Explainable Acceptance in Probabilistic Abstract Argumentation: Complexity and Approximation

Gianvincenzo Alfano[1], **Marco Calautti**[1,2], **Sergio Greco**[1], **Francesco Parisi**[1], **Irina Trubitsyna**[1]
[1]DIMES Department, University of Calabria, Italy
[2]DISI Department, University of Trento, Italy
g.alfano@dimes.unical.it, marco.calautti@unitn.it, {greco, fparisi, i.trubitsyna}@dimes.unical.it

Abstract

Recently there has been an increasing interest in probabilistic abstract argumentation, an extension of Dung's abstract argumentation framework with probability theory. In this setting, we address the problem of computing the probability that a given argument is accepted. This is carried out by introducing the concept of probabilistic explanation for a given (probabilistic) extension. We show that the complexity of the problem is $FP^{\#P}$-hard and propose polynomial approximation algorithms with bounded additive error for probabilistic argumentation frameworks where odd-length cycles are forbidden. This is quite surprising since, as we show, such kind of approximation algorithm does not exist for the related $FP^{\#P}$-hard problem of computing the probability of the credulous acceptance of an argument, even for the special class of argumentation frameworks considered in the paper.

1 Introduction

Formal argumentation has emerged as one of the important fields in Artificial Intelligence (Bench-Capon and Dunne 2007; Simari and Rahwan 2009; Atkinson et al. 2017). In particular, an abstract Argumentation Framework (AF) is a simple, yet powerful formalism for modelling disputes between two or more agents (Dung 1995). An AF consists of a set of *arguments* and a binary *attack* relation over the set of arguments that specifies the interactions between arguments: intuitively, if argument a attacks argument b, then b is acceptable only if a is not. Hence, arguments are abstract entities whose role is entirely determined by the interactions specified by the attack relation.

Recently, there has been an increasing interest in modeling uncertainty in argumentation. This has been carried out by combining probability theory with formal argumentation. One of the most popular approaches based on probability theory for modeling the uncertainty is the so called *constellations* approach (Dung and Thang 2010; Rienstra 2012; Doder and Woltran 2014; Hunter 2012; Li, Oren, and Norman 2011), where alternative scenarios, called *possible worlds*, are associated with probabilities. In particular, in a *Probabilistic Argumentation Framework* (PrAF) (Li, Oren, and Norman 2011; Fazzinga, Flesca, and Parisi 2015; Fazzinga, Flesca, and Parisi 2016; Fazzinga, Flesca, and Furfaro 2019) a probability distribution function (PDF) on the set of possible worlds is entailed by the probabilities that are associated with arguments and attacks.

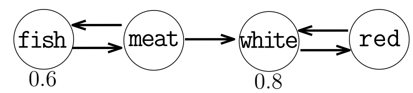

Figure 1: Probabilistic argumentation framework Δ of Example 1.

Example 1. Consider the PrAF $\Delta = \langle\{\texttt{fish}, \texttt{meat}, \texttt{white}, \texttt{red}\}, \{(\texttt{fish}, \texttt{meat}), (\texttt{meat}, \texttt{fish}), (\texttt{meat}, \texttt{white}), (\texttt{white}, \texttt{red}), (\texttt{red}, \texttt{white})\}, \{\texttt{fish}/0.6,\ \texttt{white}/0.8\}\rangle$ whose corresponding graph is shown in Figure 1, where nodes and edges represent arguments and attacks, respectively, and probabilities different from 1 are specified nearby them. For the sake of brevity, we do not specify the probabilities of certain elements in Δ (all the other elements different from `fish` and `white` have probability 1). Intuitively, Δ describes what a person is going to have for lunch as follows. (S)he will have either `fish` or `meat`, and will drink either `white` wine or `red` wine. However, if (s)he will have `meat`, then (s)he will not drink `white` wine. Furthermore, the probability that `fish` is available is 0.6, whereas that that `white` wine is available is 0.8. □

In this paper we do not address the problem of assigning probabilities to arguments or attacks, as instead done e.g. in (Hunter 2012; Hunter 2013), and assume they are given.

Several argumentation semantics—e.g. *grounded* ($\mathcal{GR}$), *preferred* ($\mathcal{PR}$), *stable* ($\mathcal{ST}$), and *semi-stable* ($\mathcal{SST}$)—have been defined for AFs, leading to the characterization of $\mathcal{S}$-*extensions*, which intuitively consist of the sets of arguments that can be collectively accepted under semantics $\mathcal{S}$. Consider for instance the deterministic version of the PrAF in Example 1, obtained by assuming that all arguments are certain (i.e., they have probability 1). Considering the stable semantics, the $\mathcal{ST}$-extensions are $E_1 = \{\texttt{fish}, \texttt{white}\}$, $E_2 = \{\texttt{fish}, \texttt{red}\}$, and $E_3 = \{\texttt{meat}, \texttt{red}\}$.

The semantics of a PrAF is given by considering all possible worlds (i.e., AFs) obtained by removing consistent subsets of the probabilistic elements. Every possible world has associated a probability value derived from the probabilities of the elements that have been kept or removed. Moreover, every possible world admits a set of $\mathcal{S}$-extensions.

Example 2. Continuing with Example 1, the (non-zero probability) possible worlds of Δ are as follows, where for the sake of brevity, arguments are denoted by their initials:

- $w_1 = \langle\{\mathtt{f},\mathtt{m},\mathtt{w},\mathtt{r}\},\{(\mathtt{f},\mathtt{m}),(\mathtt{m},\mathtt{f}),(\mathtt{m},\mathtt{w}),(\mathtt{w},\mathtt{r}),(\mathtt{r},\mathtt{w})\}\rangle$;
- $w_2 = \langle\{\mathtt{f},\mathtt{m},\mathtt{r}\},\{(\mathtt{f},\mathtt{m}),(\mathtt{m},\mathtt{f})\}\rangle$;
- $w_3 = \langle\{\mathtt{m},\mathtt{w},\mathtt{r}\},\{(\mathtt{m},\mathtt{w}),(\mathtt{w},\mathtt{r}),(\mathtt{r},\mathtt{w})\}\rangle$;
- $w_4 = \langle\{\mathtt{m},\mathtt{r}\},\{\}\rangle$.

For instance, w_1 is the AF obtained from Δ by keeping all the arguments and attacks, while w_2 is obtained from Δ by removing white and, consistently with this, the attacks towards/from it. As it will be clear later, the probabilities of w_1, w_2, w_3, and w_4 are 0.48, 0.12, 0.32, and 0.08.

Since w_1 coincides with the deterministic version of Δ, its $\mathcal{ST}$-extensions are E_1, E_2, and E_3 given earlier. The $\mathcal{ST}$-extensions of w_2 are E_2 and E_3, while w_3 and w_4 admit only E_3 as their stable extension. □

An interesting problem recently investigated in the context of probabilistic argumentation is *probabilistic credulous acceptance* (Fazzinga, Flesca, and Furfaro 2018; Fazzinga, Flesca, and Furfaro 2019): Given a probabilistic framework Δ, whose set of arguments is A, and a semantics $\mathcal{S}$, compute the probability $PrCA^{\mathcal{S}}_{\Delta}(g)$ that a goal argument $g \in A$ is credulously accepted, that is, there is a possible world w of Δ where g belongs to an $\mathcal{S}$-extension of w. However, the answer to this problem does not reflect our intuition of probability that a goal argument is accepted under a given semantics. For instance, considering the PrAF Δ of Figure 1, the probability that meat is credulously accepted, under stable semantics, is 1. This means that the person in our example will surely have meat (since meat belongs to at least one $\mathcal{ST}$-extension of every world of Δ), whereas we expect that the probability of acceptance of meat should be less than 1, as in a possible world, the presence of multiple extensions is an additional source of uncertainty, that one should take into account.

To better grasp the issue behind the probability of credulous acceptance, consider the following AF (where all elements are certain): $\Delta' = \langle\{\mathtt{fish},\mathtt{meat}\},\{(\mathtt{fish},\mathtt{meat}),(\mathtt{meat},\mathtt{fish})\}\rangle$ saying that fish and meat are mutually exclusive. Again, the probability that a person will have meat is 1, under probabilistic credulous acceptance, when considering the stable semantics, whereas we believe that the expected answer should be 0.5.

With the aim of providing more intuitive answers for probabilistic acceptance, in this paper we investigate a new problem that we call *Probabilistic Acceptance* (denoted as PrA[$\mathcal{S}$]), i.e., given a PrAF Δ and a goal argument g, compute the probability that g is accepted under semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. In our framework, acceptance still relies on $\mathcal{S}$-extensions but, differently from credulous acceptance, we get rid of the assumption that no uncertainty exists at the level of the extensions of a world (i.e., an AF). In more detail, PrA[$\mathcal{S}$] implicitly assumes that a PDF over the set of $\mathcal{S}$-extensions of any AF (and thus of any possible world of PrAF Δ) is defined. Thus, a concrete instance of PrA[$\mathcal{S}$] is obtained after defining such a PDF.

