Argument & Computation 0 (0) 1
IOS Press

An Experimental Analysis on the Similarity of Argumentation Semantics

Federico Cerutti [a], Matthias Thimm [b,*] and Mauro Vallati [c]
[a] *Department of Information Engineering, University of Brescia, Italy*
E-mail: federico.cerutti@unibs.it
[b] *Institute for Web Science and Technologies, University of Koblenz-Landau, Germany*
E-mail: thimm@uni-koblenz.de
[c] *School of Computing and Engineering, University of Huddersfield, United Kingdom*
E-mail: m.vallati@hud.ac.uk

Abstract. In this paper we ask whether approximation for abstract argumentation is useful *in practice*, and in particular whether *reasoning with grounded semantics*—which has polynomial runtime—is already an approximation approach sufficient for several practical purposes. While it is clear from theoretical results that reasoning with grounded semantics is different from, for example, skeptical reasoning with preferred semantics, we investigate how significant this difference is in actual argumentation frameworks. As it turns out, in many graphs models, reasoning with grounded semantics actually approximates reasoning with other semantics almost perfectly. An algorithm for grounded reasoning is thus a conceptually simple approximation algorithm that not only does not need a learning phase—like recent approaches—but also approximates well—in practice—several decision problems associated to other semantics.

Keywords: Approximate algorithm, Experimental analysis, Jaccard's distance

1. Introduction

Dung's theory of abstract argumentation [1] unifies a large variety of formalisms in nonmonotonic reasoning, logic programming and computational argumentation. It is based on the notion of an argumentation framework (AF) that consists of a set of arguments and of an *attack* relation between them. Different *argumentation semantics* introduce the criteria to determine which arguments emerge as *justified* from the conflicts, by identifying a number of *extensions*, i. e. sets of arguments that can *survive the conflicts together*. In [1] three *traditional* semantics are introduced, namely *grounded*, *stable*, and *preferred* semantics. Other literature proposals include *semi-stable* [2, 9] and *ideal* semantics [3]. For an introduction on the various semantics, see [4]. Several problems are associated to each semantics, notably *credulous* and *skeptical* acceptance of an argument with respect to a given argumentation framework—i. e. determining whether an argument belongs to at least one (resp. every) extension—and *enumeration* of *all* or *some* extensions given an argumentation framework. Among those semantics, grounded semantics prescribes a unique extension which can be computed in polynomial time, thus all the problems related to grounded semantics are easy to solve. Instead, decision problems associated to the other semantics are much more complex, with some at the second level of the polynomial hierarchy (see also Section 2).

*Corresponding author. E-mail: thimm@uni-koblenz.de.

To our knowledge, only a few work addressed the problem of approximating the solution of some decision or enumeration problems associated to an argumentation semantics. In [5] predictive models have been positively exploited in abstract argumentation for predicting significant aspects, such as the number, of the solution to the preferred extensions enumeration problem, where the complete knowledge of such structure would require a computationally hard problem to be solved. In [6], an approximation algorithm for credulous reasoning with preferred/complete semantics is presented. That algorithm is based on learning a graph convolutional neural network [7] from a set of correctly solved benchmark instances and then using the learned network as an approximation algorithm. The advantage is that runtime drastically decreases (basically to linear runtime, given the learned network) while classification accuracy is still at a reasonable 80 % or more in certain cases, i. e. 80 % of all arguments of a certain input argumentation framework were correctly classified as credulously accepted or not. The work [6] thus showed that it is generally feasible to employ this methodology for developing approximation algorithms for hard problems in abstract argumentation. Using more recent approaches from the deep learning community and increasing efforts in streamlining this methodology will probably increase classification accuracy further.

It has to be noted that the methodology illustrated in [6] is conceptually complex, requiring sophisticated learning algorithms and complex deep learning models, and needs additional time for the learning phase. In the present paper, we ask the question whether such a complexity is necessary *in practice*. More concretely, we ask the question whether *reasoning with grounded semantics*, which has polynomial runtime, is not already a sufficient approximation approach. While it is clear from theoretical results (Section 2) that reasoning with grounded semantics is different from, for example, skeptical reasoning with preferred semantics, we wish to investigate how significant this difference is in actual argumentation frameworks (Section 3). As it turns out, in many graphs models (Section 4) reasoning with grounded semantics actually approximates reasoning with other semantics almost perfectly (Section 5). An algorithm for grounded reasoning is thus a conceptually simple approximation algorithm that does not need an expensive learning phase but turns out to have high classification accuracy as well.

2. Background

An argumentation framework [1] consists of a set of arguments[1] and a binary attack relation between them.

Definition 1. An *argumentation framework* (*AF*) is a pair $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A}$ is a set of arguments and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. We say that $\mathbf{b}$ *attacks* $\mathbf{a}$ iff $(\mathbf{b}, \mathbf{a}) \in \mathcal{R}$, also denoted as $\mathbf{b} \rightarrow \mathbf{a}$. The set of attackers of an argument $\mathbf{a}$ is denoted as $\mathbf{a}^- \triangleq \{\mathbf{b} : \mathbf{b} \rightarrow \mathbf{a}\}$, the set of arguments attacked by $\mathbf{a}$ is denoted as $\mathbf{a}^+ \triangleq \{\mathbf{b} : \mathbf{a} \rightarrow \mathbf{b}\}$. An argument $\mathbf{a}$ without attackers, i. e. such that $\mathbf{a}^- = \emptyset$, is said *initial*. We also extend attack notations to sets of arguments, i. e. given $E, S \subseteq \mathcal{A}$, $E \rightarrow \mathbf{a}$ iff $\exists \mathbf{b} \in E$ s.t. $\mathbf{b} \rightarrow \mathbf{a}$; $\mathbf{a} \rightarrow E$ iff $\exists \mathbf{b} \in E$ s.t. $\mathbf{a} \rightarrow \mathbf{b}$; $E \rightarrow S$ iff $\exists \mathbf{b} \in E, \mathbf{a} \in S$ s.t. $\mathbf{b} \rightarrow \mathbf{a}$; $E^- \triangleq \{\mathbf{b} \mid \exists \mathbf{a} \in E, \mathbf{b} \rightarrow \mathbf{a}\}$ and $E^+ \triangleq \{\mathbf{b} \mid \exists \mathbf{a} \in E, \mathbf{a} \rightarrow \mathbf{b}\}$.

Each argumentation framework has an associated directed graph where the vertices are the arguments, and the edges are the attacks.

The basic properties of conflict–freeness, acceptability, and admissibility of a set of arguments are fundamental for the definition of argumentation semantics.

[1] In this paper we consider only *finite* sets of arguments: see [8] for a discussion on infinite sets of arguments.

