F} = (\mathcal{A}, \to)$ be an argument graph. For $\alpha, \beta \in \mathcal{A}$,*

- *$\langle \alpha, \beta \rangle$ is a defense iff $\{\alpha, \beta\}$ is conflict-free, and $\exists \gamma \in \mathcal{A}$ such that $\alpha \to \gamma$ and $\gamma \to \beta$.*
- *$\langle \emptyset, \beta \rangle$ is a defense iff β is initial.*

The set of defenses of $\mathcal{F}$ is denoted as $\mathcal{F}^{\mathrm{DEF}}$. Given a defense $\langle \alpha, \beta \rangle$ or $\langle \emptyset, \beta \rangle \in \mathcal{F}^{\mathrm{DEF}}$, we call α the *defender*, and β the *defendee*, of the defense. Given a set $D \subseteq \mathcal{F}^{\mathrm{DEF}}$, we write $\mathrm{defendee}(D) = \{\beta \mid \langle \alpha, \beta \rangle, \langle \emptyset, \beta \rangle \in D\}$ to denote the set of defendees in D, $\mathrm{defender}(D) = \{\alpha \mid \langle \alpha, \beta \rangle \in D\}$ to denote the set of defenders in D, and $\mathrm{def}(D) = \mathrm{defendee}(D) \cup \mathrm{defender}(D)$ be the set of defendees and defenders in D.

Note that not all arguments of an AF are included in the defenses. Consider the following example.

Example 3. In $\mathcal{F}_6$, $\mathcal{F}_7$ and $\mathcal{F}_8$, $\langle a_2, a_4 \rangle$, $\langle a_3, a_5 \rangle$, $\langle a_4, a_6 \rangle$, $\langle \emptyset, a_7 \rangle$, $\langle a_7, a_9 \rangle$, $\langle \emptyset, a_{11} \rangle$ and $\langle a_{11}, a_{13} \rangle$ are defenses, while some defense-like pairs, for instance (a_1, a_3) and (a_{14}, a_{13}), are not defenses since both $\{a_1, a_3\}$ and $\{a_{14}, a_{13}\}$ are not conflict-free. And, $(\emptyset, a_{10})$, $(\emptyset, a_{14})$ and $(\emptyset, a_{15})$ are not defenses, because they are not initial arguments, but either self-attacked or attacked by a self-attacked argument.

Given a defense $\langle x, \alpha \rangle$ where $x \in \mathcal{A} \cup \{\emptyset\}$ and $\alpha \in \mathcal{A}$, $\langle x, \alpha \rangle$ can be regarded as a meta-argument. Its status is affected by other defenses and/or other defense-like pairs (cf. (a_1, a_3) and $(\emptyset, a_{10})$ in Example 3). Since the pairs like (a_1, a_3) and $(\emptyset, a_{10})$ are not accepted as a defense, but may be used to hamper the acceptance of some defenses, their behavior is similar to that of *defeaters* in defeasible logic. We call them *defeaters of defenses (DoD)*.

Definition 4 (Defeaters of defenses). *Let $\mathcal{F} = (\mathcal{A}, \rightarrow)$ be an argument graph. For $\alpha, \beta \in \mathcal{A}$,*

– *(α, β) is a DoD, iff $\{\alpha, \beta\}$ is not conflict-free, and $\exists \gamma \in \mathcal{A} \setminus \{\alpha, \beta\}$ such that $\alpha \rightarrow \gamma$ and $\gamma \rightarrow \beta$.*
– *$(\emptyset, \beta)$ is a DoD, iff β is self-attacked or attacked by a self-attacked argument.*

The set of DoDs of $\mathcal{F}$ is denoted as $\mathcal{F}^{\mathrm{DoD}}$.

In this definition, note that (α, β) and $(\emptyset, \beta)$ are not accepted as a defense, and may be used to hamper the acceptance of some defenses. This does not mean that the arguments α and β in the corresponding argument graph can not be accepted, since they may in some defenses at the same time. See the following example.

Example 4. In $\mathcal{F}_9$, $\langle a, c \rangle$, while $(\emptyset, b)$, $(\emptyset, c)$ and (b, d) are DoDs. c is both in the defense $\langle a, c \rangle$ and in the DoD $(\emptyset, c)$. When $\langle a, c \rangle$ is accepted, c is accepted.

$$\mathcal{F}_9: \quad a \rightarrow b \circlearrowright \rightarrow c \rightarrow d$$
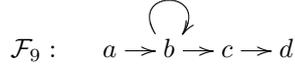
Note also that in Definition 4 when β is attacked by a self-attacked argument, it is a DoD. Consider $\mathcal{F}_8$ in Example 3. $\langle a_{11}, a_{13} \rangle$ is a defense. If $(\emptyset, a_{15})$ is not a DoD, then there is no DoD to prevent the acceptance of $\langle a_{11}, a_{13} \rangle$.

Let $\arg(\mathcal{F}^{\mathrm{DoD}}) = \{\alpha, \beta \mid (\alpha, \beta) \in \mathcal{F}^{\mathrm{DoD}}\} \cup \{\beta \mid (\emptyset, \beta) \in \mathcal{F}^{\mathrm{DoD}}\}$ be the set of arguments involved in $\mathcal{F}^{\mathrm{DoD}}$. Let $\mathrm{def}(\mathcal{F}^{\mathrm{DEF}})^{\rightarrow}$ be the set of arguments attacked by $\mathrm{def}(\mathcal{F}^{\mathrm{DEF}})$. We have the following proposition.

Proposition 1. *Let $\mathcal{F} = (\mathcal{A}, \rightarrow)$ be an argument graph. It holds that $\mathcal{A} = \arg(\mathcal{F}^{\mathrm{DoD}}) \cup \mathrm{def}(\mathcal{F}^{\mathrm{DEF}}) \cup \mathrm{def}(\mathcal{F}^{\mathrm{DEF}})^{\rightarrow}$.*

This proposition states that arguments in $\mathcal{F}$ are equivalent to the union of the arguments in defenses, arguments in defeaters of defenses, and the arguments attacked by the arguments in defenses.

Let $\mathcal{F}^{\mathrm{d}} = \mathcal{F}^{\mathrm{DEF}} \cup \mathcal{F}^{\mathrm{DoD}}$ be the set of defenses and their defeaters. The attack relation between the elements of $\mathcal{F}^{\mathrm{d}}$ can be identified according to the attack relation between the arguments involved. For convenience, we also write $[x, \beta]$ to denote a defense $\langle x, \beta \rangle$ or a defeater of defenses (x, β) where $x \in \mathcal{A} \cup \{\emptyset\}$ and $\beta \in \mathcal{A}$. Formally we have the following definition.

Definition 5 (Attacks between defenses and their defeaters). *For all $[x, \alpha], [y, \beta] \in \mathcal{F}^{\mathrm{d}}$ where $x, y \in \mathcal{A} \cup \{\emptyset\}$ and $\alpha, \beta \in \mathcal{A}$, we say that $[x, \alpha]$ attacks $[y, \beta]$, denoted as $[x, \alpha] \rightarrow^{\mathrm{d}} [y, \beta]$ iff $x \rightarrow y$, $x \rightarrow \beta$, $\alpha \rightarrow y$, or $\alpha \rightarrow \beta$.*

The set of attacks between defenses and their defeaters is denoted as $\to^{\mathrm{d}}$. Given $D \subseteq \mathcal{F}^{\mathrm{d}}$ and $X \in \mathcal{F}^{\mathrm{d}}$, we use $D \to^{\mathrm{d}} X$ to denote that $\exists Y \in D$ such that $Y \to^{\mathrm{d}} X$.

