Probabilistic Graded Semantics

Matthias THIMM [a], Federico CERUTTI [b], and Tjitze RIENSTRA [a]
[a] *University of Koblenz-Landau, Germany*
[b] *Cardiff University, UK*

Abstract. We propose a new graded semantics for abstract argumentation frameworks that is based on the constellations approach to probabilistic argumentation. Given an abstract argumentation framework, our approach assigns uniform probability to all arguments and then ranks arguments according to the probability of acceptance wrt. some classical semantics. Albeit relying on a simple idea this approach (1) is based on the solid theoretical foundations of probability theory, and (2) complies with many rationality postulates proposed for graded semantics. We also investigate an application of our approach for inconsistency measurement in argumentation frameworks and show that the measure induced by the probabilistic graded semantics also complies with the basic rationality postulates from that area.

Keywords. abstract argumentation, probabilistic argumentation, graded semantics

1. Introduction

Abstract argumentation frameworks [11] provide a simple approach for a computational model of argumentation [4] by solely focusing on the interplay of arguments through a conflict relation. An abstract argumentation framework is a directed graph where vertices are identified with arguments and a directed edge between an argument a and an argument b denotes an *attack* of a on b. Given such a framework the natural reasoning question is to determine a set of arguments E (also called *extension*) that represents a coherent point of view on a possible outcome of the argumentation represented by the framework. Building on simple desirable properties of such an extension such as *conflict-freeness*, i. e., no argument in E should attack another argument in E, a variety of different semantics can be defined [5], which assigns to every framework a set of these extensions. However, the classical notion of a semantics lacks expressivity when it comes to a more fine-grained assessment of the acceptability status of arguments as they usually only differentiate between acceptance and rejection of arguments. Recent years have seen a sparking interest in *graded semantics* (or *ranking semantics*) [9,1,13,8,3], i. e., semantics that assign a ranking relation S over arguments in abstract argumentation frameworks such that aSb expresses that a is at least as acceptable as b. The technical foundation for such a graded approach to acceptability usually stems from topological considerations and assesses arguments more acceptable than others if, e. g., they are attacked by less arguments of defended by more arguments.

We contribute to the field of graded semantics by presenting a novel approach that is based on the constellation approach to probabilistic argumentation [18,14,12], viz., a probabilistic extension of abstract argumentation frameworks where arguments and attacks can be attached with probabilities that model uncertainty on the actual presence of these components. Such probabilistic argumentation frameworks induce a probability distribution over all subgraphs of the original framework and, by applying a classical semantics on each subgraph, a probability of acceptability of each argument. We use the constellation approach to define a graded semantics by assigning to each argument the same probability and using the resulting probabilities of acceptability as a means to rank the arguments. A high probability of acceptability of an argument means that it is usually accepted when it is present, despite a changing topology of the argumentation framework. This means that the acceptability of this argument is robust wrt. the topology, leading to a high level of acceptability. On the other hand, if an argument has a low probability of acceptability it means that it is often rejected, even when subgraphs are considered. This leads to a low level of acceptability. In addition to be founded on solid theoretical foundations, our approach also turns out to comply with many rationality postulates for graded semantics [8].

We also apply our novel graded semantics on the problem of measuring inconsistency (or *disagreement*) in abstract argumentation frameworks [15,16,2]. It is motivated by the field of inconsistency measurement in classical logic [20] and aims at assessing the degree of conflicts in an abstract argumentation framework in one number. This area has certain relationships with graded semantics and we show that a straightforward application of our concrete graded semantics complies with the basic rationality postulates of inconsistency measurement as well, but also provides a new perspective.

In summary, the contributions of this paper are as follows.

1. We propose a novel graded semantics based on the constellation approach to probabilistic argumentation (Section 4)
2. We apply the probabilistic graded semantics on the problem of measuring inconsistency in an abstract argumentation framework (Section 5)

Section 2 recalls background information on abstract argumentation, Section 3 introduces notation for general graded semantics, and Section 6 concludes with a summary. Proofs are omitted due to space restrictions but an extended version can be found online.[1]

2. Preliminaries

An *abstract argumentation framework* AF is a tuple $\mathsf{AF} = (A, R)$ where A is a set of arguments and R is a relation $R \subseteq A \times A$. For two arguments $a, b \in A$ the relation aRb means that argument a attacks argument b.

A *path* from b to a, noted $P(b,a)$ is a sequence $s = (c_0, \ldots, c_n)$ of arguments such as $c_0 = a$, $c_n = b$, and $\forall i < n, (c_{i+1}, c_i) \in R$. We denote by $l_P = n$ the *length* of P. In accordance with [8], for fixed $\mathsf{AF} = (A, R)$ and for all $a \in A$ define

$$R_n^-(a) = \{b \mid \exists P(b,a) \text{ with } l_P = n \in 2\mathbb{N}+1\} \qquad \text{(attackers of } a \text{ at } n)$$

$$R_-(a) = \bigcup_n R_n^-(a) \qquad \text{(attackers of } a)$$

$$\mathscr{B}_n^-(a) = \{b \in R_n^-(a) \mid R_1^-(b) = \emptyset\} \qquad \text{(attack roots of } a)$$

$$\mathscr{B}^-(a) = \bigcup_n \mathscr{B}_n^-(a) \qquad \text{(attack branches of } a)$$

$$R_n^+(a) = \{b \mid \exists P(b,a) \text{ with } l_P = n \in 2\mathbb{N}\} \qquad \text{(defenders of } a \text{ at } n)$$