Definition 2. Given an AF $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$:

- a set $S \subseteq \mathcal{A}$ is a *conflict–free* set of Γ if $\nexists$ $\mathbf{a}, \mathbf{b} \in S$ s.t. $\mathbf{a} \to \mathbf{b}$;
- an argument $\mathbf{a} \in \mathcal{A}$ is *acceptable* in Γ with respect to a set $S \subseteq \mathcal{A}$ if $\forall \mathbf{b} \in \mathcal{A}$ s.t. $\mathbf{b} \to \mathbf{a}$, $\exists$ $\mathbf{c} \in S$ s.t. $\mathbf{c} \to \mathbf{b}$;
- the function $\mathcal{F}_\Gamma : 2^{\mathcal{A}} \to 2^{\mathcal{A}}$ such that $\mathcal{F}_\Gamma(S) = \{\mathbf{a} \mid \mathbf{a} \text{ is acceptable in } \Gamma \text{ w.r.t. } S\}$ is called the *characteristic function* of Γ;
- a set $S \subseteq \mathcal{A}$ is an *admissible* set of Γ if S is a conflict–free set of Γ and every element of S is acceptable in Γ with respect to S, i. e. $S \subseteq \mathcal{F}_\Gamma(S)$.

An argumentation semantics σ prescribes for any AF Γ a set of *extensions*, denoted as $\mathcal{E}_\sigma(\Gamma)$, namely a set of sets of arguments satisfying the conditions dictated by σ. The paper is focused on grounded (denoted as GR), stable (ST), preferred (PR) semantics, introduced in [1]; as well as on semi–stable (SST), originally introduced with the name of *admissible argumentation stage extension* in [9] and then re-named in [2, 10]; and on ideal (ID), originally introduced in [3].

Definition 3. Given an AF $\Gamma = \langle \mathcal{A}, \mathcal{R} \rangle$:

- a set $S \subseteq \mathcal{A}$ is the *grounded extension* of Γ i. e. $\{S\} = \mathcal{E}_{\mathsf{GR}}(\Gamma)$, iff S is the minimal (w.r.t. set inclusion) fixed point of $\mathcal{F}_\Gamma$;
- a set $S \subseteq \mathcal{A}$ is a *stable extension* of Γ, i. e. $S \in \mathcal{E}_{\mathsf{ST}}(\Gamma)$, iff S is a conflict-free set of Γ and $S \cup S^+ = \mathcal{A}$;
- a set $S \subseteq \mathcal{A}$ is a *preferred extension* of Γ, i. e. $S \in \mathcal{E}_{\mathsf{PR}}(\Gamma)$, iff S is a maximal (w.r.t. set inclusion) admissible set of Γ;
- a set $S \subseteq \mathcal{A}$ is a *semi–stable extension* of Γ, i. e. $S \in \mathcal{E}_{\mathsf{SST}}(\Gamma)$, iff S is a preferred extension where $S \cup S^+$ is maximal (w.r.t. set inclusion) among all preferred extensions;
- a set $S \subseteq \mathcal{A}$ is the *ideal extension* of Γ, i. e. $\{S\} = \mathcal{E}_{\mathsf{ID}}(\Gamma)$, iff S is the maximal (w.r.t. set inclusion) admissible set of Γ that is also subset of each preferred extension.

An argument $\mathbf{a}$ is *credulously* (resp. *skeptically*) accepted with regard to a given semantics σ and a given AF Γ iff $\mathbf{a}$ belongs to at least one (resp. each) extension of Γ under σ. Let denote with σ_Γ-C (resp. σ_Γ-S) the set of all the credulously (resp. skeptically) accepted arguments according to σ, i. e., if $\exists S \in \mathcal{E}_\sigma(\Gamma)$, $\mathbf{a} \in S$, then $\mathbf{a} \in \sigma_\Gamma$-C; and if $\forall S \in \mathcal{E}_\sigma(\Gamma)$ $\mathbf{a} \in S$, then $\mathbf{a} \in \sigma_\Gamma$-S. Note that in the case no stable extension exists, ST_Γ-S $= \mathcal{A}$.

With a slight abuse of notation, and begging the reader for forgiveness, we will write $\mathsf{GR}_\Gamma = \mathsf{GR}_\Gamma$-C $=$ GR_Γ-S and $\mathsf{ID}_\Gamma = \mathsf{ID}_\Gamma$-C $= \mathsf{ID}_\Gamma$-S as both grounded and ideal are unique. Also, when it applies to generic Dung's argumentation framework, or when it is clear from the context, we will also drop the reference to a given AF, hence for instance we will write PR-C to refer to PR_Γ-C where Γ is a generic, unspecified Dung's argumentation framework, or the specific Dung's argumentation framework we are discussing in a specific portion of text.

In [11] the notion of *skepticism* has been formally investigated. An extension E_1 is at least as skeptical as an extension E_2 if $E_1 \subseteq E_2$, since then E_1 supports the acceptance of no more arguments than E_2.

Definition 4. Given two extensions E_1 and E_2 of an argumentation framework Γ, E_1 is *at least as skeptical* as E_2, denoted as $E_1 \preceq E_2$ if and only if $E_1 \subseteq E_2$.

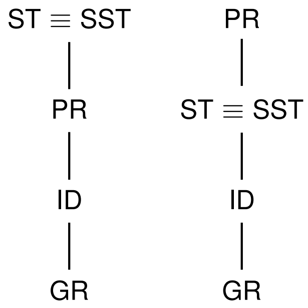


Fig. 1. $\preceq_\cap$ (left) and $\preceq_\cup$ (right) relation for frameworks where the stable extensions exist.

Table 1

Complexity of traditional decision problem on Dung's abstract argumentation.

σ	$\mathbf{a} \overset{?}{\in} \sigma$-C	$\mathbf{a} \overset{?}{\in} \sigma$-S
GR	P-complete	P-complete
ST	NP-complete	coNP-complete
PR	NP-complete	Π_2^{P}-complete
SST	Σ_2^{P}-complete	Π_2^{P}-complete
ID	Θ_2^{P}-complete	Θ_2^{P}-complete

This notion suffices in the case of grounded and ideal semantics as they are unique. In [11] the authors introduced also the following two relations between non-empty sets of extensions[2] based to skeptical and credulous acceptance.

Definition 5. Given two non-empty sets of extensions $\mathcal{E}_1$ and $\mathcal{E}_2$ of an argumentation framework Γ, $\mathcal{E}_1 \preceq_\cap \mathcal{E}_2$ if and only if $\bigcap_{E_1 \in \mathcal{E}_1} E_1 \subseteq \bigcap_{E_2 \in \mathcal{E}_2} E_2$.

Definition 6. Given two non-empty sets of extensions $\mathcal{E}_1$ and $\mathcal{E}_2$ of an argumentation framework Γ, $\mathcal{E}_1 \preceq_\cup \mathcal{E}_2$ if and only if $\bigcup_{E_1 \in \mathcal{E}_1} E_1 \subseteq \bigcup_{E_2 \in \mathcal{E}_2} E_2$.

Figure 1 summarises the skeptical relationships that exists between different semantics extensions [11].

3. Measuring relative skepticism

If we have a look at the computational complexity of decision problems associated to Dung's argumentation framework—see [12] for an extensive analysis—we can see (cf. Table 1) that many decision problems cannot be solved in deterministic polynomial time, except for the case of grounded semantics.

We already know [12] that in the case of acyclic and even-cycle free AFs all the semantics we consider in this paper are equivalent to the grounded. For the other cases, building on top of Figure 1, we can

[2] An interested reader is referred to [11] to appreciate the differences with possibly empty sets of extensions.