In this paper, we explore an instantiation of PrA[$\mathcal{S}$] where the PDF over the $\mathcal{S}$-extensions of a world relies on the concept of *explanation*. We call this problem *Explanation-based Probabilistic Acceptance*, and denote it by PrEA[$\mathcal{S}$].

	General PrAFs		PrAFs without odd cycles	
	FPRAS	FPARAS	FPRAS	FPARAS
$\mathcal{GR}$	×	✓	×	✓
$\mathcal{PR}$	×	×	×	✓
$\mathcal{ST}$	×	×	×	✓
$\mathcal{SST}$	×	×	×	✓

Table 1: Approximability of PrEA[$\mathcal{S}$], depending on the semantics $\mathcal{S}$ and on whether the input PrAF admits odd-length cycles. Non-existence (resp., existence) of an FP(A)RAS is denoted with × (resp., ✓), in the corresponding column.

Intuitively, an explanation for an $\mathcal{S}$-extension E is a *sequence of arguments* occurring in E that 'justify' E. Every explanation is associated with a probability entailed by the possible choices that can be made. These choices must be consistent with an ordering entailed by the strongly connected components of the given AF, and they are used to guide the construction of an extension. The sum of the probabilities of the explanations for an extension E gives the probability of E. Thus, we still assign to each possible world w of Δ a probability as in the standard way, but in addition propose to distinguish among extensions of a given world w by associating with them a probability based on explanations.

Example 3. Continuing with Example 1, take for instance the possible world w_1 having probability 0.48. As shown in Example 2, w_1 has 3 $\mathcal{ST}$-extensions, namely E_1, E_2 and E_3. As we shall see, in this case, for each extension there is only one explanation. In particular, $X_1 = \langle\mathtt{fish},\mathtt{white}\rangle$ is the explanation for E_1. The intuition of explanation X_1 is that, considering that the AF consists of two strongly connected components, we first choose fish (with probability 1/2 as we can only choose between fish and meat) in the first component and determine that meat cannot belong to the extension; then we choose white (with probability 1/2 as we can only choose between white and red) in the second component, obtaining that X_1 has probability $1/2 \cdot 1/2 = 1/4$. Analogously, $X_2 = \langle\mathtt{fish},\mathtt{red}\rangle$ is the only explanation for E_2 with probability $1/2 \cdot 1/2 = 1/4$. Considering explanation $X_3 = \langle\mathtt{meat}\rangle$ for extension E_3, we have that we first choose meat with probability 1/2 as it belongs to the first component, and we can only choose between fish with meat. Next, since we determine that fish and white cannot belong to the extension, whereas red does, the probability of X_3 turns out to be 1/2. Since the probabilities of X_1, X_2 and X_3 are $1/4, 1/4$ and $1/2$, respectively, the probabilities associated with E_1, E_2 and E_3 in the world w_1 are $1/4, 1/4$ and $1/2$, respectively. Moreover, since E_1 is not an extension of any other possible world, the probability of E_1 in Δ is $1/4 \cdot 0.48 = 0.12$. In Example 8, we will give the probabilities of the $\mathcal{ST}$-extensions of every possible world of the probabilistic AF Δ of Example 1, from which it turns out that the answer to PrEA[$\mathcal{ST}$] for meat is 0.70, while that for fish is 0.30. □

Contributions. In this paper we tackle a new problem that we call *Probabilistic Acceptance* of an argument.

- We first formally define the problem of Probabilistic Acceptance $\mathsf{PrA}[\mathcal{S}]$, for some semantics $\mathcal{S}$. Given a PrAF Δ and an argument g, the problem asks the probability that g is accepted in Δ, by means of some fixed PDF over the $\mathcal{S}$-extensions of the possible worlds of Δ.

- Then, we introduce our notion of explanation, and exploit it to provide a PDF over the $\mathcal{S}$-extensions of an AF. This leads to an instantiation of $\mathsf{PrA}[\mathcal{S}]$, dubbed $\mathsf{PrEA}[\mathcal{S}]$.
- We then investigate the complexity of $\mathsf{PrA}[\mathcal{S}]$, showing that it is $FP^{\#P}$-hard for $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, even for acyclic PrAFs and regardless of the way a PDF is defined on $\mathcal{S}$-extensions. This entails that $\mathsf{PrEA}[\mathcal{S}]$ is as hard as the problem of computing credulous acceptance (Fazzinga, Flesca, and Furfaro 2019).
- To deal with the intractability of $\mathsf{PrA}[\mathcal{S}]$ (and of $\mathsf{PrEA}[\mathcal{S}]$), we propose an *additive approximation algorithm* for $\mathsf{PrEA}[\mathcal{S}]$ for probabilistic AFs without odd-length cycles and semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, and an *additive approximation algorithm* for $\mathsf{PrEA}[\mathcal{GR}]$ for general PrAFs (without the restriction on odd-length cycles).
- We show that our approximation result is the best it can be achieved (under standard theoretical assumptions), since (*i*) no *relative* approximation algorithm exists for $\mathsf{PrA}[\mathcal{S}]$ (and thus for $\mathsf{PrEA}[\mathcal{S}]$) for $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, even considering acyclic PrAFs, and (*ii*) if we admit odd-length cycles, then no *additive* approximation algorithm exists for $\mathsf{PrA}[\mathcal{S}]$ (and thus for $\mathsf{PrEA}[\mathcal{S}]$) with $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$ (i.e, it exists only for $\mathcal{GR}$). Table 1 summarizes our approximability results for $\mathsf{PrEA}[\mathcal{S}]$.
- We experimentally show that our approximate algorithm performs well in practice, even on large PrAFs obtained from AFs used in the last International Competition on Computational Models of Argumentation (ICCMA) [1].

It is worth noting that the existence of an approximation algorithm for $\mathsf{PrEA}[\mathcal{S}]$ is quite surprising since—as we show, as a side contribution—no approximation algorithm of any kind (relative or additive) exists for the related problem of probabilistic credulous acceptance, even considering restrictions on odd-length cycles.

2 Argumentation Frameworks

An abstract *Argumentation Framework* (AF) is a pair $\langle A, \Sigma\rangle$, where A is a set of *arguments* and $\Sigma \subseteq A \times A$ is a set of *attacks*. An AF can be seen as a directed graph, whose nodes represent arguments and edges represent attacks; an attack $(a, b) \in \Sigma$ from a to b is represented by $a \rightarrow b$. We shall use the notations a^+ and a^- for the sets $\{b \mid (a, b) \in \Sigma\}$ and $\{b \mid (b, a) \in \Sigma\}$, respectively. Further, for any $\mathbf{S} \subseteq A$, we denote as $\mathbf{S}^+$ and $\mathbf{S}^-$ the sets $\bigcup_{a\in\mathbf{S}} a^+$ and $\bigcup_{a\in\mathbf{S}} a^-$, respectively.

Different argumentation semantics have been defined leading to the characterization of collectively acceptable sets of arguments, called *extensions* (Dung 1995).

Given an AF $\Lambda = \langle A, \Sigma\rangle$ and a set $\mathbf{S} \subseteq A$ of arguments, an argument $a \in A$ is said to be *i*) *defeated* w.r.t. $\mathbf{S}$ iff $\exists b \in \mathbf{S}$ such that $(b, a) \in \Sigma$, and *ii*) *acceptable* w.r.t. $\mathbf{S}$ iff for every argument $b \in A$ with $(b, a) \in \Sigma$, there is $c \in \mathbf{S}$ such that $(c, b) \in \Sigma$. The sets of arguments defeated and acceptable w.r.t. $\mathbf{S}$ are as follows (where Λ is understood):

- $Def(\mathbf{S}) = \{a \in A \mid \exists b \in \mathbf{S} \,.\, (b, a) \in \Sigma\}$;
- $Acc(\mathbf{S}) = \{a \in A \mid \forall b \in A \,.\, (b, a) \in \Sigma \Rightarrow b \in Def(\mathbf{S})\}$.

Given an AF $\langle A, \Sigma\rangle$, a set $\mathbf{S} \subseteq A$ of arguments is said to be *conflict-free* iff $\mathbf{S} \cap Def(\mathbf{S}) = \emptyset$. Moreover, $\mathbf{S} \subseteq A$ is said to be a *complete* extension iff it is conflict-free and $\mathbf{S} = Acc(\mathbf{S})$. Given an AF $\langle A, \Sigma\rangle$, a complete extension $\mathbf{S} \subseteq A$ is said to be:

- *preferred* ($\mathcal{PR}$) iff it is maximal (w.r.t. $\subseteq$);
- *stable* ($\mathcal{ST}$) iff it is a total preferred extension, i.e. a preferred extension such that $\mathbf{S} \cup Def(\mathbf{S}) = A$;
- *semi-stable* ($\mathcal{SST}$) iff it is a preferred extension with a maximal set of decided elements, i.e. a preferred extension such that $\mathbf{S} \cup Def(\mathbf{S})$ is maximal;
- *grounded* ($\mathcal{GR}$) iff it is minimal (w.r.t. $\subseteq$).