Since the status of a defense is determined by that of other defenses and affected by defeaters of defenses through the attacks between them, to evaluate the status of normal defenses, one possible way is to use *defense graph*, which is defined as follows.

Definition 6 (Defense graph). *Let $\mathcal{F} = (\mathcal{A}, \to)$ be an argument graph. Let $\mathcal{F}^{\mathrm{d}} = \mathcal{F}^{\mathrm{DEF}} \cup \mathcal{F}^{\mathrm{DoD}}$. A defense graph w.r.t. $\mathcal{F}$, denoted $\mathrm{d}(\mathcal{F})$, is defined as follows.*

$$\mathrm{d}(\mathcal{F}) = (\mathcal{F}^{\mathrm{d}}, \to^{\mathrm{d}}) \tag{2}$$

A defense graph can be viewed as a kind of meta-argumentation [12].

Example 5. The defense graph of $\mathcal{F}_6$ is as follows.
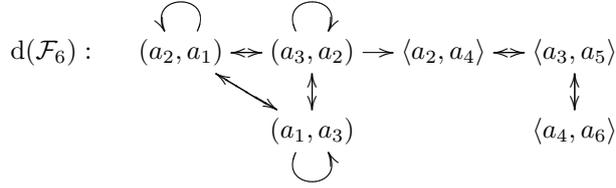
4   Defense semantics

In a defense graph $\mathrm{d}(\mathcal{F}) = (\mathcal{F}^{\mathrm{d}}, \to^{\mathrm{d}})$, nodes are defenses and/or defeaters of defenses, rather than arguments in the corresponding argument graph $\mathcal{F}$. So, when applying classical semantics to $\mathrm{d}(\mathcal{F})$, we get a set of extensions, each of which is a set of defenses. By slightly modifying the definition for classical semantics, defense semantics can be defined as follows.

Definition 7 (Defense semantics). *Defense semantics is a function Σ mapping each defense graph to a set of extensions of defenses. Given a defense graph $\mathrm{d}(\mathcal{F}) = (\mathcal{F}^{\mathrm{d}}, \to^{\mathrm{d}})$ where $\mathcal{F}^{\mathrm{d}} = \mathcal{F}^{\mathrm{DEF}} \cup \mathcal{F}^{\mathrm{DoD}}$, let $D \subseteq \mathcal{F}^{\mathrm{DEF}}$. We have:*

- *D is conflict-free iff $\nexists X, Y \in D$ such that $X \to^{\mathrm{d}} Y$.*
- *$X \in \mathcal{F}^{\mathrm{DEF}}$ is defended by D iff for all $Y \in \mathcal{F}^{\mathrm{d}}$, if $Y \to^{\mathrm{d}} X$, then $\exists Z \in D$ such that $Z \to^{\mathrm{d}} Y$.*
- *D is admissible iff D is conflict-free and each member in D is defended by D.*
- *D is a complete extension of defenses iff D is admissible, and each member in $\mathcal{F}^{\mathrm{DEF}}$ that is defended by D is in D.*
- *D is the grounded extension of defenses iff D is the minimal (w.r.t. set-inclusion) complete extension of defenses.*
- *D is a preferred extension of defenses iff D is a maximal (w.r.t. set-inclusion) complete extension of defenses.*
- *D is a stable extension of defenses iff D is conflict-free, and $\forall X \in \mathcal{F}^{\mathrm{d}} \setminus D$, $D \to^{\mathrm{d}} X$.*

The set of complete, grounded, preferred, and stable extensions of defenses of $d(\mathcal{F})$ is denoted as $CO(d(\mathcal{F}))$, $GR(d(\mathcal{F}))$, $PR(d(\mathcal{F}))$ and $ST(d(\mathcal{F}))$ respectively.

Note that the notion of defense semantics is similar to that of classical semantics. The only difference is that in a defense graph, we differentiate two kinds of nodes: defenses and defeaters of defenses. The former can be included in extensions, while the latter are only used to prevent the acceptance of some defenses.

Now, let us consider some properties of the defense semantics of an argument graph.

The first property is about the relation between defense semantics and classical semantics. Let $D \in \Sigma(d(\mathcal{F}))$ be a Σ-extension of $d(\mathcal{F})$. Now the question is whether the set of defenders and defendees in D is a σ-extension of $\mathcal{F}$. In order to verify this property, technically, we first present the follow lemma. The lemma states that for all $\langle \alpha, \beta \rangle \in D$, if α is attacked by an argument $\gamma \in \mathcal{A}$, then $\exists \eta \in \mathrm{def}(D)$ such that η attacks γ.

Lemma 1 *For all $D \in \Sigma(d(\mathcal{F}))$, for all $\langle x, y \rangle \in D$, for all $\gamma \in \mathcal{A}$, if $\gamma \to x$ or $\gamma \to y$, then $\mathrm{def}(D) \to \gamma$.*

Example 6. Consider $d(\mathcal{F}_1)$ below. Under complete semantics, $CO(d(\mathcal{F}_1)) = \{D_1, D_2, D_3\}$ where $D_1 = \{\}$, $D_2 = \{\langle a, c_2 \rangle, \langle c_2, c_3 \rangle, \langle c_3, a \rangle\}$, $D_3 = \{\langle b, c_4 \rangle, \langle c_1, b \rangle, \langle c_4, c_1 \rangle\}$. Take D_2 and $\langle a, c_2 \rangle$ in D_2 as an example. $\mathrm{def}(D_2) = \{a, c_2, c_3\}$. For a being attacked by c_4, and c_2 being attacked by c_1, it holds that $\mathrm{def}(D_2) \to c_4$ and $\mathrm{def}(D_2) \to c_1$.
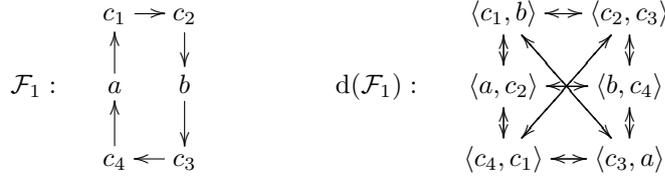
Based on Lemma 1, under complete semantics, we have the following theorem.

Theorem 1. *For all $D \in CO(d(\mathcal{F}))$, $\mathrm{def}(D) \in co(\mathcal{F})$.*

Theorem 1 makes clear that for each complete defense extension D of a defense graph, there exists a complete argument extension E of the corresponding argument graph such that E is equal to $\mathrm{def}(D)$. On the other hand, the following theorem says that for each complete argument extension E of an argument graph, there exists a complete defense extension D of the corresponding defense graph such that $D = d(E)$ where $d(E) = \{\langle x, y \rangle \in \mathcal{F}^{\mathrm{DEF}} \mid x \in E \cup \{\emptyset\}, y \in E\}$.

Theorem 2. *For all $E \in co(\mathcal{F})$, $d(E) \in CO(d(\mathcal{F}))$.*

The relation between argument extensions and defense extensions under other semantics is presented in the following corollaries.