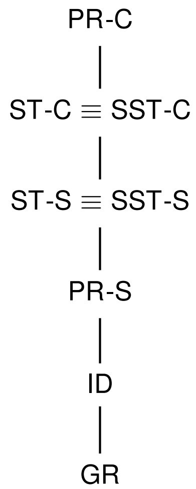


Fig. 2. Hasse diagram of the relationship between sets of credulously and skeptically accepted arguments w.r.t. GR, ST, PR, SST, and ID for argumentation frameworks admitting at least one stable extension.

easily derive a $\subseteq$ ordering between sets of credulously and skeptically accepted arguments according to the semantics we consider in this paper.

Proposition 1. *Given an argumentation framework* Γ *for which* $\mathcal{E}_{\mathsf{ST}}(\Gamma) \neq \emptyset$

$$\mathsf{GR}_\Gamma \subseteq \mathsf{ID}_\Gamma \subseteq \mathsf{PR}_\Gamma\text{-}\mathsf{S} \subseteq \mathsf{ST}_\Gamma\text{-}\mathsf{S} \equiv \mathsf{SST}_\Gamma\text{-}\mathsf{S} \subseteq \mathsf{ST}_\Gamma\text{-}\mathsf{C} \equiv \mathsf{SST}_\Gamma\text{-}\mathsf{C} \subseteq \mathsf{PR}_\Gamma\text{-}\mathsf{C}$$

Figure 2 illustrates the result of Proposition 1. We now need to be able to quantify the *distance* between such sets, so to have an indication of how much the grounded extension covers the other sets. Clearly, this can be provided only experimentally, and for this reason we rely on the statistic provided by the *Jaccard's index* [13] that quantifies the similarities between sets. It is defined as the size of the intersection divided by the size of the union of the sample sets.

Definition 7 (Jaccard's Index and Distance, derived from [13])**.** Given two sets A and B, their *Jaccard's Similarity Coefficient* is:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Their *Jaccard's distance* is then:

$$J_\delta(A, B) = 1 - J(A, B)$$

Therefore, the set of all the sets of credulously and skeptically accepted arguments for an AF form a *metric space*, independently of the existence of stable extensions.

Proposition 2. *Given an AF* Γ*,* $\langle\{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}, J_\delta\rangle$ *is a* metric space.

Proof. It follows from results in [14]. □

From Figures 1 and 2 one can say that grounded is the *most skeptical* semantics possible (among those considered). We can then define the *measure of relative (to* GR*) skepticism* of a set of credulously or skeptically accepted arguments according to a semantics as the Jaccard's distance from the grounded extension.

Definition 8 (Measure of relative skepticism)**.** Given an *AF* Γ and given $\widetilde{\sigma}_\Gamma \in \{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}$, its *measure of relative (to* GR*) skepticism* μ_S is defined as:

$$\mu_S(\widetilde{\sigma}_\Gamma) = J_\delta(\widetilde{\sigma}_\Gamma, \mathsf{GR}_\Gamma)$$

The following propositions show properties of this measure: proofs are omitted as straightforward. In particular, this measure is a function whose range is the set of real numbers between 0 and 1 (both included).

Proposition 3. *Given an AF* Γ *and given* $\widetilde{\sigma}_\Gamma \in \{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}$, $\mu_S(\widetilde{\sigma}_\Gamma) = [0, 1]$.

Also, the measure of relative skepticism has a global minimum point in correspondence of the grounded semantics.

Proposition 4. *Given an AF* Γ *,* $\mu_S(\mathsf{GR}) \leqslant \mu_S(\widetilde{\sigma}_\Gamma)$*,* $\forall \widetilde{\sigma}_\Gamma \in \{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}$.

However, such a global minimum point, in general, is not unique as the grounded extension might coincide with some other set of credulously or skeptically accepted arguments $\widetilde{\sigma}_\Gamma$.

Proposition 5. *Given an AF* Γ*,* $\forall \widetilde{\sigma}_\Gamma \in \{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}$*, if* $\mu_S(\widetilde{\sigma}_\Gamma) = 0$ *then* $\widetilde{\sigma}_\Gamma = \mathsf{GR}_\Gamma$.

Finally, it follows that in the case of acyclic and even-cycle free *AF*s, each set of credulously or skeptically accepted arguments has zero as measure of relative skepticism, consistently with the results provided in [12].

Proposition 6. *Given an acyclic or even-cycle free AF* Γ*,* $\forall \widetilde{\sigma}_\Gamma \in \{\mathsf{GR}_\Gamma, \mathsf{ST}_\Gamma\text{-}\mathsf{C}, \mathsf{ST}_\Gamma\text{-}\mathsf{S}, \mathsf{PR}_\Gamma\text{-}\mathsf{C}, \mathsf{PR}_\Gamma\text{-}\mathsf{S}, \mathsf{SST}_\Gamma\text{-}\mathsf{C}, \mathsf{SST}_\Gamma\text{-}\mathsf{S}, \mathsf{ID}_\Gamma\}$ $\mu_S(\widetilde{\sigma}_\Gamma) = 0$.

4. Benchmarks

To experimentally analyse the measure of relative skepticism (Definition 8), we considered a significantly large experimental setting with a great variety of benchmarks, so to cover the vast majority of benchmarks currently used in abstract argumentation.

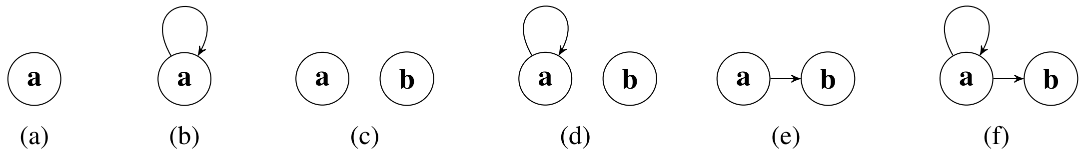


Fig. 3. The first six argumentation frameworks with increment size generated using the Tweety Generator.

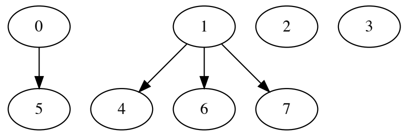


Fig. 4. One of the smallest examples of ABA-derived Dung's AFs in the ICCMA 2017 benchmarks.

4.1. A Million AFs

We exhaustively generated the first million of AFs with increasing size using the TweetyProject Generator.[3] In the following, we collectively refer to this group of AFs as aMillion. Figure 3 illustrates the first six of them.

4.2. Structured AFs

We also considered the 426 Assumption-Based argumentation frameworks translated to Dung's argumentation framework that have been submitted to the ICCMA 2017.[4] This benchmark restricted the ABA benchmark provided in [15] to those with at most 1,500 arguments. In the following, we collectively refer to this group of AFs as sABA. Figure 4 depicts one of the smallest.

We considered 300 ASPIC-like instances generated using TweetyProject[5]. In the following, we collectively refer to this group of AFs as sASPIC. Listing (1) in Appendix A lists an example of ASPIC-like theories used for generating the Dung's AF depicted in Figure 5; Appendix A also contains the necessary background on ASPIC [16].

4.3. Random AFs with Palatable Argumentative Characteristics

We randomly generated, using the `probo` [17] generator provided by the organiser of ICCMA 2015,[6] 2,400 AFs with a controllable number structure in terms of strongly connected components.[7] In the following, we collectively refer to this group of AFs as rSCC. Figure 6 depicts one of the smallest.