The set of preferred (resp. stable, semi-stable, grounded) extensions of an AF Λ will be denoted by $\mathcal{PR}(\Lambda)$ (resp. $\mathcal{ST}(\Lambda)$, $\mathcal{SST}(\Lambda)$, $\mathcal{GR}(\Lambda)$). In the following, whenever we consider a generic semantics $\mathcal{S}$, we refer to a semantics in $\{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$.

Example 4. Consider the AF Λ derived from the probabilistic AF of Example 1, where all arguments are certain (Λ coincides with the possible world w_1 of Example 2). The set of all complete extensions of Λ is $\{\emptyset, \{\texttt{fish}\}, \{\texttt{red}\}, \{\texttt{fish, white}\}, \{\texttt{fish, red}\}, \{\texttt{meat, red}\}\}$. Thus, $\mathcal{PR}(\Lambda) = \mathcal{ST}(\Lambda) = \mathcal{SST}(\Lambda) = \{\{\texttt{fish, white}\}, \{\texttt{meat, red}\}, \{\texttt{fish, red}\}\}$, whereas $\mathcal{GR}(\Lambda) = \{\emptyset\}$. □

Let $\Lambda = \langle A, \Sigma\rangle$ be an AF. A *strongly connected component (SCC)* of Λ is a maximal subset C of A such that, for every pair $a, b \in C$, there is a path[2] from a to b in Λ. An *ordering* of Λ is a sequence $C_1, \ldots, C_n$ of all SCCs of Λ, such that for each $1 \leq i < j \leq n$, no argument in C_j attacks arguments in C_i. An SCC C of Λ is *initial w.r.t. a set* $E \subseteq A$ if there is an ordering of Λ in which C is the first SCC such that $C \cap E \neq \emptyset$, We simply say that C is *initial*, if no argument in $A \setminus C$ attacks arguments in C.

For instance, the SCCs of the AF Λ of Example 4 are $\{\texttt{fish, meat}\}$ (which is initial) and $\{\texttt{white, red}\}$. Note that, differently from the standard definition, we use SCC to denote a set of nodes (i.e., arguments), not a subgraph.

In the following, given an AF $\Lambda = \langle A, \Sigma\rangle$ and a set $\mathbf{S} \subseteq A$ of arguments, we define $\Lambda\downarrow_{\mathbf{S}} = \langle \mathbf{S}, \Sigma \cap (\mathbf{S} \times \mathbf{S})\rangle$ as the *restriction* of Λ to the set $\mathbf{S}$.

3 Probabilistic Argumentation Frameworks

In general a probabilistic argumentation framework consists of probabilistic arguments and probabilistic attacks (Li, Oren, and Norman 2011; Fazzinga, Flesca, and Parisi 2015; Fazzinga, Flesca, and Furfaro 2019). However, w.l.o.g. we can focus on *Probabilistic Argumentation Frameworks* (PrAFs) where only arguments are uncertain (and attacks are certain, i.e., their probability is 1), since as shown in (Mantadelis and Bistarelli 2020) an argumentation framework

[1] http://argumentationcompetition.org

[2] If $a = b$, a path trivially exists.

with probabilities on both arguments and attacks can be transformed into an equivalent PrAF.

Definition 1. *A Probabilistic Argumentation Framework (PrAF) is a triple $\langle A, \Sigma, P\rangle$ where $\langle A, \Sigma\rangle$ is an AF, and P is a function assigning a non-zero probability value to every argument in A, that is, $P : A \rightarrow (0, 1]$.*

Observe that assigning probability equal to 0 to arguments is useless. Basically, the value assigned by P to any argument a represents the probability that a actually occurs. Moreover, every attack (a, b) occurs with conditional probability 1, that is, a attacks b whenever both a and b occur.

The meaning of a PrAF is given in terms of *possible worlds*. Formally, given a PrAF $\Delta = \langle A, \Sigma, P\rangle$, a possible world of Δ is an AF $w = \langle A', \Sigma'\rangle$ such that $A' \subseteq A$ and $\Sigma' = \Sigma \cap (A' \times A')$. We use $pw(\Delta)$ to denote the set of all possible worlds of Δ.

Thus, an argument $a \in A$ is viewed as a probabilistic event which is independent from the other events associated with other arguments $b \in A$ (with $b \neq a$).

An *interpretation* for a PrAF $\Delta = \langle A, \Sigma, P\rangle$ is a probability distribution function I over the set $pw(\Delta)$ of the possible worlds. Each $w = \langle A', \Sigma'\rangle \in pw(\Delta)$ is assigned by I the probability

$$I(w) = \prod_{a \in A} P(a) \cdot \prod_{a \in A \setminus A'} (1 - P(a)).$$

Example 5. The (non-zero probability) possible worlds of the PrAF Δ of Example 1 are w_1, w_2, w_3 and w_4 given in Example 2. Then, interpretation I is as follows:
$I(w_1) = P(\texttt{fish}) \cdot P(\texttt{white}) = 0.6 \cdot 0.8 = 0.48$,
$I(w_2) = P(\texttt{fish}) \cdot (1 - P(\texttt{white})) = 0.6 \cdot 0.2 = 0.12$,
$I(w_3) = (1 - P(\texttt{fish})) \cdot P(\texttt{white}) = 0.4 \cdot 0.8 = 0.32$,
$I(w_4) = (1 - P(\texttt{fish})) \cdot (1 - P(\texttt{white})) = 0.4 \cdot 0.2 = 0.08$,
and $I(w) = 0$ for any other world $w \in pw(\Delta)$. □

A relevant problem in the field of formal argumentation is that of *credulous acceptance*, that is checking whether a goal argument g of an AF Λ is accepted under a given semantics $\mathcal{S}$, i.e. there exists an $\mathcal{S}$-extension of Λ containing g.

The analogous problem in the context of a probabilistic AF is the following.

Definition 2 (Probabilistic credulous acceptance)**.** *Given a PrAF Δ, an argument $g \in A$, the probability $PrCA^{\mathcal{S}}_{\Delta}(g)$ that g is credulously acceptable w.r.t $\mathcal{S}$ semantics is*

$$PrCA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ \exists E \in \mathcal{S}(w) \, s.t. \, g \in E}} I(w).$$

The probability that an argument g is credulously accepted according to a semantics $\mathcal{S}$ is defined as the sum of the probabilities of the possible worlds w of a PrAF Δ for which argument g is credulously accepted. Computing $PrCA^{\mathcal{S}}_{\Delta}(g)$ is $FP^{\#P}$-hard for all semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$ (Fazzinga, Flesca, and Furfaro 2018).

As discussed in the introduction, probabilistic credulous acceptance does not express the probability that a given argument is accepted. Therefore, in this paper we study a new problem, called *Probabilistic Acceptance*, which can be intuitively stated as follows. Given a probabilistic framework Δ, a semantics $\mathcal{S}$, and a goal argument g, compute the probability that g is accepted. However, differently from previously proposed probabilistic measures, considering a possible world w having probability $I(w)$, under the given semantics $\mathcal{S}$, every extension $E \in \mathcal{S}(w)$ has associated a probability $Pr(E, w, \mathcal{S})$ so that $\sum_{E \in \mathcal{S}(w)} Pr(E, w, \mathcal{S}) = 1$ (the sum of the probabilities of the $\mathcal{S}$-extensions of w is equal to 1) and $Pr(E, w, \mathcal{S}) = 0$ for all $E \notin \mathcal{S}(w)$. In more detail, as stated next, we require a PDF over the set of extensions.

Definition 3 (Probabilistic Acceptance)**.** *Given a PrAF $\Delta = \langle A, \Sigma, P\rangle$ and an argument $g \in A$, the probability $PrA^{\mathcal{S}}_{\Delta}(g)$ that g is acceptable w.r.t. semantics $\mathcal{S}$ is*

$$PrA^{\mathcal{S}}_{\Delta}(g) = \sum_{\substack{w \in pw(\Delta) \wedge \\ E \in \mathcal{S}(w) \wedge g \in E}} I(w) \cdot Pr(E, w, \mathcal{S})$$

where $Pr(\cdot, w, \mathcal{S})$ is a PDF over the set $\mathcal{S}(w)$.

Our definition of probabilistic acceptance generalizes the notion of (probablistic) credulous acceptance for deterministic AFs proposed in (Thimm 2012), where a PDF over the set of $\mathcal{S}$-extensions is assumed to be given. Besides considering PrAFs, we propose an approach based on explanations which entails a PDF on the set of $\mathcal{S}$-extensions of a PrAF.