Corollary 1. *For all $\Sigma \in \{GR, PR, ST\}$, it holds that for all $D \in \Sigma(d(\mathcal{F}))$, $\mathrm{def}(D) \in \sigma(\mathcal{F})$.*

Proofs for Lemma 1, Theorem 1, 2 and Corollary 1 are presented in the Appendix. In the following theorems and corollaries, when we say $\Sigma \in \{\mathrm{CO}, \mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$, σ is referred to co, gr, pr and st, correspondingly. Meanwhile, when we say $\sigma \in \{\mathrm{co}, \mathrm{gr}, \mathrm{pr}, \mathrm{st}\}$, Σ is referred to CO, GR, PR and ST, correspondingly.

Corollary 2. *For all $\Sigma \in \{\mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$ it holds that for all $E \in \sigma(\mathcal{F})$, $\mathrm{d}(E) \in \Sigma(\mathrm{d}(\mathcal{F}))$.*

Proof. Under grounded semantics, we need to verify that $\mathrm{d}(E)$ is minimal (w.r.t. set-inclusion). Assume the contrary. Then $\exists D' \subsetneq \mathrm{d}(E)$ such that D' is a grounded extension. According to theorem 1, $\mathrm{def}(D')$ is a complete extension. It follows that $\mathrm{def}(D') \subsetneq \mathrm{def}(\mathrm{d}(E)) = E$. It turns out that E is not a grounded extension. Contradiction.

Under preferred semantic, it is easy to verify that $\mathrm{d}(E)$ is maximal (w.r.t. set-inclusion).

Under stable semantics, we need to prove that for all $[x, \alpha] \in \mathcal{F}^{\mathrm{d}} \setminus \mathrm{d}(E)$: $\mathrm{d}(E) \to [x, \alpha]$. Assume the contrary. Then, $\exists[x, \alpha] \in \mathcal{F}^{\mathrm{d}} \setminus \mathrm{d}(E)$ such that $\mathrm{d}(E)$ does not attack $[x, \alpha]$. So, E does not attack x and α. Since E is stable, it holds that $\{x, \alpha\} \setminus \{\emptyset\} \subseteq E$. So, $[x, \alpha] \in E$. Contradiction.

Theorems 1 and 2 and Corollaries 1 and 2 describe the relation between argument extensions and defense extensions under various semantics. This relation can be further described by two equations in the following two corollaries.

First, by overloading the notation, let $\mathrm{d}(\sigma(\mathcal{F})) = \{\mathrm{d}(E) \mid E \in \sigma(\mathcal{F})\}$, where $\sigma \in \{\mathrm{co}, \mathrm{gr}, \mathrm{pr}, \mathrm{st}\}$.

Corollary 3. *For all $\sigma \in \{\mathrm{co}, \mathrm{pr}, \mathrm{gr}, \mathrm{st}\}$, it holds that $\mathrm{d}(\sigma(\mathcal{F})) = \Sigma(\mathrm{d}(\mathcal{F}))$.*

Proof. For all $\mathrm{d}(E) \in \mathrm{d}(\sigma(\mathcal{F}))$, according to Theorem 2 and Corollary 2, $\mathrm{d}(E) \in \Sigma(\mathrm{d}(\mathcal{F}))$. For all $D \in \Sigma(\mathrm{d}(\mathcal{F}))$, according to Theorem 1 and Corollary 1, $\mathrm{def}(D) \in \sigma(\mathcal{F})$. Since $\mathrm{d}(\mathrm{def}(D)) = \{\langle \beta, \alpha \rangle \in \mathcal{F}^{\mathrm{DEF}} \mid \alpha, \beta \in \mathrm{def}(D)\} \cup \{\langle \emptyset, \alpha \rangle \in \mathcal{F}^{\mathrm{DEF}} \mid \alpha \in \mathrm{def}(D)\} = D$, it holds that $D \in \mathrm{d}(\sigma(\mathcal{F}))$.

Example 7. Consider $\mathcal{F}_{10}$ and $\mathrm{d}(\mathcal{F}_{10})$ below. Under complete semantics, we have:

- $\mathrm{co}(\mathcal{F}_{10}) = \{E_1, E_2\}$, where $E_1 = \{\}$, $E_2 = \{b\}$;
- $\mathrm{d}(\mathrm{co}(\mathcal{F}_{10})) = \{\mathrm{d}(E_1), \mathrm{d}(E_2)\}$, where $\mathrm{d}(E_1) = \{\}$, $\mathrm{d}(E_2) = \{\langle b, b \rangle\}$;
- $\mathrm{CO}(\mathrm{d}(\mathcal{F}_{10})) = \{D_1, D_2\}$, where $D_1 = \{\}$, $D_2 = \{\langle b, b \rangle\}$.

So, it holds that $\mathrm{d}(\mathrm{co}(\mathcal{F}_{10})) = \mathrm{CO}(\mathrm{d}(\mathcal{F}_{10}))$.
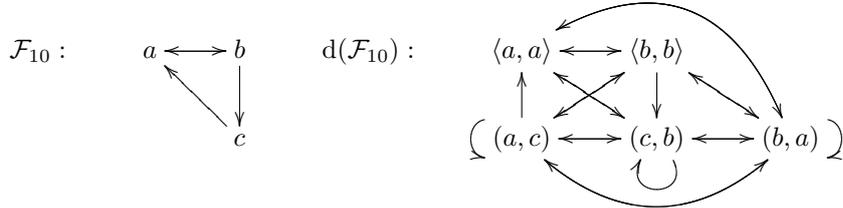
Second, by overloading the notation, let $\mathrm{def}(\Sigma(\mathrm{d}(\mathcal{F}))) = \{\mathrm{def}(D) \mid D \in \Sigma(\mathrm{d}(\mathcal{F}))\}$, where $\Sigma \in \{\mathrm{CO}, \mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$.

Corollary 4. *For all $\Sigma \in \{\mathrm{CO}, \mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$, it holds that $\sigma(\mathcal{F}) = \mathrm{def}(\Sigma(\mathrm{d}(\mathcal{F})))$.*

Proof. For all $E \in \sigma(\mathcal{F})$, according to Theorem 2 and Corollary 2, $\mathrm{d}(E) \in \Sigma(\mathrm{d}(\mathcal{F}))$. Since $\mathrm{def}(\mathrm{d}(E)) = E$, $E \in \mathrm{def}(\Sigma(\mathrm{d}(\mathcal{F})))$. For all $\mathrm{def}(D) \in \mathrm{def}(\Sigma(\mathrm{d}(\mathcal{F})))$, since $D \in \Sigma(\mathrm{d}(\mathcal{F}))$, according to Theorem 1 and Corollary 1, $\mathrm{def}(D) \in \mathrm{co}(\mathcal{F})$.

Example 8. Under compete semantics, continue Example 7, $\mathrm{def}(\mathrm{CO}(\mathrm{d}(\mathcal{F}_{10}))) = \{\mathrm{def}(D_1), \mathrm{def}(D_2)\}$, where $\mathrm{def}(D_1) = \{\}, \mathrm{def}(D_2) = \{b\}$. It holds that $\mathrm{co}(\mathcal{F}_{10}) = \mathrm{def}(\mathrm{CO}(\mathrm{d}(\mathcal{F}_{10})))$.