[3]http://tweetyproject.org/api/1.10/net/sf/tweety/arg/dung/util/EnumeratingDungTheoryGenerator.html
[4]http://argumentationcompetition.org/2017/ABA2AF.pdf
[5]http://tweetyproject.org/api/1.12/net/sf/tweety/arg/aspic/util/RandomAspicArgumentationTheoryGenerator.html
[6]http://argumentationcompetition.org/2015/results.html
[7]Parameters used: `arguments=20,40,...,200`, `numSccs=#arg/5,#arg/10,#arg/20`, `innerprob=0.6,0.8`, `outerprob=0.05,0.1`.

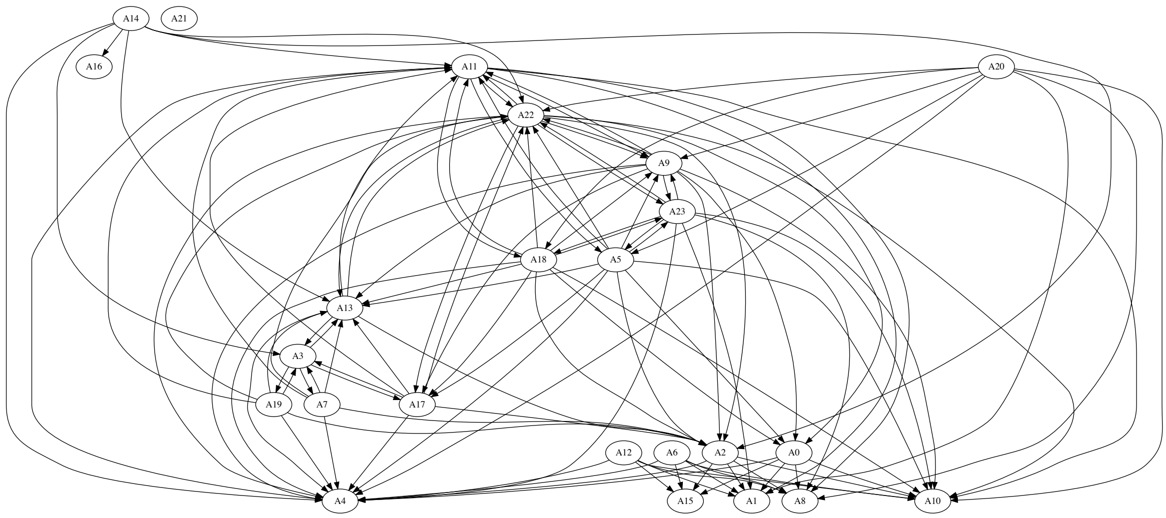


Fig. 5. Example of a Dung's *AF* generated from an ASPIC-like theory.

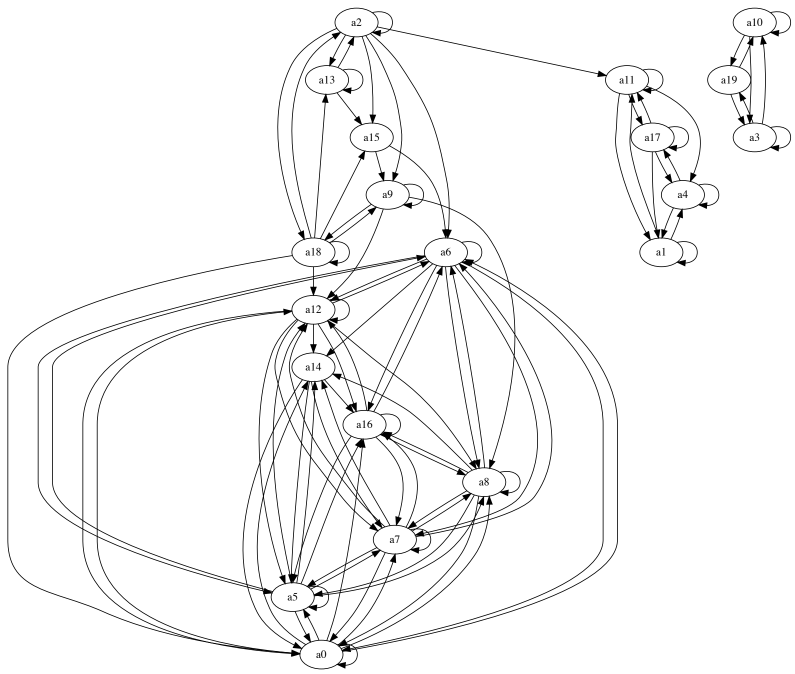


Fig. 6. One of the smallest example of Dung's *AF*s generated controlling parameters related to strongly connected components.

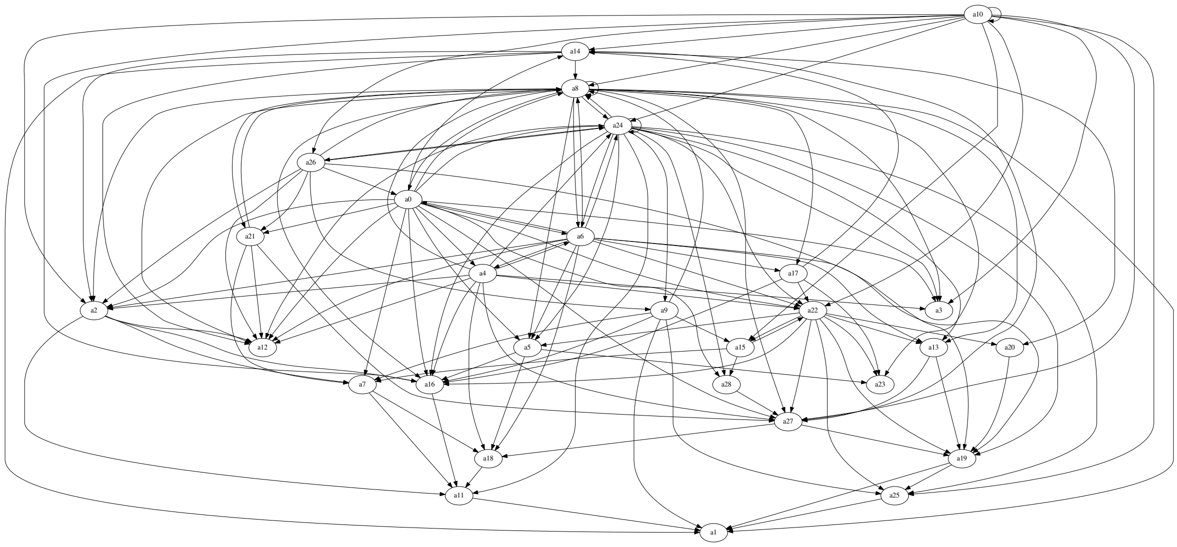


Fig. 7. An example of *AF*s featuring a large number of stable extensions.

We also randomly generated 200 *AF*s featuring a large number of stable extensions—and clearly of preferred extensions as well.[8] In the following, we collectively refer to this group of *AF*s as rStable. Figure 7 depicts an example of such *AF*s.