4 Explanations

In this section, we present how the probability of an extension E for an AF Λ under semantics $\mathcal{S}$ (denoted as $Pr(E, \Lambda, \mathcal{S})$) can be defined. Based on this, we obtain a PDF over the set of $\mathcal{S}$-extensions of every possible world w of a PrAF that will be then used to provide a concrete instantiation the probabilistic acceptance problem. It is important to note that here we are proposing a way for defining such a PDF, but other ways can be devised—our approach is not meant to be the ultimate solution to the problem of defining such a PDF, though we believe it is a reasonable one. For instance, compared with the uniform distribution, the proposed PDF is such that, as it will be clear in the following (cf. Example 8), considering the AF obtained from the deterministic version of the PrAF in Figure 1 (i.e., world w_1), the probability of having one of the two mutually exclusive arguments `fish` or `meat` (under stable/preferred/semi-stable semantics) is the same and equal to $1/2$. In contrast, considering the uniform distribution, the probability of having `fish` (resp., `meat`) is $2/3$ (resp., $1/3$). Moreover, as stated in Theorem 4, the proposed PDF based on explanations allows us to obtain a tractable sampling strategy.

To define $Pr(E, \Lambda, \mathcal{S})$ we introduce the concept of *explanation* consisting of a sequence of necessary suggestions useful to construct a given extension E, that is a sequence of choices made to obtain the extension. In particular, the choices we consider are guided by an ordering entailed by strongly connected components (SCCs) of an AF. In fact, SCCs have been exploited in several approaches in argumentation since they are inherently related to intuitive properties of AFs (Baroni, Giacomin, and Guida 2005; Cerutti et al. 2014; Baroni et al. 2014; Rienstra et al. 2018).

Given an AF $\Lambda = \langle A, \Sigma\rangle$ and its grounded extension $G \in \mathcal{GR}(\Lambda)$, we use $\hat{\Lambda}$ to denote Λ without G and G^+, i.e., the

AF $\Lambda\downarrow_{A\setminus(G\cup G^+)}$. Moreover, for an argument $a \in A$, we use $\hat{\Lambda}_a$ to denote Λ without G, G^+ and the attackers of a, i.e., the AF $\Lambda\downarrow_{A\setminus(G\cup G^+\cup\{a\}^-)}$.

An explanation for a given extension is defined as follows.

Definition 4 (Explanation). *Let $\Lambda = \langle A, \Sigma\rangle$ be an AF, and E an $\mathcal{S}$-extension of Λ, for some semantics $\mathcal{S}$. A sequence $X = \langle a_1, \ldots, a_n\rangle$ of arguments in E is an explanation for E (w.r.t. Λ), if either X is empty and $E \in \mathcal{GR}(\Lambda)$, or*

i) a_1 belongs to some initial SCC of $\hat{\Lambda}$ w.r.t. E, and

ii) $\langle a_2, \ldots, a_n\rangle$ is an explanation for $E \setminus \mathcal{GR}(\Lambda)$ w.r.t $\hat{\Lambda}_{a_1}$.

Intuitively, an explanation X for an $\mathcal{S}$-extension E of Λ is recursively defined as a sequence of chosen arguments such that *i)* every argument belongs to E, and *ii)* every argument choice is not trivial, in the sense that arguments can be chosen only among those belonging to an initial SCC of a restricted AF obtained by removing arguments determined by the grounded semantics (and by previous choices). Basically, such an initial SCC consists of arguments whose acceptance status cannot be determined on the basis of the acceptance status of arguments determined so far. It is worth noting that the definition entails that an explanation X of E is such that there is no explanation X' of E which is a proper prefix of X. Moreover, for the grounded semantics, which admits a unique extension, there exists a unique explanation which is the empty sequence $\langle\rangle$.

Example 6. Continuing with Example 4, we have that for the stable extension $E_1 = \{\mathtt{fish}, \mathtt{white}\}$ of Λ there is only one explanations $X_1 = \langle\mathtt{fish}, \mathtt{white}\rangle$. In fact, since the grounded extension G of Λ is empty, it does not help to determine any argument of the initial AF, i.e., $\hat{\Lambda} = \Lambda$, and fish can be chosen in the initial SCC of $\hat{\Lambda}$ w.r.t. E_1 (which coincides with the initial SCC of $\hat{\Lambda}$). Next, we look for an explanation for $\{\mathtt{fish}, \mathtt{white}\}$ w.r.t $\hat{\Lambda}_{\mathtt{fish}} = \langle\{\mathtt{fish}, \mathtt{white}, \mathtt{red}\}, \{(\mathtt{white}, \mathtt{red}), (\mathtt{red}, \mathtt{white})\}\rangle$, where the attackers of fish have been removed (no other argument is removed since $G = \emptyset$). Now, since the grounded extension of $\hat{\Lambda}_{\mathtt{fish}}$ consists of fish only, we consider the AF Λ' obtained from $\hat{\Lambda}_{\mathtt{fish}}$ by removing fish (no argument is attacked by fish in the current AF). Then, the initial SCC of Λ' w.r.t. $\{\mathtt{fish}, \mathtt{white}\}$ is $\{\mathtt{white}, \mathtt{red}\}$, and the only argument of our extension that we can choose from this SCC is white. Thus, we end up with the AF Λ'' obtained from Λ' by removing the attackers of white, and the set $\{\mathtt{white}\}$, which is the grounded extension of Λ''. Hence, $X_1 = \langle\mathtt{fish}, \mathtt{white}\rangle$ is an explanation of $\{\mathtt{fish}, \mathtt{white}\}$.

Consider now explanation $X_3 = \langle\mathtt{meat}\rangle$ for the stable extension $E_3 = \{\mathtt{meat}, \mathtt{red}\}$. Herein, $\hat{\Lambda} = \Lambda$ as said earlier, and meat can be chosen in the initial SCC of $\hat{\Lambda}$ w.r.t. E_3 (which again coincides with the initial SCC). Next, we look for an explanation for $\{\mathtt{meat}, \mathtt{red}\}$ w.r.t $\hat{\Lambda}_{\mathtt{meat}} = \langle\{\mathtt{meat}, \mathtt{white}, \mathtt{red}\}, \{(\mathtt{meat}, \mathtt{white}), (\mathtt{white}, \mathtt{red}), (\mathtt{red}, \mathtt{white})\}\rangle$, where the attackers of meat have been removed. Now, since the grounded extension of $\hat{\Lambda}_{\mathtt{meat}}$ is $\{\mathtt{meat}, \mathtt{red}\}$, we conclude that $X_3 = \langle\mathtt{meat}\rangle$ is an explanation for E_3. □

Proposition 1. *Let $\Lambda = \langle A, \Sigma\rangle$ be an AF, $\mathcal{S}$ a semantics, and $X = \langle a_1, \ldots, a_n\rangle$ an explanation for some $\mathcal{S}$-extension of Λ. Then, i) there exists a unique extension in $\mathcal{S}(\Lambda)$ for which X is an explanation and ii) for each $\mathcal{S}$-extension E of Λ there always exists at least one explanation for E.*

Proof. (Sketch) For $\mathcal{S} = \mathcal{GR}$, the claim follows by Definition 4. It suffices to prove the claim for $\mathcal{S} = \mathcal{PR}$. We can show that (*) for an AF Λ, a preferred extension E of Λ, and for any argument $a \in E$ belonging to some initial SCC of $\hat{\Lambda}$ w.r.t. E, then $E \setminus G$ is a preferred extension of $\hat{\Lambda}_a$, where $G \in \mathcal{GR}(\Lambda)$. Moreover, if no initial SCC of $\hat{\Lambda}$ w.r.t. E exists, then $E \in \mathcal{GR}(\Lambda)$.

With the above, to prove *i*) suffices to note that any two preferred extensions E_1, E_2, explained by some X, always get removed a set of arguments which is common to both, at each recursive step of Definition 4 (i.e., the set $G \in \mathcal{GR}(\Lambda)$, by claim (*)). Moreover, at the last step, since X is an explanation of both E_1 and E_2 w.r.t. Λ, both extensions necessarily coincide with the grounded extension, and thus $E_1 = E_2$. For proving *ii*), it suffices to note that Definition 4 naturally induces a recursive procedure for constructing a sequence X of arguments, starting from a preferred extension E of some AF Λ. The constructed sequence X is an explanation of E as at the last recursive step of Definition 4, no initial SCC of the remaining AF w.r.t. the remaining extension exists, and by claim (*), it follows that the remaining extension is the grounded extension of the remaining AF. □

In the following, the set of explanations for an $\mathcal{S}$-extension E of an AF Λ is denoted by $Exp^{\mathcal{S}}_{\Lambda}(E)$. We assume that whenever $\mathbf{S} \notin \mathcal{S}(\Lambda)$, $Exp^{\mathcal{S}}_{\Lambda}(\mathbf{S}) = \emptyset$. Moreover, $Exp^{\mathcal{S}}(\Lambda) = \bigcup_{E\in\mathcal{S}(\Lambda)} Exp^{\mathcal{S}}_{\Lambda}(E)$ is the set of explanations of AF Λ under semantics $\mathcal{S}$.