The second property formulated in Theorems 3 and 4 is about the equivalence of argument graphs under defense semantics, called *defense equivalence of argument graphs*.

Definition 8 (Defense equivalence of AFs). *Let $\mathcal{F}$ and $\mathcal{G}$ be two argument graphs. $\mathcal{F}$ and $\mathcal{G}$ are of defense equivalence w.r.t. a semantics Σ, denoted as $\mathcal{F} \equiv_{\mathrm{d}}^{\Sigma} \mathcal{G}$, iff $\Sigma(\mathrm{d}(\mathcal{F})) = \Sigma(\mathrm{d}(\mathcal{G}))$.*

Concerning the relation between defense equivalence and standard equivalence of argument graphs, we have the following theorem.

Theorem 3. *Let $\mathcal{F}$ and $\mathcal{G}$ be two argument graphs, and $\Sigma \in \{\mathrm{CO}, \mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$ be a semantics. If $\mathcal{F} \equiv_{\mathrm{d}}^{\Sigma} \mathcal{G}$, then $\mathcal{F} \equiv^{\sigma} \mathcal{G}$.*

Proof. If $\mathcal{F} \equiv_{\mathrm{d}}^{\Sigma} \mathcal{G}$, then $\Sigma(\mathrm{d}(\mathcal{F})) = \Sigma(\mathrm{d}(\mathcal{G}))$. According to Corollary 4, it follows that $\sigma(\mathcal{F}) = \mathrm{def}(\Sigma(\mathrm{d}(\mathcal{F}))) = \mathrm{def}(\Sigma(\mathrm{d}(\mathcal{G}))) = \sigma(\mathcal{G})$. Since $\sigma(\mathcal{F}) = \sigma(\mathcal{G})$, $\mathcal{F} \equiv^{\sigma} \mathcal{G}$.

Note that $\mathcal{F} \equiv^{\sigma} \mathcal{G}$ does not imply $\mathcal{F} \equiv_{\mathrm{d}}^{\Sigma} \mathcal{G}$ in general. Consider the following example under complete semantics.

Example 9. Since $\mathrm{co}(\mathcal{F}_3) = \mathrm{co}(\mathcal{F}_4) = \{\{a, c\}\}$, it holds that $\mathcal{F}_3 \equiv^{\mathrm{co}} \mathcal{F}_4$. Since $\mathrm{CO}(\mathrm{d}(\mathcal{F}_3)) = \{\{\langle \emptyset, a \rangle\}, \langle a, c \rangle\}$ and $\mathrm{CO}(\mathrm{d}(\mathcal{F}_4)) = \{\{\langle \emptyset, a \rangle\}, \langle \emptyset, c \rangle\}$, $\mathrm{CO}(\mathrm{d}(\mathcal{F}_3)) \neq \mathrm{CO}(\mathrm{d}(\mathcal{F}_4))$. So, it is not the case that $\mathcal{F}_3 \equiv_{\mathrm{d}}^{\mathrm{CO}} \mathcal{F}_4$.

$$\mathcal{F}_3: \quad a \twoheadrightarrow b \twoheadrightarrow c \qquad \mathrm{d}(\mathcal{F}_3): \quad \langle \emptyset, a \rangle \quad \langle a, c \rangle$$

$$\mathcal{F}_4: \quad a \twoheadrightarrow b \quad c \qquad \mathrm{d}(\mathcal{F}_4): \quad \langle \emptyset, a \rangle \quad \langle \emptyset, c \rangle$$

About the relation between defense equivalence and strong equivalence of argument graphs, under complete semantics, we have the following lemma and theorem.

Lemma 2 *It holds that $\mathrm{CO}(\mathrm{d}(\mathcal{F})) = \mathrm{CO}(\mathrm{d}(\mathcal{F}^{\mathrm{ck}}))$.*

Proof. According to Corollary 3, $\mathrm{d}(\mathrm{co}(\mathcal{F})) = \mathrm{CO}(\mathrm{d}(\mathcal{F})), \mathrm{d}(\mathrm{co}(\mathcal{F}^{\mathrm{ck}})) = \mathrm{CO}(\mathrm{d}(\mathcal{F}^{\mathrm{ck}}))$. Since $\mathrm{co}(\mathcal{F}) = \mathrm{co}(\mathcal{F}^{\mathrm{ck}})$, it holds that $\mathrm{CO}(\mathrm{d}(\mathcal{F})) = \mathrm{d}(\mathrm{co}(\mathcal{F})) = \mathrm{d}(\mathrm{co}(\mathcal{F}^{\mathrm{ck}})) = \mathrm{CO}(\mathrm{d}(\mathcal{F}^{\mathrm{ck}}))$.

Theorem 4. *Let $\mathcal{F}$ and $\mathcal{G}$ be two argument graphs. If $\mathcal{F} \equiv_s^{\mathrm{co}} \mathcal{G}$, then $\mathcal{F} \equiv_{\mathrm{d}}^{\mathrm{CO}} \mathcal{G}$.*

Proof. If $\mathcal{F} \equiv_s^{\mathrm{co}} \mathcal{G}$, then $\mathcal{F}^{\mathrm{ck}} = \mathcal{G}^{\mathrm{ck}}$. So, $\mathrm{CO}(\mathrm{d}(\mathcal{F}^{\mathrm{ck}})) = \mathrm{CO}(\mathrm{d}(\mathcal{G}^{\mathrm{ck}}))$. According to Lemma 2, $\mathrm{CO}(\mathrm{d}(\mathcal{F})) = \mathrm{CO}(\mathrm{d}(\mathcal{F}^{\mathrm{ck}})), \mathrm{CO}(\mathrm{d}(\mathcal{G})) = \mathrm{CO}(\mathrm{d}(\mathcal{G}^{\mathrm{ck}}))$. So, we have $\mathrm{CO}(\mathrm{d}(\mathcal{F})) = \mathrm{CO}(\mathrm{d}(\mathcal{G}))$, i.e., $\mathcal{F} \equiv_{\mathrm{d}}^{\mathrm{CO}} \mathcal{G}$.

Note that $\mathcal{F} \equiv_{\mathrm{d}}^{\mathrm{CO}} \mathcal{G}$ does not imply $\mathcal{F} \equiv_{s}^{\mathrm{co}} \mathcal{G}$ in general. Consider the following example.

Example 10. Since $\mathrm{CO}(\mathrm{d}(\mathcal{F}_3)) = \mathrm{CO}(\mathrm{d}(\mathcal{F}_{11})) = \{\{\langle \o, a\rangle\}, \langle a, c\rangle\}$, $\mathcal{F}_3 \equiv_{\mathrm{d}}^{\mathrm{CO}} \mathcal{F}_{11}$. However, since $\mathcal{F}_3^{\mathrm{ck}} \neq \mathcal{F}_{11}^{\mathrm{ck}}$, $\mathcal{F}_3 \not\equiv_{s}^{\mathrm{co}} \mathcal{F}_{11}$.

$$\mathcal{F}_{11}: \quad a \rightarrowtail b \rightarrowtail c \qquad \mathrm{d}(\mathcal{F}_{11}): \quad \langle \o, a\rangle \quad \langle a, c\rangle$$
$$\searrow \ d \ \nearrow$$

5   Encoding reasons for accepting arguments

Defense semantics can be used to encode reasons for accepting arguments. Consider the following example.