4.4. Random Graphs as AFs

Finally, we consider random graphs generators proposed in literature as a way to generate random *AF*s as well.

We generated 599 *AF*s according to the Erdös-Rényi [18] model—with edges between arguments randomly selected according to a uniform distribution—varying the number of arguments between 20 and 200, with an increment of 20, and with probability of attacks fixed as $\{0.01, 0.05, 0.1\}$.[9] In the following, we collectively refer to this group of *AF*s as rER. Figure 8 depicts an example of Erdös-Rényi-like *AF*s.

We also generated 360 *AF*s according to the Barabasi-Albert [20] model, varying the number of arguments between 20 and 200 with an increment of 20; and enforcing the probability to have at least one argument belonging to a cycle in the range $\{0.1, 0.2, 0.3\}$.[10] The Barabasi-Albert model enforces the common property of many large networks that the node connectivities follow a scale-free power-law distribution. This is generally the case when: (i) networks expand continuously by the addition of new nodes, and (ii) new nodes attach preferentially to sites that are already well connected. In the following, we collectively refer to this group of *AF*s as rBA. Figure 9 depicts an example Barabasi-Albert-like *AF*s.

---

[8] Parameters used: `arguments=20,40,...,200`, `minNum=#arg/20`, `maxNum=#arg/2`, `minSize=#arg/10`, `maxSize=#arg/2`, `minGround=0`, `maxGround=#arg/10`.

[9] AFBenchGen2 [19] parameters `numargs=20,40,...,200`, and `ER_probAttacks=0.01,0.05,0.1`.

[10] AFBenchGen2 [19] parameters `numargs=20,40,...,200`, and `BA_WS_probCycles=0.1,0.2,0.3`.

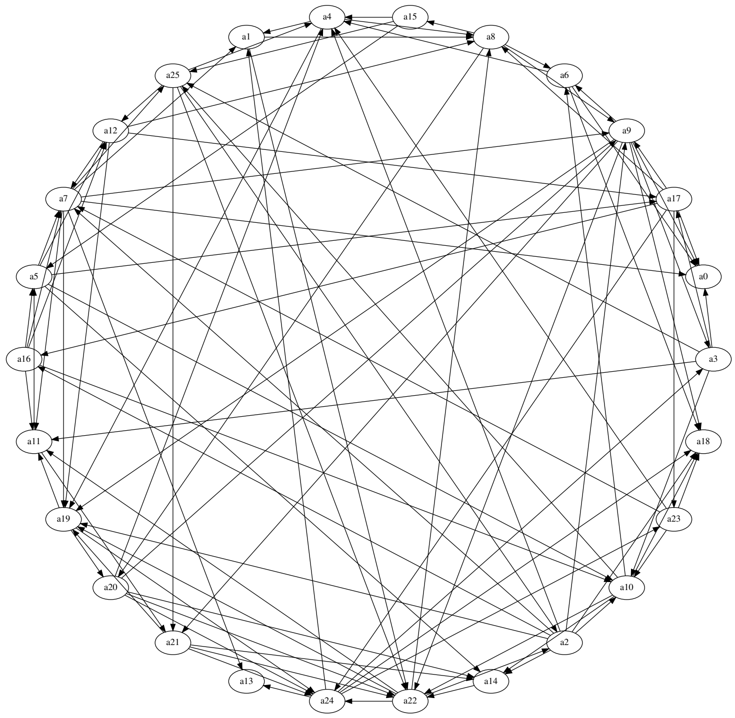


Fig. 8. An example of Erdös-Rényi-like *AF*s.

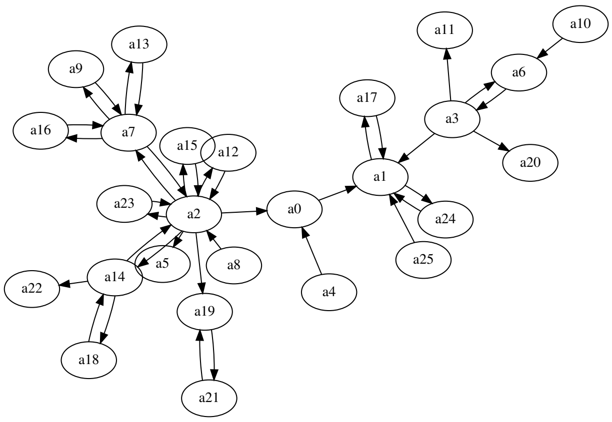


Fig. 9. An example of Barabasi-Albert-like *AF*s.

Finally, we considered the Watts-Strogatz [21] model, where a ring of n arguments where each argument is connected to its k nearest neighbors in the ring. k must satisfy $n > k > \log(n) > 1$ to ensure a connected graph. Also, each edge is randomly rewired with a probability β. Indeed, Watts and Strogatz [21] show that many biological, technological and social networks are neither completely regular nor completely random, but something in the between. They thus explore simple models of networks that can be tuned through this middle ground: regular networks *rewired* to introduce increasing amounts of disorder. These systems can be highly clustered, like regular lattices, yet have small characteristic path

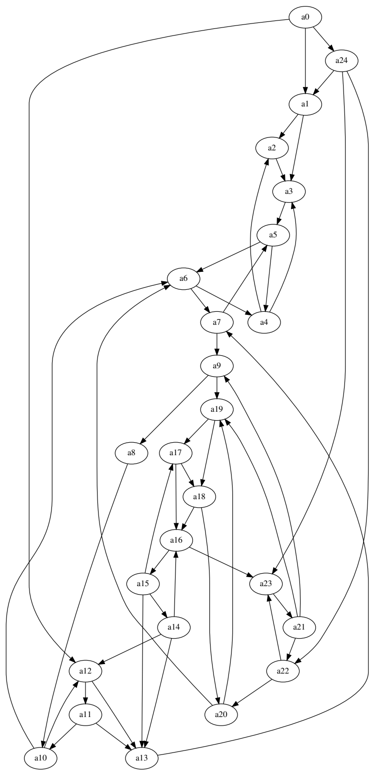


Fig. 10. An example of Watts-Strogatz-like AFs.

lengths, like random graphs, and they are named *small-world* networks by analogy with the small-world phenomenon. We generated 1,800 AFs according to the Watts-Strogatz model varying varying the number of arguments between 20 and 200 with an increment of 20; enforcing the probability to have at least one argument belonging to a cycle in the range $\{0.1, 0.2, 0.3\}$; setting k equal to half of the number of arguments; and varying β in $\{0.2, 0.4, 0.6\}$.[11] In the following, we collectively refer to this group of AFs as $\boxed{\mathsf{rWS}}$. Figure 10 depicts an example of Watts-Strogatz-like AFs.

5. Empirical analysis

To inform useful consideration, let us introduce the averaged measure of relative skepticism over a set of AFs.