Since a given extension may have multiple explanations of different length, it is reasonable to assume that some explanations are preferred to others. We now introduce probabilities for explanations. As said before, the grounded semantics has a unique explanation which has probability 1. To define probabilities of explanations, we exploit the concept of probabilistic trie.

Definition 5. *Given an AF Λ and a semantics $\mathcal{S}$, the probabilistic trie for Λ under semantics $\mathcal{S}$ is the triple $\mathcal{T}^{\mathcal{S}}_{\Lambda} = \langle N, H, \pi\rangle$ of nodes N and edges H where $\langle N, H\rangle$ is the trie of all sequences in $Exp^{\mathcal{S}}(\Lambda)$, and $\pi : N \to (0, 1]$ is the function inductively defined as follows:*

- $\pi(\langle\rangle) = 1$,
- $\pi(x) = \pi(parent(x)) \cdot 1/|children(parent(x))|$,

where $parent(x)$ denotes the parent of x, whereas $|children(x)|$ denotes the number of children of x.

Since the set of leaves of the probabilistic trie $\mathcal{T}^{\mathcal{S}}_{\Lambda} = \langle N, H, \pi\rangle$ coincides with $Exp^{\mathcal{S}}(\Lambda)$, hereafter, with a little abuse of notation, we assume that π is a function from $Exp^{\mathcal{S}}(\Lambda)$ to $(0, 1]$.

As defined next, the probability that a set $\mathbf{S}$ of arguments is an $\mathcal{S}$-extension of an AF Λ is given by the sum of the probabilities of the explanations for $\mathbf{S}$ under semantics $\mathcal{S}$.

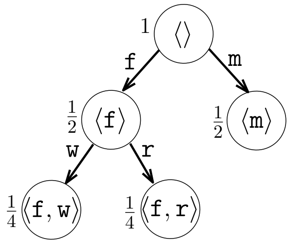

Figure 2: Probabilistic trie for the AF Λ of Example 4 under preferred/stable/semi-stable semantics.

Definition 6. *Given an AF $\Lambda = \langle A, \Sigma \rangle$, a semantics $\mathcal{S}$, and a set of arguments $\mathbf{S} \subseteq A$, then $Pr(\mathbf{S}, \Lambda, \mathcal{S}) = \sum_{X \in Exp^{\mathcal{S}}_{\Lambda}(\mathbf{S})} \pi(X)$.*

Observe that $Pr(\mathbf{S}, \Lambda, \mathcal{S}) = 0$ whenever $\mathbf{S}$ is not an $\mathcal{S}$-extension of Λ. With a little effort, it can be checked that function $Pr(\cdot, \Lambda, \mathcal{S})$ of Definition 6 is a PDF over the set of $\mathcal{S}$-extensions of Λ. Therefore, we define $PrEA^{\mathcal{S}}_{\Delta}(g)$ as the probability obtained using such a PDF in Definition 3. That is, we have obtained an instantiation of our Probabilistic Acceptance problem, that we call *Explanation-based Probabilistic Acceptance* problem, whose output is $PrEA^{\mathcal{S}}_{\Delta}(g)$.

Example 7. Let Λ be the AF of Example 4. The explanations for the preferred (stable and semi-stable) extensions are represented by the leaf nodes of the trie shown in Fig. 2 (where arguments are denoted by their initials). For instance, the probability of explanation $\langle \mathtt{f}, \mathtt{w} \rangle$ for extension $E_1 = \{\mathtt{f}, \mathtt{w}\}$ (cf. Example 6) is $1/4$. Therefore the probability of E_1 is $1/4$. □

Example 8. Consider the PrAF Δ of Example 1. As shown in Example 2, for Δ there are four (non-zero probability) possible worlds whose probabilities are given in Example 5.

Let $E_1 = \{\mathtt{f}, \mathtt{w}\}$, $E_2 = \{\mathtt{f}, \mathtt{r}\}$ and $E_3 = \{\mathtt{m}, \mathtt{r}\}$. We have that $\mathcal{ST}(w_1) = \{E_1, E_2, E_3\}$, $\mathcal{ST}(w_2) = \{E_2, E_3\}$, $\mathcal{ST}(w_3) = \{E_3\}$, and $\mathcal{ST}(w_4) = \{E_3\}$. The following table reports for each world w the probability $I(w)$ (second column) and, for each pair $\langle$world w, stable extension $E\rangle$, the probability $Pr(E, w, \mathcal{ST})$ w.r.t. the AF w (last three columns). Finally, the last row of the table reports, for each set E, the probability that E is an $\mathcal{ST}$-extension of the PrAF Δ, defined for a semantics $\mathcal{S}$ as $\sum_{w \in pw(\Delta)} I(w) \cdot Pr(E, w, \mathcal{S})$.

w	$I(w)$	$E_1 = \{\mathtt{f}, \mathtt{w}\}$ $Pr(E, w, \mathcal{ST})$	$E_2 = \{\mathtt{f}, \mathtt{r}\}$ $Pr(E, w, \mathcal{ST})$	$E_3 = \{\mathtt{m}, \mathtt{r}\}$ $Pr(E, w, \mathcal{ST})$
w_1	0.48	1/4	1/4	1/2
w_2	0.12	0	1/2	1/2
w_3	0.32	0	0	1
w_4	0.08	0	0	1
		0.12	0.18	0.70

Using Definition 3, the probability of acceptance of a goal argument in Δ is as follows. $PrEA^{\mathcal{ST}}_{\Delta}(\mathtt{fish}) = 0.48 \times 1/4 + 0.48 \times 1/4 + 0.12 \times 1/2 = 0.30$, $PrEA^{\mathcal{ST}}_{\Delta}(\mathtt{meat}) = 0.48 \times 1/2 + 0.12 \times 1/2 + 0.32 \times 1 + 0.08 \times 1 = 0.70$. Similarly, we obtain $PrEA^{\mathcal{ST}}_{\Delta}(\mathtt{white}) = 0.12$ and $PrEA^{\mathcal{ST}}_{\Delta}(\mathtt{red}) = 0.88$.

5 Exact and Approximate Complexity

This section discusses the complexity of the following problems; $\mathcal{S}$ is a semantics in $\{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$.

PROBLEM :	$\mathsf{PrA}[\mathcal{S}]$
INPUT :	A PrAF Δ and an argument g.
OUTPUT :	The number $PrA^{\mathcal{S}}_{\Delta}(g)$.

PROBLEM :	$\mathsf{PrEA}[\mathcal{S}]$
INPUT :	A PrAF Δ and an argument g.
OUTPUT :	The number $PrEA^{\mathcal{S}}_{\Delta}(g)$.

We recall that $\mathsf{PrA}[\mathcal{S}]$ is defined after choosing an arbitrary but fixed PDF over the set of extensions of an AF, while $\mathsf{PrEA}[\mathcal{S}]$ uses the specific PDF $Pr(\cdot, \Lambda, \mathcal{S})$ of Definition 6.

We show that for all semantics, the above problems are intractable. In particular, we show that $\mathsf{PrA}[\mathcal{S}]$ is $FP^{\#P}$-hard, regardless of the chosen PDF, from which it follows that $\mathsf{PrEA}[\mathcal{S}]$ is $FP^{\#P}$-hard as well.

Theorem 1. *For $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, $\mathsf{PrA}[\mathcal{S}]$ is $FP^{\#P}$-hard, even for acyclic PrAFs and for any chosen PDF.*

Proof. (Sketch) We reduce from the $FP^{\#P}$-hard problem #P2CNF (Welsh and Gale 2001) of counting the number of satisfying assignments $\#\phi$ of a CNF formula ϕ, where each clause consists of exactly 2 positive literals.

Let $\phi = C_1 \wedge C_2 \wedge \cdots \wedge C_k$ be a P2CNF and X the set of its n propositional variables. We define the (acyclic) PrAF $\Delta = \langle A, \Sigma, P \rangle$ as follows. The set A contains an argument a_x for each $x \in X$; an argument c_i for each clause C_i; and an argument φ. Function P assigns probability $\frac{1}{2}$ to arguments a_x, $\forall x \in X$, and 1 otherwise. Relation Σ contains, for each clause $C_i = x \vee y$, an attack (c_i, φ), and two attacks (a_x, c_i) and (a_y, c_i). Finally, we let the goal argument be φ.