[1] `http://mthimm.de/misc/comma18pgs.pdf`

$$R_+(a) = \bigcup_n R_n^+(a) \qquad \text{(defenders of } a\text{)}$$

$$\mathscr{B}_n^+(a) = \{b \in R_n^+(a) \mid R_1^-(b) = \emptyset\} \qquad \text{(defence roots of } a\text{)}$$

$$\mathscr{B}^+(a) = \bigcup_n \mathscr{B}_n^+(a) \qquad \text{(defense branches of } a\text{)}$$

$$Anc_{\mathsf{AF}}(a) = (A', \{(b,c) \in R \mid b \in A', c \in A'\})$$
$$\text{with } A' = \{a\} \cup R_+(a) \cup R_-(a) \qquad \text{(ancestor graph of } a\text{)}$$

For $a \in A$, the *defense branch added to* a is $P^+(a) = (A', R')$ where $A' = \{a_0, \ldots, a_n\}, n \in 2\mathbb{N}, a_0 = a, A \cap A' = \{a\}$ and $R' = \{(a_i, a_{i-1}) \mid i \leq n\}$. The *attack branch added to* a is $P^-(a) = (A', R')$ where $A' = \{a_0, \ldots, a_n\}, n \in 2\mathbb{N}+1, a_0 = a, A \cap A' = \{a\}$ and $R' = \{(a_i, a_{i-1}) \mid i \leq n\}$.

Two abstract argumentation frameworks $\mathsf{AF} = (A, R)$ and $\mathsf{AF}' = (A', R')$ are *isomorphic*, written $\mathsf{AF} \equiv \mathsf{AF}'$, if there is a bijective function $\gamma : A \to A'$ such that aRb iff $\gamma(a)R'\gamma(b)$ for all $a, b \in A$ (γ is then called an *isomorphism*).

A *connected component* of an $\mathsf{AF} = (A, R)$ is a maximal subgraph $\mathsf{AF}' = (A', R')$ such that every two arguments $a, b \in A'$ are connected through a path while *ignoring edge directions*. Let $cc(\mathsf{AF})$ be the set of all connected components of AF.

An extension E is a set $E \subseteq A$ that contains a set of arguments that are *mutually acceptable*. We say that

- E is *conflict-free* iff for no $a, b \in E$, aRb;
- E *defends* $a \in A$ iff for all $c \in A$ with cRa there is $b \in E$ with bRc;
- E is *admissible* iff E is conflict-free and for all $a \in E$, E defends a;
- E is *complete* (co) iff E is admissible and for all $b \in A$ s. t. E defends b, $b \in E$;
- E is *grounded* (gr) iff E is complete and E is minimal (wrt. set inclusion);
- E is *preferred* (pr) if E is complete and E is maximal (wrt. set inclusion).

Note that the grounded extension is uniquely determined and every argumentation framework possess at least one complete, grounded, and preferred extension [11]. Note that we do not consider stable semantics as existence of stable extensions is not universally guaranteed [11] and this would lead to some case differentiations in our definitions.

Let $\sigma \in \{\mathsf{co}, \mathsf{gr}, \mathsf{pr}\}$ be any semantics. An argument $a \in A$ is *credulously accepted* in AF wrt. σ, written $\mathsf{AF} \mid\!\sim^c_\sigma a$, iff $a \in E$ for some σ-extension E. An argument $a \in A$ is *skeptically accepted* in AF wrt. σ, written $\mathsf{AF} \mid\!\sim^s_\sigma a$, iff $a \in E$ for all σ-extensions E. We use $\circ \in \{s, c\}$ as a symbol to refer to any inference mode.

3. Graded Semantics

In contrast to the classical semantics introduced above, *graded semantics* [9,1,13,8,3] take a more fine-grained perspective on the acceptability of an argument by assigning numerical values.

Definition 1. A *graded semantics* G assigns to each argumentation framework $\mathsf{AF} = (A, R)$ a function $G_{\mathsf{AF}} : A \to \mathbb{R}$.

The intuition behind the value $G_{\mathsf{AF}}(a)$ of an argument a is that larger values of $G_{\mathsf{AF}}(a)$ indicate larger acceptability of a. For every classical semantics $\sigma \in \{\mathsf{co}, \mathsf{gr}, \mathsf{pr}\}$

and inference type $\circ \in \{s,c\}$ we can define a simple graded semantics G^{σ} via

$$G^{\sigma,\circ}_{\mathsf{AF}}(a) = \begin{cases} 1 \text{ iff } \mathsf{AF} \mid\!\sim^{\circ}_{\sigma} a \\ 0 \text{ otherwise} \end{cases}$$

However, the idea behind graded semantics is to obtain a more fine-grained perspective on the acceptability of arguments. An example of a more elaborate graded semantics is the *categoriser function* [7,19] $G^{\text{Cat}}_{\mathsf{AF}}$ defined by the (unique) solution to the set of equations

$$G^{\text{Cat}}_{\mathsf{AF}}(a) = \frac{1}{1+\sum_{b\in R^{-}_{1}(a)} G^{\text{Cat}}_{\mathsf{AF}}(b)}$$

for all $a \in A$ with the usual convention that the value of the empty sum is zero. In other words, $G^{\text{Cat}}_{\mathsf{AF}}$ assigns to an argument 1 if it is not attacked. Otherwise, the more and stronger arguments are attacking that argument, the lower the score. We refer to [8,10] for a more thorough discussion of various approaches.