Definition 9. Let $\mathcal{B}$ be a set of AFs, given a semantic $\sigma \in \{\mathsf{GR}, \mathsf{ST}, \mathsf{PR}, \mathsf{SST}, \mathsf{ID}\}$, $\mu_S^{\mathcal{B}}$ is the averaged measure of relative skepticism w.r.t. $\mathcal{B}$ defined as follows:
$$\mu_S^{\mathcal{B}}(\sigma\text{-}\mathsf{C}) = \frac{\sum_{\Gamma \in \mathcal{B}} \mu_S(\sigma_\Gamma\text{-}\mathsf{C})}{|\mathcal{B}|}$$
and

[11] AFBenchGen2 [19] parameters `numargs=20,40,...,200`, `BA_WS_probCycles=0.1,0.2,0.3`, `WS_baseDegree=(#arg/2)`, and `WS_beta=0.2,0.4,0.6`.

Fig. 11. Measure of relative skepticism μ_S of sets of credulously or skeptically accepted arguments according to the semantics discussed in Definition 3 and over the benchmarks described in Section 4. GR is omitted as $\mu_S(\mathsf{GR}_\Gamma) = 0$ for any AF Γ.

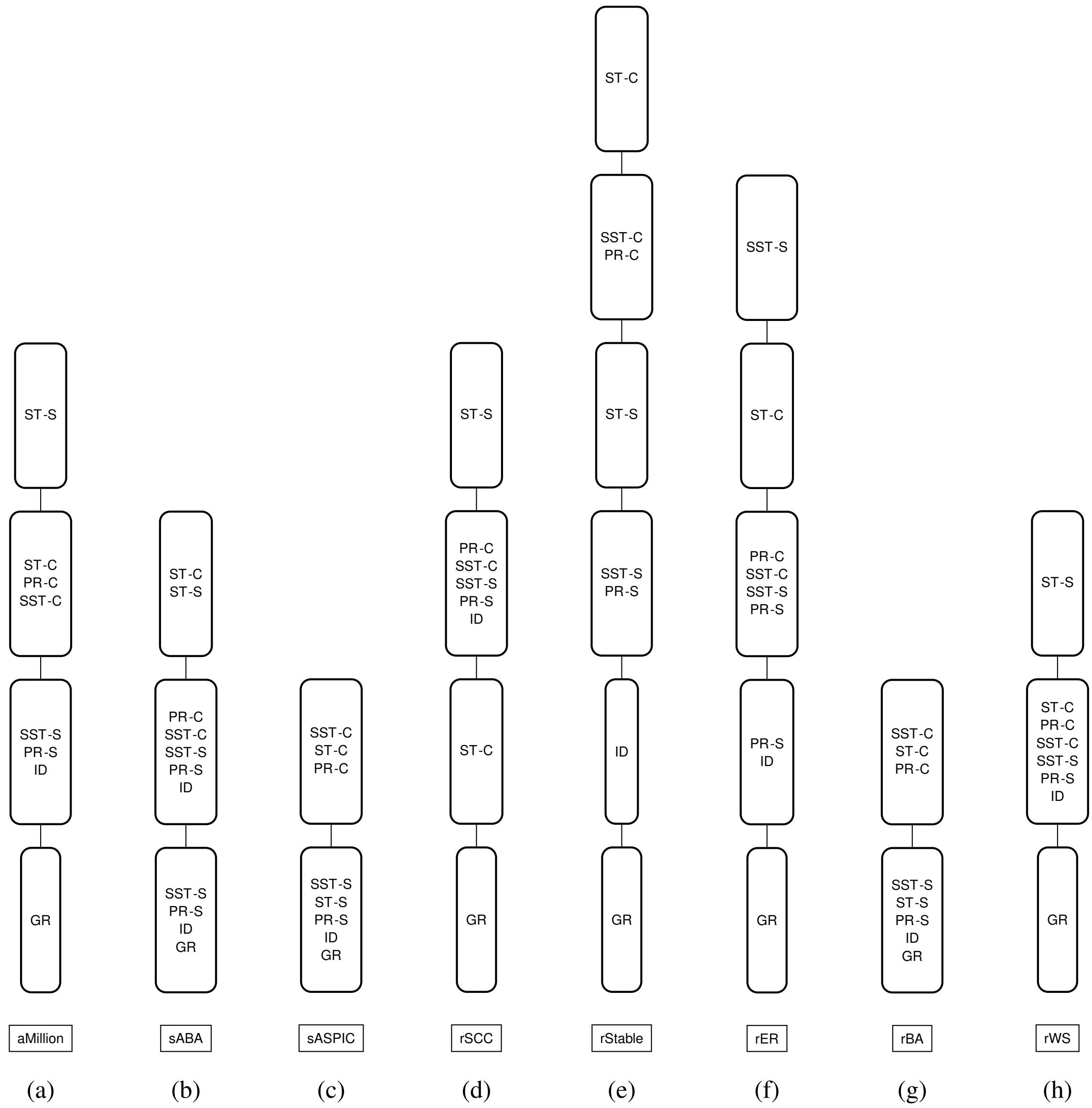


Fig. 12. 0.05-clusters of averaged measures of relative skepticism μ_S of sets of credulously or skeptically accepted arguments according to the semantics discussed in Definition 3 and over the benchmarks described in Section 4.

$$\mu_S^{\mathcal{B}}(\sigma\text{-S}) = \frac{\sum_{\Gamma \in \mathcal{B}} \mu_S(\sigma_\Gamma\text{-S})}{|\mathcal{B}|}$$

Figure 11 graphically summarises the averages of the measure of relative skepticism μ_S over the benchmarks we introduced in Section 4: full results for each of the benchmark groups are provided in Appendix B. First of all, it is worth mentioning that—except for the case of ST-S—the $\subseteq$ ordering discussed in Proposition 1 is naturally maintained. Also, the measure of relative skepticism is bounded between 0 and 1 (cf. Proposition 3): since $\mu_S(\mathsf{GR}_\Gamma) = 0$ for any AF Γ (cf. Proposition 4), we omit it from Figure 11.

As we can see in Figure 11, there are case where the averaged measures of relative skepticism appear to cluster closely together. To investigate this further, let us introduce the concept of a ε-cluster, as the set of credulously or skeptically accepted arguments with averaged measure of skepticism all within a chosen ε.

Definition 10. Let $\mathcal{B}$ be a set of AFs, and $\widetilde{\sigma}_1^{\mathcal{B}}, \widetilde{\sigma}_2^{\mathcal{B}} \in \{\mathsf{GR}, \mathsf{ST\text{-}C}, \mathsf{ST\text{-}S}, \mathsf{PR\text{-}C}, \mathsf{PR\text{-}S}, \mathsf{SST\text{-}C}, \mathsf{SST\text{-}S}, \mathsf{ID}\}$, we say that $\widetilde{\sigma}_1^{\mathcal{B}}$ and $\widetilde{\sigma}_2^{\mathcal{B}}$ belong to the same ε-cluster $\mathcal{C}^{\varepsilon}$ with $\varepsilon \geqslant 0$ if $|\mu_S^{\mathcal{B}}\left(\widetilde{\sigma}_1^{\mathcal{B}}\right) - \mu_S^{\mathcal{B}}\left(\widetilde{\sigma}_2^{\mathcal{B}}\right)| \leqslant \varepsilon$.

Figure 12 provides a qualitative interpretation of Figure 11 in terms of 0.05-clusters of sets of credulously or skeptically accepted arguments.