We can show that there is a bijection between the truth assignments of ϕ and the worlds of Δ. Moreover, since the PrAF $\Delta = \langle A, \Sigma, P \rangle$ is acyclic, every AF $w \in pw(\Delta)$ is acyclic. Thus the grounded extension of w is its unique $\mathcal{S}$-extension, and any PDF $Pr(\cdot, w, \mathcal{S})$ trivially assigns 1 to that extension. Finally, for each world w of Δ, φ belongs to the (only) $\mathcal{S}$-extension of w iff the corresponding truth assignment satisfies ϕ. Hence, $\#\phi = 2^n \cdot PrA^{\mathcal{S}}_{\Delta}(\varphi)$, and the claim follows. □

The high computational complexity of $\mathsf{PrA}[\mathcal{S}]$ (and thus of $\mathsf{PrEA}[\mathcal{S}]$), for all semantics $\mathcal{S}$, even for very simple settings, such as acyclic PrAFs, suggests that one would need to focus on finding efficient algorithms that solve the problem approximately. Next, we present a complete picture of the approximability landscape of our problems, under different semantics and approximation schemes.

We start by defining the kind of approximation schemes we are going to target.

Definition 7 **((Arora and Barak 2009)).** Consider a function $f : \{0, 1\}^* \to \mathbb{Q}$. A *fully polynomial-time randomized approximation scheme* (FPRAS) for f is a randomized algorithm A that given as input $x \in \{0, 1\}^*$, and numbers $\epsilon > 0$,

$0 < \delta < 1$, outputs a random variable $A(x, \epsilon, \delta)$ such that:

$$\mathsf{Pr}\left(|A(x, \epsilon, \delta) - f(x)| \leq \epsilon \cdot f(x)\right) \geq 1 - \delta,$$

and A runs in polynomial time in $|x|$, $1/\epsilon$, and $\ln(1/\delta)$.

Similarly, we can define the notion of *additive* FPRAS. A *fully polynomial-time additive randomized approximation scheme* (FPARAS) for a function f is defined as in Definition 7, where the inequality $|A(x, \epsilon, \delta) - f(x)| \leq \epsilon \cdot f(x)$ is replaced with $|A(x, \epsilon, \delta) - f(x)| \leq \epsilon$.

5.1 Inapproximability Results

We start by discussing inapproximability results for $\mathsf{PrA}[\mathcal{S}]$ (and $\mathsf{PrEA}[\mathcal{S}]$) in the FP(A)RAS sense.

For this, we need to recall the class of decision problems BPP. A decision problem Π is in BPP iff there exists a polynomial-time randomized decision procedure A such that, for every instance $x \in \{0,1\}^*$ of Π, if x is a yes (resp., no) instance of Π, then $Pr(A(x) = \text{yes})$ (resp., $Pr(A(x) = \text{no})$) is greater than or equal to $2/3$. It is known that NP $\subseteq$ BPP implies that the polynomial-time hierarchy collapses (Ko 1982).

Theorem 2. *Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPRAS for* $\mathsf{PrA}[\mathcal{S}]$, *even for acyclic PrAFs and for any chosen PDF.*

Proof. (Sketch) The claim follows from the fact that the reduction in the proof of Theorem 1 is from the problem #P2CNF, which has no FPRAS (unless NP $\subseteq$ BPP) (Welsh and Gale 2001), and that the constructed PrAF Δ and goal φ are such that $PrA^{\mathcal{S}}_{\Delta}(\varphi)$ equals the number of satisfying assignments of the formula, modulo a multiplying factor. ☐

Next, we show that for all semantics except for $\mathcal{GR}$, even approximation algorithms with bounded additive error cannot be devised, for PrAFs of general shape. For the proof, we rely on a technical lemma that shows a certain gap property of the problem of credulously accepting an argument. We first introduce some notions.

We say that a pair (Λ, g) of an AF Λ and argument g is $\mathcal{S}$-*uniform*, for a semantics $\mathcal{S}$, if the existence of an extension $E \in \mathcal{S}(\Lambda)$, such that $g \in E$, implies that every extension $E \in \mathcal{S}$ is such that $g \in E$.[3]

Let us now consider the following restriction of the classical credulous acceptance problem, where $\mathcal{S}$ is a semantics.

PROBLEM :	$\mathsf{UnCA}[\mathcal{S}]$
INPUT :	An $\mathcal{S}$-uniform pair (Λ, g).
QUESTION :	Is there $E \in \mathcal{S}(\Lambda)$ such that $g \in E$?

We show that even when restricting our attention to $\mathcal{S}$-uniform pairs of AFs and arguments, credulous acceptance is NP-hard, for all semantics in $\{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. The result is proved by providing a reduction from 3SAT by exploiting and adapting the construction (of an AF Λ) known in the literature for the credulous acceptance (Dunne and Bench-Capon 2002). Particularly, we prove that Λ has exclusively non-empty $\mathcal{S}$-extensions if the given formula is satisfiable, and exclusively empty $\mathcal{S}$-extensions otherwise.

[3]In other words, when a pair (Λ, g) is $\mathcal{S}$-uniform, then credulous and skeptical acceptance of g over Λ coincide.

Algorithm 1 Apx

Require: A PrAF $\Delta = \langle A, \Sigma, P \rangle$, a semantics $\mathcal{S}$, a goal argument $g \in A$, error parameter $\epsilon > 0$, and uncertainty parameter $0 < \delta < 1$.
Ensure: a random number p such that $PrEA^{\mathcal{S}}_{\Delta}(g) \in [p - \epsilon, p + \epsilon]$ with probability $1 - \delta$.

```
n := ⌈ 1/(2ε²) × ln(2/δ) ⌉;
c := 0;
for i ∈ {1, ..., n} do
    Choose w ∈ pw(Δ) with probability I(w);
    Choose E ∈ S(w) with probability Pr(E, w, S);
    if g ∈ E then
        c := c + 1;
return c/n;
```

Lemma 1. *For each $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$,* $\mathsf{UnCA}[\mathcal{S}]$ *is NP-hard.*

We can now exploit the above lemma to prove our inapproximability result.

Theorem 3. *Let $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPARAS for* $\mathsf{PrA}[\mathcal{S}]$, *for any chosen PDF.*

Proof. (Sketch). The main idea is to show that for $\mathcal{S}$-uniform pairs (Λ, g), one can always construct a PrAF Δ from Λ such that if (Λ, g) is a "yes" instance of $\mathsf{UnCA}[\mathcal{S}]$, then $PrA^{\mathcal{S}}_{\Delta}(g) = 1$, and if (Λ, g) is a "no" instance, $PrA^{\mathcal{S}}_{\Delta}(g) = 0$. Thus, by setting ϵ to be a sufficiently small error, an FPARAS for $\mathsf{PrA}[\mathcal{S}]$ can be used to distinguish between "yes" and "no" instances of an NP-hard problem (i.e., $\mathsf{UnCA}[\mathcal{S}]$) with high probability, and in polynomial time. The latter implies NP $\subseteq$ BPP. ☐

Theorems 1, 2, and 3 rule out the existence of polynomial-time algorithms for solving $\mathsf{PrA}[\mathcal{S}]$. In terms of exact and approximate computation via FPRASes, this is not possible even for acyclic PrAFs, whereas in terms of approximate computation via FPARASes, this is not possible for general PrAFs and for all semantics, besides $\mathcal{GR}$. Notably, our results highlight an intrinsic difficulty in providing efficient procedures (either exact or approximate) for *any* approach assigning a probability to an argument by means of a probability distribution over the extensions.

From the above discussion, it is clear that our efforts should be towards approximation schemes with bounded additive error guarantees, i.e., FPARASes. In particular, in the light of Theorem 3, one could still provide an FPARAS either when $\mathcal{S} = \mathcal{GR}$, or when some restriction on the input PrAF is assumed. In fact, we are going to show that either when $\mathcal{S} = \mathcal{GR}$ or when the input PrAF has no odd-length cycles, the use of explanations for devising a PDF over extensions allows us to construct an FPARAS.

5.2 Devising an FPARAS

We report an FPARAS for the problem $\mathsf{PrEA}[\mathcal{S}]$, when either $\mathcal{S} = \mathcal{GR}$ or the input PrAF has no odd-length cycles.

The general structure of our algorithm is presented in Algorithm 1. Consider a PrAF Δ, a semantics $\mathcal{S}$ and an argument g. The high-level idea is to perform a number of iterations n, and at each iteration sample a world w of Δ and an explanation X in $Exp^{\mathcal{S}}(w)$, and count the fraction of iterations for which the given argument g is in the $\mathcal{S}$-extension E explained by X.

We point out that, besides line 5, all steps of our algorithm can be easily implemented in polynomial time regardless of the shape of the input PrAF and the semantics. Particularly, to prove that Algorithm 1 leads to an FPARAS in the cases described above, it suffices to prove that i) line 5 can be implemented in polynomial time when either $\mathcal{S} = \mathcal{GR}$ or Δ has no odd-length cycles. This is done via Algorithm 2, and ii) Algorithm 1 enjoys the probabilistic and error guarantees of an FPARAS (this can be proved via standard probabilistic inequalities (Hoeffding 1963)). We first show that Algorithm 2 efficiently implements line 5 of Algorithm 1.