A different perspective on graded semantics can be obtained by focusing solely on the relationships between arguments wrt. acceptability and not on the actual numbers.

Definition 2. A *ranking semantics* S assigns to each argumentation framework $\mathsf{AF} = (A,R)$ a relation $S_{\mathsf{AF}} \subseteq A \times A$.

Similarly as for graded semantics, $aS_{\mathsf{AF}}b$ is interpreted as a is at least as acceptable as b. Every graded semantics can be trivially translated into a ranking semantics S^G via $aS^G_{\mathsf{AF}}b$ iff $G_{\mathsf{AF}}(a) \geq G_{\mathsf{AF}}(b)$ but there are instances of ranking semantics that cannot be formalised as graded semantics, in particular in cases where arguments are not comparable, see e. g. [13]. However, when there is no risk of confusion we will use the terms *ranking semantics* and *graded semantics* interchangeably.

A ranking S_{AF} can be extended to sets of arguments $X,X' \subseteq A$ as follows [1]. We write $XS_{\mathsf{AF}}X'$ iff there is an injective function $f : X' \to X$ such that for all $a \in X'$, $f(a)S_{\mathsf{AF}}a$. We write $XS^{>}_{\mathsf{AF}}X'$ iff $XS_{\mathsf{AF}}X'$ and ($|X'| < |X|$ or there is $a \in X'$ with $f(a)S_{\mathsf{AF}}a$ and not $aS_{\mathsf{AF}}f(a)$).

4. Probabilistic Argumentation and Graded Semantics

We now propose a graded semantics that is based on probabilistic abstract argumentation [18,14,17]. We briefly recall probabilistic argumentation frameworks in Section 4.1 and continue with our approach and its analysis in Section 4.2.

4.1. Probabilistic Argumentation Frameworks

In order to define our graded semantics, we will consider a simplified version of *probabilistic argumentation frameworks* due to [18]. In our simplification, we only consider probabilities of arguments (in contrast to the general case of [18]) where probabilities of attacks are allowed as well).

Definition 3. A *probabilistic argumentation framework* PAF is a triple $\mathsf{PAF} = (A,R,P)$ where (A,R) is an abstract argumentation framework and P is a function $P : A \to [0,1]$.

For every argument $a \in A$ of a probabilistic argumentation framework PAF the value $P(a)$ is the probability that a is actually present in the argumentation framework. By

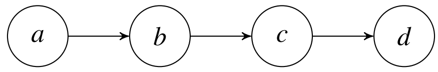

Figure 1. The argumentation framework from Example 1

X	$\emptyset$	a	b	a,b	c	a,c	b,c	a,b,c	d	a,d	b,d	a,b,d	c,d	a,c,d	b,c,d	a,b,c,d
Accepted	$\emptyset$	a	b	a	c	a,c	b	a,c	d	a,d	b,d	a,d	c	a,c	b,d	a,c

Table 1. Choices of X and corresponding accepted arguments in Example 1

assuming probabilistic independence between the presence of different arguments, we obtain a probability distribution over sets of arguments. By abuse of notation we denote this probability distribution P as well, which is defined as

$$P(X) = \prod_{a \in X} P(a) \prod_{a \notin X} (1 - P(a))$$

for all $X \subseteq A$. It can be easily shown that $\sum_{X \subseteq A} P(X) = 1$, so P is indeed a probability distribution. Given a set $X \subseteq A$ of arguments, we denote by AF_X the induced subgraph of X, i. e. $\mathsf{AF}_X = (X, R \cap (X \times X))$.

Let now $\sigma \in \{\mathsf{co}, \mathsf{gr}, \mathsf{pr}\}$ be a semantics and $\circ \in \{s, c\}$ be an inference mode. The *probability of acceptance of* a, denoted by $P^{\mathsf{PAF}}_{\circ,\sigma}(a)$, is then defined via

$$P^{\mathsf{PAF}}_{\circ,\sigma}(a) = \sum_{a \in X \subseteq A, \mathsf{AF}_X \mathrel{|\!\sim}^{\circ}_{\sigma} a} P(X)$$

In other words, $P^{\mathsf{PAF}}_{\circ,\sigma}(a)$ is the sum of the probabilities of the subgraphs of (A, R) where a is accepted wrt. σ and $\circ$.