Example 11. $\mathrm{CO}(\mathrm{d}(\mathcal{F}_{12})) = \{D_1, D_2\}$, where $D_1 = \{\langle b, b\rangle, \langle b, d\rangle, \langle g, d\rangle, \langle e, g\rangle, \langle \o, e\rangle\}$, $D_2 = \{\langle a, a\rangle, \langle g, d\rangle, \langle e, g\rangle, \langle \o, e\rangle\}$. One way to capture reasons for accepting arguments is to relate each reason to an extension of defenses. For instance, concerning the reasons for accepting d w.r.t. D_1, we differentiate the following reasons:

- Direct reason: accepting $\{b, g\}$ is a direct reason for accepting d. This reason can be identified in terms of defenses $\langle b, d\rangle$ and $\langle g, d\rangle$ in D_1.
- Root reason: accepting $\{e, b\}$ is a root reason for accepting d, in the sense that each element of a root reason is either an initial argument, or an argument without further defenders except itself. This reason can be identified by means of viewing each defense as a binary relation, and allowing this relation to be transitive. Given $\langle e, g\rangle$ and $\langle g, d\rangle$ in D_1, we have $\langle e, d\rangle$. Since e is an initial argument, it is an element of the root reason. Given $\langle b, d\rangle$ in D_1, since b's defender is b itself, b is an element of the root reason.

The informal notions in Example 11 are formulated as follows.

Definition 9 (Direct reasons for accepting arguments). *Let $\mathcal{F} = (\mathcal{A}, \rightarrow)$ be an argument graph. Direct reasons for accepting arguments in $\mathcal{F}$ under a semantics Σ is a function, denoted $\mathrm{dr}_{\Sigma}^{\mathcal{F}}$, mapping from $\mathcal{F}$ to sets of arguments, such that for all $\alpha \in \mathcal{A}$,*

$$\mathrm{dr}_{\Sigma}^{\mathcal{F}}(\alpha) = \{\mathcal{DR}(\alpha, D) \mid D \in \Sigma(\mathrm{d}(\mathcal{F}))\} \tag{3}$$

where $\mathcal{DR}(\alpha, D) = \{\beta \mid \langle \beta, \alpha\rangle \in D\}$, if α is not an initial argument; otherwise, $\mathcal{DR}(\alpha, D) = \{\o\}$.

Example 12. Continue Example 11. According to Definition 9, $\mathrm{dr}_{\mathrm{CO}}^{\mathcal{F}_{12}}(d) = \{R_1, R_2\}$, where $R_1 = \{b, g\}$, $R_2 = \{g\}$. $\mathrm{dr}_{\mathrm{CO}}^{\mathcal{F}_{12}}(f) = \{R_3, R_4\}$, where $R_3 = R_4 = \{\}$.

For all $D \in \Sigma(\mathrm{d}(\mathcal{F}))$, we view D as a transitive relation, and let D^+ be the transitive closure of D.

Definition 10 (Root reasons for accepting arguments). *Let $\mathcal{F} = (\mathcal{A}, \rightarrow)$ be an argument graph. Root reasons for accepting arguments in $\mathcal{F}$ under a semantics Σ is a function, denoted $\mathrm{rr}_{\Sigma}^{\mathcal{F}}$, mapping from $\mathcal{F}$ to sets of arguments, such that for all $\alpha \in \mathcal{A}$,*

$$\mathrm{rr}_{\Sigma}^{\mathcal{F}}(\alpha) = \{\mathcal{RR}(\alpha, D) \mid D \in \Sigma(\mathrm{d}(\mathcal{F}))\} \tag{4}$$

where $\mathcal{RR}(\alpha, D) = \{\beta \in \mathcal{A} \mid \langle \beta, \beta \rangle \in D^+, \beta = \alpha\} \cup \{\beta \in \mathcal{A} \mid (\langle \beta, \alpha \rangle \in D^+), (\langle \beta, \beta \rangle \in D^+ \vee \beta^{\leftarrow} = \emptyset)\}$, if α is not an initial argument; otherwise, $\mathcal{RR}(\alpha, D) = \{\emptyset\}$.

According Definition 10, we say that a set of arguments $\mathcal{RR}(\alpha, D)$ is a root reason of an argument α if and only if for all $\beta \in \mathcal{RR}(\alpha, D)$, β is either equal to α when α (partially) defends itself directly or indirectly through a transitive relation of defenses in D, or an initial argument, or an argument that can (partially) defend itself directly or indirectly.

Example 13. Continue Example 11. $D_1^+ = D_1 \cup \{\langle e, d \rangle, \langle \emptyset, g \rangle, \langle \emptyset, d \rangle\}$; $D_2^+ = D_2 \cup \{\langle e, d \rangle, \langle \emptyset, g \rangle, \langle \emptyset, d \rangle\}$. According to Definition 10, $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}_{12}}(d) = \{R_1, R_2\}$, where $R_1 = \{b, e\}$, $R_2 = \{e\}$. $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}_{12}}(f) = \{R_3, R_4\}$, where $R_3 = R_4 = \{\}$.

Motivated by the first example in Section 1 (regarding $\mathcal{F}_1$), based on the notion of root reasons, we propose as follows a notion of root equivalence of AFs.

Definition 11 (Root equivalence of AFs). *Let $\mathcal{F} = (\mathcal{A}_1, \rightarrow_1)$ and $\mathcal{H} = (\mathcal{A}_2, \rightarrow_2)$ be two argument graphs. For all $B \subseteq \mathcal{A}_1 \cap \mathcal{A}_2$, if $B \neq \emptyset$, we say that $\mathcal{F}$ and $\mathcal{H}$ are equivalent w.r.t. the root reasons for accepting B under semantics Σ, denoted $\mathcal{F}|B \equiv_{rr}^{\Sigma} \mathcal{H}|B$, iff for all $\alpha \in B$, $\mathrm{rr}_{\Sigma}^{\mathcal{F}}(\alpha) = \mathrm{rr}_{\Sigma}^{\mathcal{H}}(\alpha)$.*

When $B = \mathcal{A}_1 = \mathcal{A}_2$, we write $\mathcal{F} \equiv_{rr}^{\Sigma} \mathcal{H}$ for $\mathcal{F}|B \equiv_{rr}^{\Sigma} \mathcal{H}|B$.

Example 14. Consider $\mathcal{F}_1$ and $\mathcal{F}_2$ in Section 1 again. Under complete semantics, $\mathrm{CO}(\mathrm{d}(\mathcal{F}_1)) = \{D_1, D_2, D_3\}$ where $D_1 = \{\}$, $D_2 = \{\langle a, c_2 \rangle, \langle c_2, c_3 \rangle, \langle c_3, a \rangle\}$, $D_3 = \{\langle b, c_4 \rangle, \langle c_1, b \rangle, \langle c_4, c_1 \rangle\}$. $\mathrm{CO}(\mathrm{d}(\mathcal{F}_2)) = \{D_4, D_5, D_6\}$ where $D_4 = \{\}$, $D_5 = \{\langle a, a \rangle\}$, $D_6 = \{\langle b, b \rangle\}$. Let $B = \{a, b\}$. $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}_1}(a) = \{\{\}, \{a\}, \{\}\}$, $\mathrm{rr}_{co}^{\mathcal{F}_2}(a) = \{\{\}, \{a\}, \{\}\}$, $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}_1}(b) = \{\{\}, \{\}, \{b\}\}$, $\mathrm{rr}_{co}^{\mathcal{F}_2}(b) = \{\{\}, \{\}, \{b\}\}$. So, it holds that $\mathcal{F}_1|B =_{rr}^{\mathrm{CO}} \mathcal{F}_2|B$.