Theorem 4. *Algorithm 2 with input an AF Λ and a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$ is such that:*

- *For each $E \in \mathcal{S}(\Lambda)$, it outputs E with probability $Pr(E, \Lambda, \mathcal{S})$, and*
- *it runs in polynomial time,*

whenever i) $\mathcal{S} = \mathcal{GR}$, or ii) Λ has no odd-length cycles.

Proof. (Sketch). The algorithm runs in polynomial time as the number of iterations is linear in the number of arguments and each iteration performs polynomial time operations. When $\mathcal{S} = \mathcal{GR}$ the claim follows trivially, as $E = \mathcal{GR}(\Lambda)$ is the grounded extension of Λ. Otherwise, considering AFs Λ without odd-length cycles, we focus w.l.o.g. on the case $\mathcal{S} = \mathcal{ST}$. We can show that every argument of every initial SCC of $\hat{\Lambda}$ belongs to some stable extension of $\hat{\Lambda}$, and hence of Λ, and every stable extension of Λ contains at least one argument of every initial SCC of $\hat{\Lambda}$. The probability of outputting an extension E coincides with the probability of choosing an argument for each execution of the cycle, which form an explanation X for E w.r.t. Λ whose probability coincides with the value associated with the leaf node corresponding to X in the probabilistic trie $\mathcal{T}_{\Lambda}^{\mathcal{ST}}$. □

By exploiting the above result, and standard probabilistic inequalities (Hoeffding 1963), we can prove our main approximability result.

Theorem 5. *Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. The problem* $\mathsf{PrEA}[\mathcal{S}]$ *has an FPARAS if either i) $\mathcal{S} = \mathcal{GR}$, or ii) the input PrAF has no odd cycles.*

6 Inapproximability For Previous Approaches

In this section, we consider the problem of Probabilistic Credulous Acceptance known from the literature. That is, for a semantics $\mathcal{S}$, we consider the problem:

PROBLEM :	$\mathsf{PrCA}[\mathcal{S}]$
INPUT :	A PrAF Δ and an argument g.
OUTPUT :	The number $PrCA_{\Delta}^{\mathcal{S}}(g)$.

Algorithm 2

Require: An AF $\Lambda = \langle A, \Sigma \rangle$ and a semantics $\mathcal{S}$.
Ensure: An $\mathcal{S}$-extension E.

```
Let E ∈ GR(Λ)
while E is not an S-extension of Λ do
    Let C be the union of initial SCCs of Λ̂;
    Choose a ∈ C with probability 1/|C|;
    E := E ∪ E' s.t. E' ∈ GR(Λ̂_a);
    Λ := Λ̂_a;
return E
```

It is known that $\mathsf{PrCA}[\mathcal{S}]$ is $FP^{\#P}$-hard (Fazzinga, Flesca, and Furfaro 2018), and this holds even for acyclic PrAFs. Moreover, with the same argument given for Theorem 2, we can also show that $\mathsf{PrCA}[\mathcal{S}]$ admits no FPRAS, for all semantics $\mathcal{S}$.

Theorem 6. *Consider a semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless NP $\subseteq$ BPP, there is no FPRAS for* $\mathsf{PrCA}[\mathcal{S}]$*, even for acyclic PrAFs.*

The main difference with $\mathsf{PrEA}[\mathcal{S}]$ lies in the approximability via FPARASes. In particular, we can show that $\mathsf{PrCA}[\mathcal{S}]$ is harder than $\mathsf{PrEA}[\mathcal{S}]$ in this regard, as no FPARAS exists, for all semantics $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{STT}\}$, even when PrAFs have no odd-length cycles.

For the proof of this theorem, we use the fact that checking whether a given argument g is credulously accepted over an AF Λ is NP-hard, when $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, even when AFs have no odd-length cycles (Simari and Rahwan 2009). With this in place, it is easy to apply a similar argument to the one given in the proof of Theorem 3.

Theorem 7. *Consider a semantics $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$. Unless $NP \subseteq BPP$, there is no FPARAS for* $\mathsf{PrCA}[\mathcal{S}]$*, even for PrAFs without odd-length cycles.*

The above result highlights an additional difficulty in approximating $\mathsf{PrCA}[\mathcal{S}]$ w.r.t. $\mathsf{PrEA}[\mathcal{S}]$, since Theorem 5 states that an FPARAS for $\mathsf{PrEA}[\mathcal{S}]$ exists for $\mathcal{S} \in \{\mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$, when we consider PrAFs without odd-length cycles.

The only positive result one can obtain for the problem $\mathsf{PrCA}[\mathcal{S}]$ is when $\mathcal{S} = \mathcal{GR}$. In this case, $\mathsf{PrEA}[\mathcal{S}] = \mathsf{PrCA}[\mathcal{S}]$, and thus the following is a corollary of Theorem 5.

Corollary 1. *The problem* $\mathsf{PrCA}[\mathcal{GR}]$ *admits an FPARAS.*

7 Experimental Analysis

We implemented a Python prototype of Algorithm 1 for approximating $\mathsf{PrEA}[\mathcal{S}]$, and tested it over the dataset used during the last ICCMA competition. The dataset consists of 326 AFs having a number of arguments in $[4,\ 10K]$ and a number of attacks in $[8,\ 1M]$. For each AF in the dataset, each odd-length cycle was broken by randomly removing an attack until an AF having no odd-length cycles was obtained. The resulting AF has been transformed into a PrAF by randomly assigning a probability to every argument. For each PrAF, we have chosen five distinct goal arguments, and executed our implementation with $\epsilon = \delta = 5\%$, for each semantics $\mathcal{S} \in \{\mathcal{GR}, \mathcal{PR}, \mathcal{ST}, \mathcal{SST}\}$.

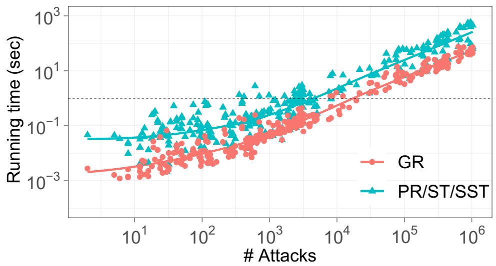

Figure 3: Running time of Apx versus the number of attacks.

Figure 3 reports the running time of our implementation versus the number of attacks of each PrAF, for the grounded semantics (red circles) and preferred/stable/semi-stable semantics (blue triangles), that coincide for AFs without odd-length cycles. Each data point refers to the mean of 25 runs. For the sake of readability, LOESS regression is also shown.

The experiments show that the running time depends almost linearly on the number of attacks of the input PrAF. Further experiments showed that the running time increases quadratically (resp., logarithmically) as ϵ (resp., δ) decreases, confirming the theoretical behavior of our algorithm, where ϵ and δ define the number of iterations.

Figure 3 also shows that the running time for the grounded semantics is lower than those for the other semantics. This is due to the fact that the execution of line 5 of Algorithm 1 is faster for $\mathcal{GR}$, as under $\mathcal{GR}$, only one extension exists, which has probability 1. However, the running time for the other semantics is not much higher than that for the grounded semantics: it is only 5.53 times that of the grounded semantics, on average. This is due to the fact that most of the PrAFs have a very large SCC containing 85% of the arguments on average, and thus the probabilistic trie of a world is not very deep. Overall, the experiments show that our approximation algorithm performs well on quite large PrAFs—for PrAFs having up to $10K$ attacks (almost 60% of PrAFs in the dataset), execution ends in less than 1 second (see the dotted line on Figure 3).

8 Conclusions and Future Work

We have explored the problem of Probability of Acceptance of a goal argument in probabilistic argumentation frameworks. Our approach steams from the fact that, in our view, Probabilistic Credulous Acceptance may not provide intuitive answers as it generalizes the classical credulous acceptance problem for AFs in one dimension only, that is, via probabilities over possible worlds. Our approach considers also another dimension, i.e, assigns probabilities also to the extensions of each possible world. As shown in our running example, this enables more intuitive answers, e.g. the probabilities of acceptance of mutually conflicting arguments (e.g. `fish` and `meat`) sum up to 1 (this is not the case for Probabilistic Credulous Acceptance). Thus, we introduced the problem $\mathsf{PrA}[\mathcal{S}]$, where a PDF is assumed over the set of extensions, and devised $\mathsf{PrEA}[\mathcal{S}]$ as a concrete instance, where the PDF leverages our notion of explanations for extensions.

Integrating explanations in argumentation systems is important for enhancing the argumentation and persuasion capabilities of software agents (Moulin et al. 2002; Bex and Walton 2016; Cyras et al. 2019; Miller 2019). For this reasons, several researchers explored how to deal with explanations in formal argumentation. Significant work in this field includes (Fan and Toni 2015), where a new argumentation semantics is proposed for capturing explanations in AF, and (Craven and Toni 2016) that focuses on ABA framework (Dung, Kowalski, and Toni 2009). They treat an explanation as a semantics to answer why an argument is accepted or not. Thus, an explanation is viewed as a set of arguments, instead of a sequence of arguments, needed for explaining such an extension. In (Fan and Toni 2015) an explanation is as a set of arguments justifying a given argument by means of a proponent-opponent dispute-tree (Dung, Mancarella, and Toni 2007). A similar approach based on debate trees as proof procedure for computing grounded, ideal, and preferred semantics, has been proposed in (Thang, Dung, and Hung 2009). However, in our perspective, explanations provide a tool to assign probabilities to extensions, and an explanation can be viewed as a sequence of choices to be made to justify how an extension is obtained.