Example 1. Let $\mathsf{AF} = (A, R)$ be the AF shown in Figure 1 and consider credulous reasoning wrt. grounded semantics. Let $\mathsf{PAF} = (A, R, P)$ be a probabilistic argumentation framework with $P(x) = 0.5$ for all $x \in A$. Table 1 lists each subset of $X \subseteq A$, together with the set of arguments $x \in A$ such that $\mathsf{AF}_X \mathrel{|\!\sim}^{c}_{\mathsf{gr}} x$. For each $X \subseteq A$ we have $P(X) = 0.5^4 = 0.0625$. Thus, for each $x \in A$ we can calculate the probability $P^{\mathsf{PAF}}_{c,\mathsf{gr}}(x)$ by multiplying the number of subsets of A that make x accepted by 0.0625. This yields

$$P^{\mathsf{PAF}}_{c,\mathsf{gr}}(a) = 0.5 \qquad P^{\mathsf{PAF}}_{c,\mathsf{gr}}(b) = 0.25 \qquad P^{\mathsf{PAF}}_{c,\mathsf{gr}}(c) = 0.375 \qquad P^{\mathsf{PAF}}_{c,\mathsf{gr}}(d) = 0.3125$$

4.2. Probabilistic Graded Semantics

We now turn to the main contribution of this paper, namely a graded semantics based on probabilistic argumentation frameworks. The main idea is to assign to all arguments of an argumentation framework uniform probability and interpret the obtained justification probabilities as scores in a graded semantics approach.

Definition 4. Let $\mathsf{AF} = (A, R)$ be an argumentation framework and $p \in [0, 1]$. We denote by $\mathsf{AF}[p]$ the probabilistic argumentation framework $\mathsf{PAF} = (A, R, P)$ with $P(a) = p$ for all $a \in A$.

Definition 5. Let $\mathsf{AF} = (A, R)$ be an argumentation framework, $p \in [0, 1]$, $\sigma \in \{\mathsf{co}, \mathsf{gr}, \mathsf{pr}\}$, and $\circ \in \{s, c\}$. The *probabilistic graded semantics* $G^{\sigma,\circ,p}$ is defined through

$$G^{\sigma,\circ,p}_{\mathsf{AF}}(a) = P^{\mathsf{AF}[p]}_{\circ,\sigma}(a)$$

for every $a \in A$.

Our idea is simple, yet effective (as we will see below) and based on probabilistic foundations. We assess the strength (or level of acceptability) of an argument on the effect this argument has globally in varying scenarios. In our setting, a high score of an argument means that it is usually accepted, given that it is present in some subgraph. That means that even under a changing topology, i. e., when other arguments are ignored, the argument under consideration is usually accepted and thus, more or less, independent of other arguments. On the other hand, if an argument has a low score then it is usually rejected in subgraphs, i. e., even in subgraphs where some of its attackers are not even present. Using a uniform probability for all arguments in the graph is a natural way of assessing the impact of each argument in changing scenarios. It shows that some arguments, even when starting with the same initial probability, behave differently and thus should be assigned different levels of acceptability.

Example 2. Consider again the argumentation framework AF in Figure 1. Using the results obtained in Example 1 it is straightforward to determine that

$$G^{\mathsf{gr},c,0.5}_{\mathsf{AF}}(a) = 0.5 \quad G^{\mathsf{gr},c,0.5}_{\mathsf{AF}}(b) = 0.25 \quad G^{\mathsf{gr},c,0.5}_{\mathsf{AF}}(c) = 0.375 \quad G^{\mathsf{gr},c,0.5}_{\mathsf{AF}}(d) = 0.3125.$$

We can see that our graded semantics complies with some basic intuitions on graded acceptability. First, argument a has the largest score, which is intuitive as it is the only argument that is not attacked. Argument c has the second highest score, which is clear as though it is attacked by b it is also defended by a. Argument d is attacked and defended, but ultimately defeated by a. Finally, b is attacked but undefended, thus resulting in the lowest score.

We now establish some basic properties of our approach.

Theorem 1. *Let* $\mathsf{AF} = (A,R)$ *be an argumentation framework,* $p \in [0,1]$*,* $\sigma \in \{\mathit{co},\mathit{gr},\mathit{pr}\}$*, and* $\circ \in \{s,c\}$*.*

1. $0 \leq G^{\sigma,\circ,p}_{\mathsf{AF}}(a) \leq p$*, for all* $a \in A$
2. $G^{\sigma,\circ,p}_{\mathsf{AF}}(a) = p$ *iff* $R^{-}_{1}(a) = \emptyset$*, for all* $a \in A$
3. $G^{\sigma,\circ,p}_{\mathsf{AF}}(a) = 0$ *iff* $p = 0$ *or* aRa*, for all* $a \in A$
4. *If* $p = 0$ *then* $G^{\sigma,\circ,p}_{\mathsf{AF}}(a) = 0$*, for all* $a \in A$
5. $G^{\sigma,\circ,1} = G^{\sigma,\circ}$

Property 1 above establishes general bounds for the scores in our approach and thus can be used for normalisation purposes. Property 2 states that unattacked arguments always have the highest score, while property 3 states that self-attacking arguments always have the lowest score of zero. Property 4 shows that the semantics trivialises to a uniform score for all arguments in the case $p = 0$ and property 5 shows that the semantics trivialises to classical semantics in the case $p = 1$. Due to properties 4 and 5 we will assume $p \in (0,1)$ for the rest of the paper.

A question one might ask is about the implications of choosing a particular uniform probability p over another one. In particular, one might ask the question whether different choices of p lead to the same qualitative ranking in the end.