Theorem 5. *Let $\mathcal{F} = (\mathcal{A}_1, \rightarrow_1)$ and $\mathcal{H} = (\mathcal{A}_2, \rightarrow_2)$ be two argument graphs. If $\mathcal{F} \equiv_{rr}^{\mathrm{CO}} \mathcal{H}$, then $\mathcal{F} \equiv^{co} \mathcal{H}$.*

Proof. According to Definition 10, the number of extensions of $\mathrm{co}(\mathcal{F})$ is equal to the number of $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}}(\alpha)$, where $\alpha \in \mathcal{A}_1$. Since $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}}(\alpha) = \mathrm{rr}_{\mathrm{CO}}^{\mathcal{H}}(\alpha)$, $\mathcal{A}_1 = \mathcal{A}_2$. Let $\mathrm{rr}_{\mathrm{CO}}^{\mathcal{F}}(\alpha) = \mathrm{rr}_{\mathrm{CO}}^{\mathcal{H}}(\alpha) = \{R_1, \dots, R_n\}$. Let $\mathrm{co}(\mathcal{F}) = \{E_1, \dots, E_n\}$ be the set of extensions of $\mathcal{F}$, where $n \geq 1$. For all $\alpha \in \mathcal{A}_1$, for all R_i, $i = 1, \dots, n$, we have $\alpha \in E_i$ iff

$R_i \neq \{\}$, in that in terms of Definition 10,when $R_i \neq \{\}$, there is a reason to accept α. On the other hand, let $\mathrm{co}(\mathcal{H}) = \{S_1, \ldots, S_n\}$ be the set of extensions of $\mathcal{H}$. For all $\alpha \in \mathcal{A}_2 = \mathcal{A}_1$, for all R_i, $i = 1, \ldots, n$, for the same reason, we have $\alpha \in S_i$ iff $R_i \neq \{\}$. So, it holds that $E_i = S_i$ for $i = 1, \ldots, n$, and hence $\mathrm{co}(\mathcal{F}) = \mathrm{co}(\mathcal{H})$, i.e., $\mathcal{F} \equiv^{\mathrm{co}} \mathcal{H}$.

Note that $\mathcal{F} \equiv^{\mathrm{co}} \mathcal{H}$ does not imply $\mathcal{F} \equiv^{\mathrm{CO}}_{rr} \mathcal{H}$ in general. This can be easily verified by considering $\mathcal{F}_3$ and $\mathcal{F}_4$ in Example 9.

The notion of root equivalence of argument graphs can be used to capture a kind of summarization in the graphs. Consider the following example borrowed from [8].

Example 15. Let $\mathcal{F}_{13} = (\mathcal{A}, \rightarrow)$ and $\mathcal{F}_{13} = (\mathcal{A}', \rightarrow')$, illustrated below. Under complete semantics, $\mathcal{F}_{13}$ is a summarization of $\mathcal{F}_{13}$ in the sense that $\mathcal{A}' \subseteq \mathcal{A}$, and the root reason of each argument in $\mathcal{F}_{13}$ is the same as that of each corresponding argument in $\mathcal{F}_{13}$. More specifically, it holds that $\mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_3) = \mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_3) = \{\{e_1, e_2\}\}$, $\mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_2) = \mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_2) = \{\{\o\}\}$, and $\mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_1) = \mathrm{rr}^{\mathcal{F}_{13}}_{\mathrm{CO}}(e_1) = \{\{\o\}\}$.

$$\mathcal{F}_{13}: \quad e_1 \rightarrowtail a_1 \rightarrowtail a_2 \rightarrowtail o \rightarrowtail e_3 \qquad \mathcal{F}_{13}: \quad e_1 \rightarrowtail o \rightarrowtail e_3$$
$$e_2 \rightarrowtail b_1 \rightarrowtail b_2 \qquad\qquad\qquad\qquad e_2$$

Definition 12 (Summarization of AFs). *Let $\mathcal{F} = (\mathcal{A}_1, \rightarrow_1)$ and $\mathcal{H} = (\mathcal{A}_2, \rightarrow_2)$ be two argument graphs. $\mathcal{F}$ is a summarization of $\mathcal{H}$ under a semantics σ iff $\mathcal{A}_1 \subset \mathcal{A}_2$, and $\mathcal{F}|\mathcal{A}_1 \equiv^{\sigma}_{rr} \mathcal{H}|\mathcal{A}_1$.*

Now, a property of summarization of argument graphs under complete semantics is as follows.

Theorem 6. *Let $\mathcal{F} = (\mathcal{A}_1, \rightarrow_1)$ and $\mathcal{H} = (\mathcal{A}_2, \rightarrow_2)$ be two argument graphs. If $\mathcal{F}$ is a summarization of $\mathcal{H}$ under complete semantics* CO, *then* $\mathrm{CO}(\mathcal{F}) = \{E \cap \mathcal{A}_2 \mid E \in \mathrm{CO}(\mathcal{H})\}$.

Proof. Let $\mathrm{co}(\mathcal{F}) = \{E_1, \ldots, E_n\}$, $\mathrm{co}(\mathcal{H}) = \{S_1, \ldots, S_n\}$. According to the proof of Theorem 5, $E_1 = S_1 \cap \mathcal{A}_2$. Therefore, we have $\mathrm{co}(\mathcal{F}) = \{E \cap \mathcal{A}_2 \mid E \in \mathrm{co}(\mathcal{H})\}$.

The property looks similar to that of directionality of argumentation [13]. However, they are conceptually different. Specifically, it is said that if a semantics σ satisfies the property of directionality if and only if $\forall \mathcal{F} = (\mathcal{A}, \rightarrow)$, $\forall U \subseteq \mathcal{A}$, if U is an unattacked set, then $\sigma(\mathcal{F} \downarrow U) = \{E \cap U \mid E \in \sigma(\mathcal{F})\}$ where $\mathcal{F} \downarrow U = (U, \rightarrow \cap (U \times U))$. So, the property of directionality is about the relation between an argument graph and its subgraph induced by an unattacked set of arguments. By contrast, the property of summarization of argument graphs is about the relation between two root equivalent argument graphs.

6   Conclusions

In this paper, we have proposed a defense semantics of argumentation based on a novel notion of defense graphs, and used it to encode reasons for accepting arguments. By introducing two new kinds of equivalence relation between argument graphs, i.e., defense

equivalence and root equivalence, we have shown that defense semantics can be used to capture the equivalence of argument graphs from the perspective of reasons for accepting arguments. In addition, we have defined a notion of summarization of argument graphs by exploiting root equivalence.

Under complete semantics, defense equivalence is located inbetween strong and standard equivalence. It is interesting to further investigate its position in the so-called equivalence zoo where further equivalence notions inbetween the two extremal versions are compared too [14], and to study how defense equivalence, root equivalence and strong equivalence are related. We will present this part of work in an extended version of the present paper.