Analogously to several other computational approaches in formal argumentation, our approach suffers from high computational complexity (Dunne and Wooldridge 2009; Dvorák and Woltran 2010; Dvorák et al. 2014; Kröll, Pichler, and Woltran 2017; Alfano, Greco, and Parisi 2017; Alfano, Greco, and Parisi 2018; Alfano et al. 2018; Alfano, Greco, and Parisi 2019a; Alfano, Greco, and Parisi 2019b; Alfano et al. 2020). However, after showing that $\mathsf{PrA}[\mathcal{S}]$ and $\mathsf{PrEA}[\mathcal{S}]$ are $FP^{\#P}$-hard, even for acyclic PrAFs, we investigated the existence of polynomial-time algorithms for $\mathsf{PrEA}[\mathcal{S}]$ in terms of approximate computation via FP(A)RASes. This is analogous to what done in (Fazzinga, Flesca, and Parisi 2016; Fazzinga, Flesca, and Parisi 2016), where Monte-Carlo techniques are proposed to estimate the probability that a set of arguments is an extension in probabilistic AFs and structured argumentation frameworks.

We also found that approximate computation via FPARASes is not possible for general PrAFs and for all the considered semantics, besides the grounded. Thus, we proposed an additive approximation algorithm for solving $\mathsf{PrEA}[\mathcal{S}]$ in all the cases where it exists: probabilistic AFs without odd-length cycles and any semantics $\mathcal{S}$, and for $\mathsf{PrEA}[\mathcal{GR}]$ in general PrAFs. Our results immediately apply to probabilistic frameworks with uncertain attacks, thanks to the results of (Mantadelis and Bistarelli 2020).

To the best of our knowledge, this is the first piece of work investigating probabilistic acceptance in combination with explanations for probabilistic AFs. As a first direction for future work, we plan to extend our notion of explanation, and investigate the counterparts of our problems, in the context of structured argumentation, such as p-ASPIC (Rienstra 2012), a probabilistic version of a general fragment of ASPIC (Prakken 2010). As a second direction, we plan to investigate other ways of defining a PDF over the set of extensions which enable other instantiations of $\mathsf{PrA}[\mathcal{S}]$, not necessary defined using explanations.

References

Alfano, G.; Greco, S.; Parisi, F.; Simari, G. I.; and Simari, G. R. 2018. An incremental approach to structured argumentation over dynamic knowledge bases. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 78–87.

Alfano, G.; Cohen, A.; Gottifredi, S.; Greco, S.; Parisi, F.; and Simari, G. 2020. Dynamics in abstract argumentation frameworks with recursive attack and support relations. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*.

Alfano, G.; Greco, S.; and Parisi, F. 2017. Efficient computation of extensions for dynamic abstract argumentation frameworks: An incremental approach. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 49–55.

Alfano, G.; Greco, S.; and Parisi, F. 2018. A meta-argumentation approach for the efficient computation of stable and preferred extensions in dynamic bipolar argumentation frameworks. *Intelligenza Artificiale* 12(2):193–211.

Alfano, G.; Greco, S.; and Parisi, F. 2019a. An efficient algorithm for skeptical preferred acceptance in dynamic argumentation frameworks. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 18–24.

Alfano, G.; Greco, S.; and Parisi, F. 2019b. On scaling the enumeration of the preferred extensions of abstract argumentation frameworks. In *Proceedings of ACM/SIGAPP Symposium on Applied Computing (SAC)*, 1147–1153.

Arora, S., and Barak, B. 2009. *Computational Complexity - A Modern Approach*. Cambridge University Press.

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

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

Baroni, P.; Giacomin, M.; and Guida, G. 2005. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence* 168(1-2):162–210.

Bench-Capon, T., and Dunne, P. E. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171:619 – 641.

Bex, F., and Walton, D. 2016. Combining explanation and argumentation in dialogue. *Argument & Computation* 7(1):55–68.

Cerutti, F.; Giacomin, M.; Vallati, M.; and Zanella, M. 2014. An SCC recursive meta-algorithm for computing preferred labellings in abstract argumentation. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

Craven, R., and Toni, F. 2016. Argument graphs and assumption-based argumentation. *Artificial Intelligence* 233:1–59.

Cyras, K.; Birch, D.; Guo, Y.; Toni, F.; Dulay, R.; Turvey, S.; Greenberg, D.; and Hapuarachchi, T. 2019. Explanations by arbitrated argumentative dispute. *Expert Systems with Applications* 127:141–156.

Doder, D., and Woltran, S. 2014. Probabilistic argumentation frameworks-A logical approach. In *Proceedings of International Conference on Scalable Uncertainty Management (SUM)*, 134–147.

Dung, P. M., and Thang, P. M. 2010. Towards (probabilistic) argumentation for jury-based dispute resolution. In *Proceeding of International Conference on Computational Models of Argument (COMMA)*, 171–182.

Dung, P. M.; Kowalski, R. A.; and Toni, F. 2009. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*. Springer. 199–218.

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

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77:321–358.

Dunne, P. E., and Bench-Capon, T. J. M. 2002. Coherence in finite argument systems. *Artificial Intelligence* 141(1/2):187–203.

Dunne, P. E., and Wooldridge, M. 2009. Complexity of abstract argumentation. In *Argumentation in Artificial Intelligence*. Springer. 85–104.

Dvorák, W., and Woltran, S. 2010. Complexity of semi-stable and stage semantics in argumentation frameworks. *Information Processing Letters* 110(11):425–430.

Dvorák, W.; Järvisalo, M.; Wallner, J. P.; and Woltran, S. 2014. Complexity-sensitive decision procedures for abstract argumentation. *Artificial Intelligence* 206:53–78.

Fan, X., and Toni, F. 2015. On computing explanations in argumentation. In *Proceedings of AAAI Conference on Artificial Intelligence*, 1496–1502.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2018. Credulous and skeptical acceptability in probabilistic abstract argumentation: complexity results. *Intelligenza Artificiale* 12(2):181–191.

Fazzinga, B.; Flesca, S.; and Furfaro, F. 2019. Complexity of fundamental problems in probabilistic abstract argumentation: Beyond independence. *Artificial Intelligence* 268:1–29.

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

Fazzinga, B.; Flesca, S.; and Parisi, F. 2016. On efficiently estimating the probability of extensions in abstract argumentation frameworks. *International Journal of Approximate Reasoning* 69:106–132.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58(301):13–30.

Hunter, A. 2012. Some foundations for probabilistic abstract argumentation. In *Proceeding of International Conference on Computational Models of Argument (COMMA)*, 117–128.

Hunter, A. 2013. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning* 54(1):47–81.

Ko, K. 1982. Some observations on the probabilistic algorithms and NP-hard problems. *Information Processing Letters* 14(1):39–43.

Kröll, M.; Pichler, R.; and Woltran, S. 2017. On the complexity of enumerating the extensions of abstract argumentation frameworks. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, 1145–1152.

Li, H.; Oren, N.; and Norman, T. J. 2011. Probabilistic argumentation frameworks. In *Proceeding of International Workshop on Theorie and Applications of Formal Argumentation (TAFA)*, 1–16.

Mantadelis, T., and Bistarelli, S. 2020. Probabilistic abstract argumentation frameworks, a possible world view. *International Journal of Approximate Reasoning* 119:204–219.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267:1–38.

Moulin, B.; Irandoust, H.; Bélanger, M.; and Desbordes, G. 2002. Explanation and argumentation capabilities: Towards the creation of more persuasive agents. *Artificial Intelligence Review* 17(3):169–222.

Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument & Computation* 1(2):93–124.

Rienstra, T.; Thimm, M.; Liao, B.; and van der Torre, L. W. N. 2018. Probabilistic abstract argumentation based on SCC decomposability. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 168–177.

Rienstra, T. 2012. Towards a probabilistic dung-style argumentation system. In *Proceedings of International Conference on Agreement Technologies (AT)*, 138–152.

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

Thang, P. M.; Dung, P. M.; and Hung, N. D. 2009. Towards a common framework for dialectical proof procedures in abstract argumentation. *Journal of Logic and Computation* 19(6):1071–1109.

Thimm, M. 2012. A probabilistic semantics for abstract argumentation. In *Proceedings of European Conference on Artificial Intelligence (ECAI)*, 750–755.

Welsh, F., and Gale, A. 2001. The complexity of counting problems. In *Aspects of complexity*.