Example 3. Consider the argumentation framework AF in Figure 2. Using credulous reasoning with grounded semantics and $p_1 = 0.1$ and $p_2 = 0.9$ we get

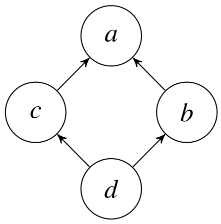

Figure 2. The argumentation framework from Example 3

$$G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(a) = 0.0829 \qquad G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(b) = G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(c) = 0.09 \qquad G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(d) = 0.1$$

$$G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(a) = 0.8109 \qquad G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(b) = G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(c) = 0.09 \qquad G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(d) = 0.9$$

As one can see, $G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(a) < G^{\mathsf{gr},c,p_1}_{\mathsf{AF}}(b)$ but $G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(a) > G^{\mathsf{gr},c,p_2}_{\mathsf{AF}}(b)$, so a is less acceptable than b wrt. $p_1 = 0.1$ but more acceptable wrt. $p_2 = 0.9$.

As the previous example illustrated, the parameter p (roughly) controls whether attackers or defenders are of more importance to assess the score of an argument. For large p the semantics behaves more like classical semantics and makes an argument more acceptable if it has defenders. For low p the existence of many attackers can lower the score, even if the argument is eventually accepted in classical semantics. We leave a thorough analysis of the role of p for future work.

We will now analyse the properties of our approach and thus demonstrate its suitability as a graded semantics. We follow the practice established by Amgoud and Ben-Naim [1] of evaluating graded semantics by checking the compliance to a series of *rationality postulates*, see also [10] for an overview. We state these properties for ranking semantics, so let S be an arbitrary ranking semantics and S_{AF} its induced relation on an argumentation framework AF.

Abstraction For every pair $\mathsf{AF} = (A,R), \mathsf{AF}' = (A',R')$ of isomorphic frameworks and every isomorphism $\gamma : A \to A'$, for all $a,b \in A$, $aS_{\mathsf{AF}}b$ iff $\gamma(a)S_{\mathsf{AF}}\gamma(b)$. (*The ranking on arguments should be defined only on the basis of the attacks between them.*)

Independence For every $\mathsf{AF} = (A,R)$ and $\mathsf{AF}' = (A',R') \in cc(\mathsf{AF})$, for all $a,b \in A'$, $aS_{\mathsf{AF}'}b$ iff $aS_{\mathsf{AF}}b$. (*The ranking between two arguments a and b should be independent of any argument that is neither connected to a nor to b.*)

Void precedence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $R^-_1(a) = \emptyset$ and $R^-_1(b) \neq \emptyset$ then $aS_{\mathsf{AF}}b$ and not $bS_{\mathsf{AF}}a$. (*A non-attacked argument should be ranked strictly higher than any attacked argument.*)

Self-contradiction For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if not aRa but bRb then $aS_{\mathsf{AF}}b$ and not $bS_{\mathsf{AF}}a$. (*A self-attacking argument should be ranked lower than any non self-attacking argument.*)

Cardinality precedence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $|R^-_1(a)| < |R^-_1(b)|$ then $aS_{\mathsf{AF}}b$ and not $bS_{\mathsf{AF}}a$. (*The greater the number of direct attackers for an argument, the weaker the level of acceptability of this argument.*)

Quality precedence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if there is $c \in R^-_1(b)$ such that for all $d \in R^-_1(a)$, $cS_{\mathsf{AF}}d$ but not $dS_{\mathsf{AF}}c$, then $aS_{\mathsf{AF}}b$ and not $bS_{\mathsf{AF}}a$. (*The greater the acceptability of one direct attacker for an argument, the weaker the level of acceptability of this argument.*)

Counter-Transitivity For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $R_1^-(b) S_{\mathsf{AF}} R_1^-(a)$ then $a S_{\mathsf{AF}} b$. (*If the direct attackers of b are at least as numerous and acceptable as those of a, then a is at least as acceptable as b.*)

Strict Counter-Transitivity For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $R_1^-(b) S_{\mathsf{AF}}^{>} R_1^-(a)$ then $a S_{\mathsf{AF}} b$ and not $b S_{\mathsf{AF}} a$. (*The direct attackers of b are strictly more numerous or acceptable than those of a, then a is strictly more acceptable than b.*)

Defense precedence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $|R_1^-(a)| = |R_1^-(b)|$ and $R_2^+(a) \neq \emptyset$ but $R_2^+(b) = \emptyset$ then $a S_{\mathsf{AF}} b$ and not $b S_{\mathsf{AF}} a$. (*For two arguments with the same number of direct attackers, a defended argument is ranked higher than a non-defended argument.*)

Distributed Defense precedence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if $|R_1^-(a)| = |R_1^-(b)|$ and $|R_2^+(a)| = |R_2^+(b)|$, if the defence of a is simple—every direct defender of a directly attacks exactly one direct attacker of a—and distributed—every direct attacker of a is attacked by at most one argument—and the defence of b is simple but not distributed, then $a S_{\mathsf{AF}} b$ and not $b S_{\mathsf{AF}} a$. (*The best defense is when each defender attacks a distinct attacker (*distributed defense*).*)

Addition of an Attack Branch For every $\mathsf{AF} = (A,R)$, for all $a \in A$, for every isomorphism γ such that $\mathsf{AF} = \gamma(\mathsf{AF})$, if $\mathsf{AF}^* = \mathsf{AF} \cup \gamma(\mathsf{AF}) \cup P^-(\gamma(a))$, then $\gamma(a) S_{\mathsf{AF}^*} a$ and not $a S_{\mathsf{AF}^*} \gamma(a)$. (*Adding a new attack line to any argument degrades its ranking.*)