Since defense semantics explicitly represents a defense relation in extensions and can be used to encoded reasons for accepting arguments, it provides a new way to investigate topics such as summarization in argumentation [8], dynamics of argumentation [9, 15, 10], dialogical argumentation [16, 17], etc. Further work on these topics is promising. Meanwhile, it might be interesting to study defense semantics beyond Dung's argumentation, including ADFs [18], bipolar frameworks [19], structured argumentation [20], etc. In particular, it would be interesting to apply defense semantics to modeling the explanation of why a conclusion can be reached. In [21], in order to increase the trust of the users for the Semantic Web applications, a system was proposed to automatically generate an explanation for every answer about why the answer has been produced. The notion of proof trace in [21] for explanation is closer to the notion of support relation between arguments. So, combining the defense relation (which is based on attack relation) and support relation would be useful to model the explanation of conclusions of a structured argumentation system.

Acknowledgements

References

1. Dunne, P.E., Dvorák, W., Linsbichler, T., Woltran, S.: Characteristics of multiple viewpoints in abstract argumentation. Artif. Intell. **228** (2015) 153–178
2. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. The Knowledge Engineering Review **26**(4) (2011) 365–410
3. Charwat, G., Dvorák, W., Gaggl, S.A., Wallner, J.P., Woltran, S.: Methods for solving reasoning problems in abstract argumentation - A survey. Artif. Intell. **220** (2015) 28–63
4. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77**(2) (1995) 321–358

5. Oikarinen, E., Woltran, S.: Characterizing strong equivalence for argumentation frameworks. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010. (2010)

6. Baumann, R.: Characterizing equivalence notions for labelling-based semantics. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016. (2016) 22–32

7. Baumann, R., Strass, H.: An abstract logical approach to characterizing strong equivalence in logic-based knowledge representation formalisms. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016. (2016) 525–528

8. Baroni, P., Boella, G., Cerutti, F., Giacomin, M., van der Torre, L., Villata, S.: On the input/output behavior of argumentation frameworks. Artificial Intelligence **217** (2014) 144–197

9. Cayrol, C., de Saint-Cyr, F.D., Lagasquie-Schiex, M.: Change in abstract argumentation frameworks: Adding an argument. J. Artif. Intell. Res. (JAIR) **38** (2010) 49–84

10. Ferretti, E., Tamargo, L.H., García, A.J., Errecalde, M.L., Simari, G.R.: An approach to decision making based on dynamic argumentation systems. Artif. Intell. **242** (2017) 107–131

11. Bench-Capon, T.J.M., Prakken, H., Sartor, G.: Argumentation in legal reasoning. In: Argumentation in Artificial Intelligence. (2009) 363–382

12. Boella, G., Gabbay, D.M., van der Torre, L.W.N., Villata, S.: Meta-argumentation modelling I: methodology and techniques. Studia Logica **93**(2-3) (2009) 297–355

13. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artif. Intell. **171**(10-15) (2007) 675–700

14. Baumann, R., Brewka, G.: The equivalence zoo for Dung-style semantics. J. Log. Comput. (2015) doi.org/10.1093/logcom/exv001.

15. Liao, B.S., Jin, L., Koons, R.C.: Dynamics of argumentation systems: A division-based method. Artif. Intell. **175**(11) (2011) 1790–1814

16. Hunter, A., Thimm, M.: Optimization of dialectical outcomes in dialogical argumentation. Int. J. Approx. Reasoning **78** (2016) 73–102

17. Fan, X., Toni, F.: On the interplay between games, argumentation and dialogues. In: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, May 9-13, 2016. (2016) 260–268

18. Brewka, G., Woltran, S.: Abstract dialectical frameworks. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010. (2010)

19. Cayrol, C., Lagasquie-Schiex, M.: Bipolarity in argumentation graphs: Towards a better understanding. Int. J. Approx. Reasoning **54**(7) (2013) 876–899

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

21. Antoniou, G., Bikakis, A., Dimaresis, N., Genetzakis, M., Georgalis, G., Governatori, G., Karouzaki, E., Kazepis, N., Kosmadakis, D., Kritsotakis, M., Lilis, G., Papadogiannakis, A., Pediaditis, P., Terzakis, C., Theodosaki, R., Zeginis, D.: Proof explanation for a nonmonotonic semantic web rules language. Data & Knowledge Engineering **64**(3) (2008) 662–687

Appendix

1. Proof of Lemma 1

Proof. For all $\Sigma \in \{\mathrm{CO}, \mathrm{GR}, \mathrm{PR}, \mathrm{ST}\}$, it holds that $D \in \Sigma(\mathrm{d}(\mathcal{F}))$ is a complete extension.

With respect to γ there are the following four possible cases. Let us analyze them one by one.

First, γ is initial. In this case, $\langle x, y \rangle$ is attacked by $\langle \varnothing, \gamma \rangle$ that is unattacked. So, D cannot defend $\langle x, y \rangle$, contradicting D being a complete extension.

Second, γ is self-attacked. In this case, $(\varnothing, \gamma) \in \mathcal{F}^{\mathrm{DoD}}$, and $(\varnothing, \gamma) \rightarrow^{\mathrm{d}} \langle x, y \rangle$. Since $\langle x, y \rangle$ is defended by D, $\exists \langle \eta, \eta' \rangle \in D$ such that $\eta \rightarrow \gamma$ or $\eta' \rightarrow \gamma$. In other word, it holds that $\mathrm{def}(D) \rightarrow \gamma$.

Third, γ is attacked by $\eta \in \mathcal{A} \setminus \{\gamma\}$, there are the following situations:

– η is initial or all attackers of η are attacked by $\mathrm{def}(D)$: In this case, η does not attack x or y. Otherwise, $\langle x, y \rangle \notin D$. Contradiction. Meanwhile, since $\{\langle \eta, x \rangle\} \cup D$ (reps., $\{\langle \eta, y \rangle\} \cup D$) is conflict-free, and D defends $\langle \eta, x \rangle$ (reps., $\{\langle \eta, y \rangle\}$). Since D is complete, $\langle \eta, x \rangle \in D$ (reps., $\langle \eta, y \rangle \in D$). So, it holds that $\mathrm{def}(D) \rightarrow \gamma$.
– η is self-attacked. In this case, $(\varnothing, \gamma) \in \mathcal{F}^{\mathrm{DoD}}$. According to the second point above, it holds that $\mathrm{def}(D) \rightarrow \gamma$.
– η is attacked by $\eta' \in \mathcal{A} \setminus \{\eta\}$ such that η' is not attacked by $\mathrm{def}(D)$: In this case, $(\eta', \gamma) \in \mathcal{F}^{\mathrm{d}}$, and $(\eta', \gamma) \rightarrow^{\mathrm{d}} \langle x, y \rangle$. Since $\langle x, y \rangle$ is defended by D, $\exists \langle \theta, \theta' \rangle \in D$ such that $\langle \theta, \theta' \rangle \rightarrow^{\mathrm{d}} (\eta', \gamma)$. Since η' is not attacked by θ or θ', γ is attacked by θ or θ'. In other words, it holds that $\mathrm{def}(D) \rightarrow \gamma$.