From Figures 11 and 12, we can observe the following:

(1) GR on average coincides with the skeptically accepted arguments of any semantics for sASPIC, and rBA. In the same sets it appears that all the credulously accepted set of arguments have the same averaged measure of relative skepticism;
(2) GR on average is a remarkably good estimator of PR-S, SST-S, and ID for sABA, as they all belong to the same 0.05-cluster;
(3) GR on average is a good estimator of all the sets of credulously and skeptically accepted arguments w.r.t. any semantics except for ST-S for rWS;
(4) ST-S is furthest apart from GR when considering rSCC and rWS;
(5) PR-C seems consistently to have the same measure of skepticism of SST-C, except in rER;
(6) GR on average seems to be a bad predictor for any set of credulously or skeptically arguments for any semantics in the case of rSCC, rStable, and rER.

Figure 13 depicts the distribution of the averages of Jaccard's distances across all the dataset comparing all the combinations of sets of credulously or skeptically accepted arguments. We can thus see:

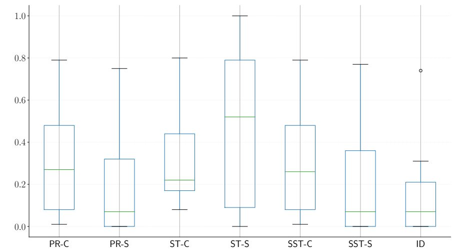
(1) overall GR is a good aproximator for ID and, yet less, for PR-S in our dataset;
(2) PR-S almost always coincide with ID and SST-S in our dataset;
(3) PR-C almost always coincide with SST-C in our dataset.

We then question whether there are some specific characteristics that substantially impact the measure of relative skepticism. To this end, following traditional machine learning approaches, we note that information about the structure of an AF can be extracted under the form of *features*. Each feature summarises a potentially important property of the considered framework, and the whole set of features can be seen as the fingerprint of the AF at hand.

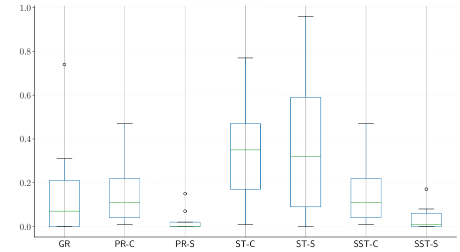
Consistently with other approaches exploiting predictive models [5, 22–26], we extracted a large set of features from each of the AFs. In total, the feature set includes 147 values, which exploit the representation of AFs as a graph or as a matrix.[12]

Given such a large set of features, we performed linear regression on the whole set of features, then feature selection and linear regression again. This double linear regression step helps us checking if the subset of features is actually informative. Feature selection has been performed using CorrelationAttribute and Ranker in Weka [27]. In a nutshell, this identifies features that are relevant with regards to the prediction task. The use of linear regression can also help in gaining an understanding of the relevance of features, on the basis of the assigned weight. The double linear regression step is performed to help us checking if the subset of selected features is actually informative.

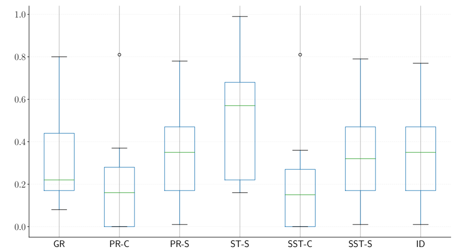
[12]The interested reader is referred to [5] for a detailed description of the extracted features.



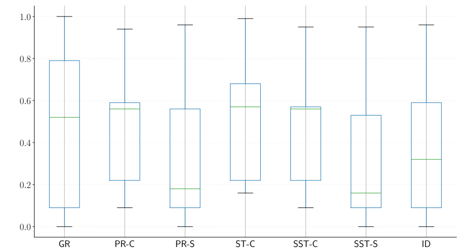
(a) GR



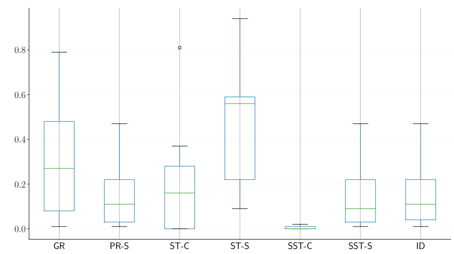
(b) ID



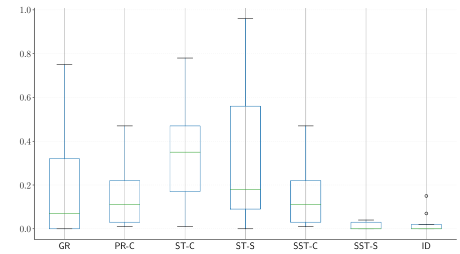
(c) ST-C



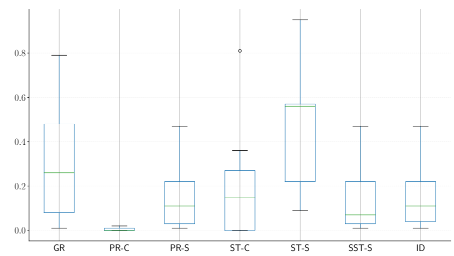
(d) ST-S



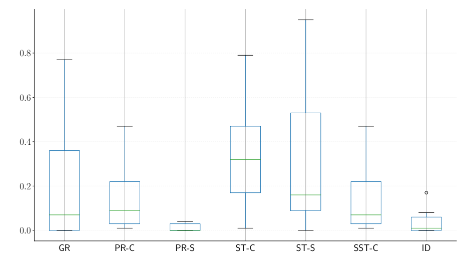
(e) PR-C



(f) PR-S



(g) SST-C



(h) SST-S

Fig. 13. Distributions of Jaccard's distance—aggregating the averages for each dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).

From this analysis, it appears that the most important aspects are: (1) flow hierarchy;[13] (2) the aperiodicity of the graph;[14] and (3) the number of SCCs. There are also some informative matrix-related features, but those are much harder to relate to characteristics of the AFs. Section C provides additional details broken down per benchmark set.

6. Conclusion

In this paper we provide a first answer to the question whether *reasoning with grounded semantics*, which has polynomial runtime, is not already a sufficient approximation approach. As it turns out, in many graphs models reasoning with grounded semantics actually approximates reasoning with other semantics almost perfectly. Indeed, our extensive experimental analysis (Section 5) shows that the grounded extension is a very good—sometimes a perfect—estimator of skeptical acceptance of arguments in AFs derived from the two approaches to structured argumentation we considered in this paper, as well as from graphs obeying to the Barabasi-Albert and Watts-Strogatz models, according to all the semantics except for stable. The main reason for this is that stable semantics is the only semantics for which existence is not guaranteed, and this has clearly a substantial effect on skeptical acceptance. Instead, it appears to be not a good predictor in the case of argumentation frameworks with a substantial number of SSCs, with a large number of stable extensions, and derived from graphs obeying to the Erdös-Renyi model. The AFs derived from graphs obeying to the Erdös-Renyi model also seem to be the ones for which the set of arguments credulously accepted according to preferred semantics is different from the set of arguments credulously accepted according to semi-stable semantics, albeit they both belong to the same 0.05-cluster. We also performed an analysis of graph features that could be mostly informative for predicting the correlation results. Unsurprisingly it appears that the most informative ones are connected to the presence of cycles in the graph.