Strict addition of a Defense Branch For every $\mathsf{AF} = (A,R)$, for all $a \in A$, for every isomorphism γ such that $\mathsf{AF} = \gamma(\mathsf{AF})$, if $\mathsf{AF}^* = \mathsf{AF} \cup \gamma(\mathsf{AF}) \cup P^+(\gamma(a))$, then $\gamma(a) S_{\mathsf{AF}^*} a$ and not $a S_{\mathsf{AF}^*} \gamma(a)$. (*Adding a defence branch to any argument improves its ranking.*)

Addition of a Defense Branch For every $\mathsf{AF} = (A,R)$, for all $a \in A$, for every isomorphism γ such that $(A',R') = \gamma(\mathsf{AF})$ and $A' \cap A = \emptyset$, if $\mathsf{AF}^* = \mathsf{AF} \cup \gamma(\mathsf{AF}) \cup P^+(\gamma(a))$ and $R_1^-(a) \neq \emptyset$, then $\gamma(a) S_{\mathsf{AF}^*} a$ and not $a S_{\mathsf{AF}^*} \gamma(a)$. (*Adding a defence branch to an attacked argument should improve its ranking.*)

Increase of an Attack branch For every $\mathsf{AF} = (A,R)$, for all $a \in A$, for every isomorphism γ such that $(A',R') = \gamma(\mathsf{AF})$ and $A' \cap A = \emptyset$, if $\exists b \in \mathscr{B}_-(a)$, $b \notin \mathscr{B}_+(a)$ and $\mathsf{AF}^* = \mathsf{AF} \cup \gamma(\mathsf{AF}) \cup P^+(\gamma(b))$, then $\gamma(a) S_{\mathsf{AF}} a$ and not $a S_{\mathsf{AF}} \gamma(a)$. (*Increasing the length of an attack branch of an argument improves its ranking.*)

Increase of a Defense branch For every $\mathsf{AF} = (A,R)$, for all $a \in A$, for every isomorphism γ such that $(A',R') = \gamma(\mathsf{AF})$ and $A' \cap A = \emptyset$, if $\exists b \in \mathscr{B}_+(a)$, $b \notin \mathscr{B}_-(a)$ and $\mathsf{AF}^* = \mathsf{AF} \cup \gamma(\mathsf{AF}) \cup P^+(\gamma(b))$, then $a S_{\mathsf{AF}} \gamma(a)$ and not $\gamma(a) S_{\mathsf{AF}} a$. (*Increasing the length of a defense branch of an argument degrades its ranking.*)

Total For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, $a S_{\mathsf{AF}} b$ or $b S_{\mathsf{AF}} a$. (*All pairs of arguments can be compared.*)

Non-attacked Equivalence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, $R_1^-(a) = \emptyset$ and $R_1^-(b) = \emptyset$ then $a S_{\mathsf{AF}} b$ and $b S_{\mathsf{AF}} a$. (*All the non-attacked arguments have the same rank.*)

Argument Equivalence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, for every isomorphism γ such that $Anc_{\mathsf{AF}}(a) = \gamma(Anc_{\mathsf{AF}}(b))$, then $a S_{\mathsf{AF}} b$ and $b S_{\mathsf{AF}} a$. (*If two arguments have the same ancestors' graph, then they should be equally acceptable.*)

Ordinal Equivalence For every $\mathsf{AF} = (A,R)$, for all $a,b \in A$, if there exists a bijective function f from $R_1^-(a)$ to $R_1^-(b)$ such that $\forall c \in R_1^-(a)$, $c S_{AF} f(c)$ and $f(c) S_{\mathsf{AF}} c$, then $a S_{\mathsf{AF}} b$ and $b S_{\mathsf{AF}} a$. (*If two arguments a and b have the same number of direct attackers and, for each direct attacker of a there is an equally acceptable direct attacker of b, then a and b should also be equally acceptable.*)

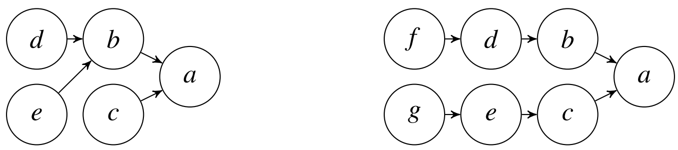

Figure 3. The argumentation frameworks AF_1 (left) and AF_2 (right) from Example 4

Attack vs Full Defense For every acyclic $\mathsf{AF} = (A, R)$, for all $a, b \in A$, $|\mathscr{B}_-(a)| = 0$, $|R_1^-(b)| = 1$, and $|R_2^+(b)| = 0$ then $aS_{\mathsf{AF}}b$ and not $bS_{\mathsf{AF}}a$. (*An argument without any attack branch should be ranked higher than an argument only attacked by one non-attacked argument.*)

Note that not all of the postulates are independent, some follow from others and some are incompatible, see [10] for a thorough analysis.

As the following result shows, our approach satisfies many of the above postulates.