2. Proof of Theorem 1

Proof. Let $E = \mathrm{def}(D)$. Under complete semantics, we need to prove: 1) E is conflict-free, 2) E defends each member of E, and 3) each argument in $\mathcal{A}$ that is defended by E is in E. Details:

– For all $\alpha, \beta \in E$, α and β are defenders or defendees of defenses in D. Since D is conflict-free, according to Definition 5, it is obvious that E is conflict-free.
– For all $\alpha \in E$, $\exists \langle x, \alpha \rangle \in D$ or $\langle \alpha, x \rangle \in D$ where $x \in \mathcal{A} \cup \{\varnothing\}$. For all $\gamma \in \mathcal{A}$, if γ attacks α, according to Lemma 2, $E \rightarrow \gamma$. So, α is defended by E.
– For all $\alpha \in \mathcal{A}$, if α is defended by E, we have the following possible cases:
  • α is unattacked: In this case, $\langle \varnothing, \alpha \rangle$ is in D. That is, $\alpha \in E$.
  • α is attacked by some arguments in $\mathcal{F}$: For all $\gamma \in \alpha^{\leftarrow}$, since α is defended by E, there exists $\delta \in E$ such that $\delta \rightarrow \gamma$. It follows that $\langle \delta, \alpha \rangle \in \mathcal{F}^{\mathrm{DEF}}$, and $\exists \langle x, \delta \rangle \in D$ or $\langle \delta, x \rangle \in D$ where $x \in E \cup \{\varnothing\}$. . Then, we have the following:
    ∗ if $\langle \delta, \alpha \rangle$ is unattacked, then since D is complete, $\langle \delta, \alpha \rangle \in D$, i.e., $\alpha \in E$; otherwise,
    ∗ for all $[u, \eta] \in \mathcal{F}^{\mathrm{d}}$: $[u, \eta] \rightarrow^{\mathrm{d}} \langle \delta, \alpha \rangle$, if u or η attacks α, then since α is defended by E, there exists $\langle \psi, \psi' \rangle \in D$ such that ψ or ψ' attacks u or η. In other words, $\langle \psi, \psi' \rangle$ attacks $[u, \eta]$; if u or η attacks δ, then $[u, \eta]$ attacks $\langle x, \delta \rangle$ or $\langle \delta, x \rangle$. Since D is complete, there exists $\langle \theta, \theta' \rangle \in D$ such that $\langle \theta, \theta' \rangle$ attacks $[u, \eta]$. So, $\langle \delta, \alpha \rangle$ is defended by D. Since D is complete, it holds that $\langle \delta, \alpha \rangle \in D$, and therefore $\alpha \in E$.

3. Proof of Theorem 2

Proof. For all $E \in \mathrm{co}(\mathcal{F})$, since it is obvious that $\mathrm{d}(E)$ is conflict-free, we need to verify: 1) $\mathrm{d}(E)$ defends each member of $\mathrm{d}(E)$, and 2) each defense in $\mathcal{F}^{\mathrm{DEF}}$ that is defended by $\mathrm{d}(E)$ is in $\mathrm{d}(E)$. Details:

  – For all $\langle \beta, \alpha \rangle \in \mathrm{d}(E)$, for all $[x, y] \in \mathcal{F}^{\mathrm{d}}$, if $[x, y]$ attacks $\langle \beta, \alpha \rangle$ such that $x \to \beta$ or $y \to \beta$, or $x \to \alpha$ or $y \to \alpha$, since E is a complete extension, $\exists \eta \in E$ such that $\eta \to x$ or $\eta \to y$. So, $\langle \eta, \beta \rangle$ or $\langle \eta, \alpha \rangle$ is in $\mathrm{d}(E)$, and $\langle \eta, \beta \rangle$ or $\langle \eta, \alpha \rangle$ attacks $[x, y]$. In other words, $\mathrm{d}(E)$ defends each member of $\mathrm{d}(E)$.
  – For all $\langle \alpha, \beta \rangle \in \mathcal{F}^{\mathrm{DEF}}$, if $\langle \alpha, \beta \rangle$ is defended by $\mathrm{d}(E)$, then both α and β are defended by $\mathrm{def}(\mathrm{d}(E)) = E$. Since E is a complete extension, $\alpha, \beta \in E$. So, $\langle \alpha, \beta \rangle \in \mathrm{d}(E)$.

4. Proof of Corollary 1

Proof. For $E = \mathrm{def}(D)$, under grounded semantics, we need to prove that E is minimal (w.r.t. set-inclusion). Assume the contrary. Then $\exists E' \subsetneq E$ such that E' is a grounded extension. According to Theorem 2, it holds that $\mathrm{d}(E) \in \mathrm{CO}(\mathrm{d}(\mathcal{F}))$ and $\mathrm{d}(E') \in \mathrm{CO}(\mathrm{d}(\mathcal{F}))$. Since $E' \subsetneq E$, it holds that $\mathrm{d}(E') \subsetneq \mathrm{d}(E)$. Since $\mathrm{d}(E) = \mathrm{d}(\mathrm{def}(D)) = D$, $\mathrm{d}(E') \subsetneq D$. It turns out that D is not a minimal complete extension, contradicting $D \in \mathrm{GR}(\mathrm{d}(\mathcal{F}))$.

Under preferred semantic, similarly, it is easy to verify that E is maximal (w.r.t. set-inclusion). So, for all $D \in \mathrm{PR}(\mathrm{d}(\mathcal{F}))$, $\mathrm{def}(D) \in \mathrm{pr}(\mathcal{F})$.

Under stable semantics, we need to prove that for all $\alpha \in \mathcal{A} \setminus E$: $E \to \alpha$. Assume the contrary. Then, $\exists \alpha \in \mathcal{A} \setminus E$ such that E does not attack α. There are the following possible cases:

  – α self-attacks. In this case, $(\emptyset, \alpha) \in \mathcal{F}^{\mathrm{d}} \setminus D$. Since D is a stable extension, D attacks $(\emptyset, \alpha)$. So, E attacks α. Contradiction.
  – α does not self-attack. Since α can not be initial, α is attacked by some argument $\beta \in \mathcal{A}$. It follows that $\beta \notin E$ and β does not self-attack. So, β is attacked by some argument $\gamma \in \mathcal{A}$. So, $[\gamma, \alpha] \in \mathcal{F}^{\mathrm{d}}$. Since $\alpha \notin E$, $[\gamma, \alpha] \notin D$. Since D is a stable extension, D attacks $[\gamma, \alpha]$. Since $E = \mathrm{def}(D)$ does not attack α, $\exists \eta \in E$ such that η attacks γ, and η does not attack α. Since α, β and η do not self-attack, we have the following possible cases:
    • $\{\eta, \beta\}$ is conflict-free: In this case, $\langle \eta, \beta \rangle \in \mathcal{F}^{\mathrm{DEF}}$. Since E cannot attack η, if $\langle \eta, \beta \rangle \notin D$, $\exists \psi \in E$ such that ψ attacks β. So, $[\psi, \alpha] \in \mathcal{F}^{\mathrm{d}}$. Since $[\psi, \alpha] \notin D$, $[\psi, \alpha]$ is attacked by D. Since E does not attack ψ, E attacks α. Contradiction.
    • $\{\eta, \beta\}$ is not conflict-free: If η attacks β, $[\eta, \alpha] \in \mathcal{F}^{\mathrm{d}}$. It follows that E attacks α. Contradiction. If η does not attack β, but β attacks η, $\langle \eta, \beta \rangle \in \mathcal{F}^{\mathrm{DEF}}$. This case also leads to a contradiction.