Theorem 2. *For $p \in (0,1)$, $\sigma \in \{\mathit{co}, \mathit{gr}, \mathit{pr}\}$, and $\circ \in \{s, c\}$, $G^{\sigma,\circ,p}$ satisfies* Abstraction*,* Independence*,* Void precedence*,* Self-contradiction*,* Defense precedence*,* Increase of an Attack Branch*,* Addition of an Attack Branch*,* Total*,* Argument Equivalence*, and* Non-attacked Equivalence*.*

Counter-examples for the other postulates can be constructed but are omitted due to space restrictions. We will, however, look a bit closer at the case of *Distributed Defense Precedence*.

Example 4. Consider the two argumentation frameworks AF_1 and AF_2 in Figure 3 and

$$G^{\mathsf{gr},c,0.5}_{\mathsf{AF}_1}(a) \approx 0.219 \qquad\qquad G^{\mathsf{gr},c,0.5}_{\mathsf{AF}_2}(a) \approx 0.195$$

showing that *Distributed Defense Precedence* is violated. However, we argue that a in AF_2 should be indeed ranked lower than a in AF_1. In AF_1 argument a is defended against b and all its defenders are unattacked. However, in AF_2 all defenders are attacked themselves. *Distributed Defense Precedence* would require that a in AF_2 should be ranked higher than a in AF_1. The problem with that requirements here is that only direct attackers and direct defenders are considered, but whether the defenders are acceptable or not is unimportant. The case could be made more extreme by adding an arbitrary number of attackers on d and/or e in AF_2. *Distributed Defense Precedence* would still require a in AF_2 to be more acceptable than a in AF_1 but our semantics will further degrade the score of a in AF_2.

It is worthwhile noting that no other graded semantics investigated in [8] satisfies the exact same set of rationality postulates nor a superset of it.

5. Application: Inconsistency Measurement

In [15,16] and independently in [2] inconsistency measures on abstract argumentation frameworks are investigated. These are functions I that map any abstract argumentation framework AF to a real value $I(\mathsf{AF})$ with the intuition that larger values indicate larger inconsistency (or rather *disagreement*) in AF. The idea is that a framework without any attacks should be assessed consistent, as all arguments are mutually compatible. Having more attacks indicates less compatibility and thus an increased value of inconsistency/disagreement. While the number of attacks is thus already a simple implementation of this idea [16], more advanced techniques can be used to measure further subtleties.

We can use our probabilistic graded semantics to define a new measure of inconsistency/disagreement as follows.

Definition 6. Let $p \in (0,1)$, $\sigma \in \{\mathsf{co},\mathsf{gr},\mathsf{pr}\}$, and $\circ \in \{s,c\}$. Define $I^{p,\sigma,\circ}$ through

$$I^{p,\sigma,\circ}(\mathsf{AF}) = p|A| - \sum_{a\in A} G^{\sigma,\circ,p}_{\mathsf{AF}}(a)$$

for $\mathsf{AF} = (A,R)$.

Our measure $I^{p,\sigma,\circ}$ sums up all probabilities of each argument in the framework. The idea here is that large probabilities for arguments indicate little conflict between the arguments. In order to obtain a measure of inconsistency/disagreement we subtract this number from its maximal possible value. Indeed, the following is a direct implication from Theorem 1 item 1.

Theorem 3. *Let* $p \in (0,1)$, $\sigma \in \{\mathit{co},\mathit{gr},\mathit{pr}\}$, *and* $\circ \in \{s,c\}$. *For every* $\mathsf{AF} = (A,R)$, $0 \leq I^{p,\sigma,\circ}(\mathsf{AF}) \leq p|A|$.

Furthermore, our definition complies with Hunter's basic constraints for such measures [16].

Theorem 4. *Let* $p \in (0,1)$, $\sigma \in \{\mathit{co},\mathit{gr},\mathit{pr}\}$, *and* $\circ \in \{s,c\}$. *The measure* $I^{p,\sigma,\circ}$ *satisfies*

Consistency $I^{p,\sigma,\circ}((A,R)) = 0$ *iff* $R = \emptyset$.
Freeness $I^{p,\sigma,\circ}((A,R)) = I^{p,\sigma,\circ}((A\cup\{a\},R))$ *for* $a \notin A$.

The basic intuition behind the property *Consistency* is that an argumentation framework without any attacks has no conflict and should receive measure zero. The basic intuition behind the property *Freeness* is that removing an argument that is not involved in any conflict does not change the measure.

Hunter considers a series of further postulates in [15,16], two of which are satisfied by our approach as well.

Theorem 5. *Let* $p \in (0,1)$, $\sigma \in \{\mathit{co},\mathit{gr},\mathit{pr}\}$, *and* $\circ \in \{s,c\}$. *The measure* $I^{p,\sigma,\circ}$ *satisfies*

Isomorphic Invariance $I^{p,\sigma,\circ}(\mathsf{AF}) = I^{p,\sigma,\circ}(\mathsf{AF}')$ *if* AF *and* AF' *are isomorphic.*
Disjoint Additivity $I^{p,\sigma,\circ}((A\cup A',R\cup R')) = I^{p,\sigma,\circ}((A,R)) + I^{p,\sigma,\circ}((A',R'))$ *if* $A\cap A' = \emptyset$, $R \subseteq A\times A$, *and* $R' \subseteq A'\times A'$.

However, our approach does not satisfy the property *monotonicity* and its generalisation *super-additivity*, which are defined as

Monotonicity $I^{p,\sigma,\circ}((A,R)) \leq I^{p,\sigma,\circ}((A',R'))$ if $A \subseteq A'$ and $R \subseteq R'$.
Super-Additivity $I^{p,\sigma,\circ}((A\cup A',R\cup R')) \geq I^{p,\sigma,\circ}((A,R)) + I^{p,\sigma,\circ}((A',R'))$ for $R \subseteq A\times A$, and $R' \subseteq A'\times A'$.

But as abstract argumentation is a non-monotonic formalism we also argue that a measure of inconsistency/disagreement should not necessarily behave monotonic as well, see a recent discussion in [21].

Figure 4. The argumentation frameworks AF_1 (left) and AF_2 (right) from Example 5

Example 5. Consider the two argumentation frameworks AF_1 and AF_2 in Figure 4 where

$$I^{0.5,\mathsf{gr},c}(\mathsf{AF}_1) = 2.5 - (0.0 + 0.25 + 0.25 + 0.25 + 0.5) = 1.25$$
$$I^{0.5,\mathsf{gr},c}(\mathsf{AF}_2) = 2.5 - (0.0 + 0.375 + 0.375 + 0.375 + 0.5) = 0.875$$

therefore violating *Monotonicity*. However, we argue that there is actually less disagreement in AF_2 than in AF_1. The self-attack of a in AF_1 is unresolved and shows some "issue" with the framework. However, in AF_2 this issue is resolved by an outside attack and thus AF_2 should be regarded as less inconsistent as AF_1 (in fact, a stable extension exists for AF_2 but not AF_1).

Another property from [16] which is not satisfied by our approach is *Inversion*. For $\mathsf{AF} = (A,R)$ define $Inv(\mathsf{AF}) = (A, \{(a,b) \mid bRa\})$, i. e., $Inv(\mathsf{AF})$ is obtained from AF by inverting all attacks. *Inversion* states that both frameworks should possess the same measure. However, here we can argue similarly as for the case of monotonicity and do not think that this is a reasonable demand. In fact, *Inversion* is also violated by many measures investigated in [16].

In [2], Amgoud and Ben-Naim independently make a similar study and also propose a series of rationality postulates. In fact, some of their postulates are the same as in [15,16]. In particular, *anonymity* in [2] is the same as *isomorphic invariance*, *agreement* is the same as *consistency*, *dummy* is the same as *freeness*, and *monotony* is (essentially) the same as *monotonicity*. Amgoud and Ben-Naim propose three further properties: *reinforcement*, *cycle precedence*, and *size sensitivity*. However, we leave a discussion of these properties and a deeper analysis of $I^{p,\sigma,\circ}$ for future work.

6. Summary

We proposed a novel graded semantics based on the constellations approach to probabilistic argumentation. We showed that this semantics complies with many rationality postulates for graded semantics from the literature and provides an intuitive ranking based on probability theory. An implementation of the semantics is available in Tweety.[2] We applied the semantics to the field of inconsistency measurement in abstract argumentation frameworks and showed that the approach is compatible with this setting as well.

For future work, we plan to analyse our approach further and, in particular, characterise the exact role of the parameter p and its influences. Note that p has no influence on the compliance with the rationality postulates. This suggests that further general properties are needed to describe the role of this parameter.

[2] `http://mthimm.de/r/?r=tweety-pgs`

Acknowledgements The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1686/3-1).

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In *Proceedings of SUM'13*, 2013.

[2] L. Amgoud and J. Ben-Naim. Measuring disagreement in argumentation graphs. In *Proceedings of SUM'17*, 2017.

[3] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In *Proceedings of IJCAI'17*, 2017.

[4] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G.R. Simari, M. Thimm, and S. Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017.

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

[6] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171:675–700, 2007.

[7] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2):203–235, 2001.

[8] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of AAAI'16*, 2016.

[9] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[10] J. Delobelle. *Ranking-based Semantics for Abstract Argumentation*. PhD thesis, Centre de Recherche en Informatique de Lens, 2017.

[11] P.M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–358, 1995.

[12] B. Fazzinga, S. Flesca, and F. Parisi. On the complexity of probabilistic abstract argumentation frameworks. *ACM Transactions on Computational Logic*, 16(3):22:1–22:39, 2015.

[13] D. Grossi and S. Modgil. On the graded acceptability of arguments. In *Proceedings of IJCAI'15*, 2015.

[14] A. Hunter. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55:607–638, 2014.

[15] A. Hunter. Measuring inconsistency in argument graphs. Technical report, arXiv:1708.02851, 2017.

[16] A. Hunter. Measuring inconsistency in argument graphs. In *Measuring Inconsistency in Information*. College Publications, 2018.

[17] A. Hunter and M. Thimm. Optimization of dialectical outcomes in dialogical argumentation. *International Journal of Approximate Reasoning*, 78:73–102, July 2016.

[18] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings of TAFA'11*, 2011.

[19] F. Pu, J. Luo, Y. Zhang, and G. Luo. Argument ranking with categoriser function. In *Proceedings of KSEM'14*, 2014.

[20] M. Thimm. On the evaluation of inconsistency measures. In *Measuring Inconsistency in Information*, College Publications, 2018.

[21] M. Ulbricht, M. Thimm, and G. Brewka. Measuring strong inconsistency. In *Proceedings of AAAI'18*, 2018.