The extensive experimental section of this paper supports the claim that an algorithm for grounded reasoning is thus a conceptually simple approximation algorithm that does not need an expensive learning phase while having a good performance on several instances of AFs, including some of those closely linked to structured argumentation. On the one hand, this can thus help the development of real-world tools using abstract argumentation, where results, albeit approximate, might need to be provided in a near-real-time setting. On the other hand, it helps shed some light about the benchmarks currently used in the community, and provides effective guidance on the hardness of AFs instances. Indeed, one could argue that benchmarks for which the 0.05-clusters of averaged measures of relative skepticism (i.e. Figure 12) should resemble as much as possible the theoretical results associated to skeptical relationships between semantics, i.e. Figure 2. From visual inspection, the set of AFs exhibiting a large number of stable extensions, as well as those derived from graphs obeying to the Erdös-Renyi model appear to have the 0.05-clusters distributed in a way similar to the theoretical results we know from the analysis of skepticisms of the various semantics.

References

[1] 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) (1995), 321–357.

[13] The fraction of edges not participating in cycles in a directed graph.

[14] A graph is aperiodic if there is no $k > 1$ that is a integer divisor of the length of each cycle in the graph.

[2] M. Caminada, Semi-Stable Semantics, in: *Proceedings of COMMA 2006*, 2006, pp. 121–130.
[3] P.M. Dung, P. Mancarella and F. Toni, Computing ideal sceptical argumentation, *Artificial Intelligence* **171**(10) (2007), 642–674, Argumentation in Artificial Intelligence, ISSN 0004-3702.
[4] P. Baroni, M. Caminada and M. Giacomin, An introduction to argumentation semantics, *Knowledge Engineering Review* **26**(4) (2011), 365–410.
[5] M. Vallati, F. Cerutti and M. Giacomin, Predictive models and abstract argumentation: the case of high-complexity semantics, *The Knowledge Engineering Review* **34** (2019), 6. doi:10.1017/S0269888918000036.
[6] I. Kuhlmann and M. Thimm, Using Graph Convolutional Networks for Approximate Reasoning with Abstract Argumentation Frameworks: A Feasibility Study, in: *Proceedings of the 13th International Conference on Scalable Uncertainty Management (SUM'19)*, Vol. 11940, N.B. Amor, B. Quost and M. Theobald, eds, Springer International Publishing, 2019, pp. 24–37.
[7] T.N. Kipf and M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, *arXiv:1609.02907 [cs, stat]* (2017).
[8] P. Baroni, F. Cerutti, P.E. Dunne and M. Giacomin, Automata for infinite argumentation structures, *Artificial Intelligence* **203**(0) (2013), 104–150, ISSN 0004-3702.
[9] B. Verheij, Two approaches to dialectical argumentation:admissible sets and argumentation stages, in: *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC'96)*, J.-J.C. Meyer and L.C. van der Gaag, eds, Utrecht, NL, 1996, pp. 357–368.
[10] M. Caminada and D.M. Gabbay, A Logical Account of Formal Argumentation, *Studia Logica (Special issue: new ideas in argumentation theory)* **93**(2–3) (2009), 109–145.
[11] P. Baroni and M. Giacomin, Skepticism relations for comparing argumentation semantics, *Int. J. of Approximate Reasoning* **50**(6) (2009), 854–866.
[12] W. Dvořák and P.E. Dunne, Computational Problems in Formal Argumentation and their Complexity, in: *Handbook of Formal Argumentation*, P. Baroni, D. Gabbay, M. Giacomin and L. van der Torre, eds, College Publications, 2018, Chap. 14.
[13] P. Jaccard, Etude de la distribution florale dans une portion des Alpes et du Jura, *Bulletin de la Societe Vaudoise des Sciences Naturelles* **37** (1901), 547–579.
[14] M. Levandowsky and D. Winter, Distance between Sets, *Nature* **234**(5323) (1971), 34–35, ISSN 1476-4687. doi:10.1038/234034a0. https://doi.org/10.1038/234034a0.
[15] R. Craven and F. Toni, Argument graphs and assumption-based argumentation, *Artificial Intelligence* **233** (2016), 1–59, ISSN 0004-3702. doi:https://doi.org/10.1016/j.artint.2015.12.004. http://www.sciencedirect.com/science/article/pii/S0004370215001800.
[16] S. Modgil and H. Prakken, The ASPIC+ framework for structured argumentation: A tutorial, *Argument & Computation* **5** (2014), 31–62.
[17] F. Cerutti, N. Oren, H. Strass, M. Thimm and M. Vallati, A benchmark framework for a computational argumentation competition, in: *Proc. of COMMA*, 2014, pp. 459–460.
[18] P. Erdös and A. Rényi, On random graphs. I, *Publ. Math-Debrecen* **6** (1959), 290–297.
[19] F. Cerutti, M. Giacomin and M. Vallati, Generating Structured Argumentation Frameworks: AFBenchGen2., in: *COMMA*, P. Baroni, T.F. Gordon, T. Scheffler and M. Stede, eds, Frontiers in Artificial Intelligence and Applications, Vol. 287, IOS Press, 2016, pp. 467–468. ISBN 978-1-61499-686-6. http://dblp.uni-trier.de/db/conf/comma/comma2016.html#CeruttiGV16.
[20] A. Barabasi and R. Albert, Emergence of scaling in random networks, *Science* **286**(5439) (1999), 11.
[21] D.J. Watts and S.H. Strogatz, Collective dynamics of 'small-world' networks., *Nature* **393**(6684) (1998), 440–442.
[22] F. Cerutti, M. Vallati and M. Giacomin, On the impact of configuration on abstract argumentation automated reasoning, *Int. J. Approx. Reason.* **92** (2018), 120–138. doi:10.1016/j.ijar.2017.10.002. https://doi.org/10.1016/j.ijar.2017.10.002.
[23] M. Vallati, F. Cerutti and M. Giacomin, On the Combination of Argumentation Solvers into Parallel Portfolios, in: *AI 2017: Advances in Artificial Intelligence - 30th Australasian Joint Conference, Melbourne, VIC, Australia, August 19-20, 2017, Proceedings*, W. Peng, D. Alahakoon and X. Li, eds, Lecture Notes in Computer Science, Vol. 10400, Springer, 2017, pp. 315–327. doi:10.1007/978-3-319-63004-5_25. https://doi.org/10.1007/978-3-319-63004-5_25.
[24] F. Cerutti, M. Vallati and M. Giacomin, Where Are We Now? State of the Art and Future Trends of Solvers for Hard Argumentation Problems, in: *Computational Models of Argument - Proceedings of COMMA 2016, Potsdam, Germany, 12-16 September, 2016*, P. Baroni, T.F. Gordon, T. Scheffler and M. Stede, eds, Frontiers in Artificial Intelligence and Applications, Vol. 287, IOS Press, 2016, pp. 207–218. doi:10.3233/978-1-61499-686-6-207. https://doi.org/10.3233/978-1-61499-686-6-207.
[25] F. Cerutti, M. Giacomin and M. Vallati, Algorithm Selection for Preferred Extensions Enumeration, in: *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, S. Parsons, N. Oren, C. Reed and F. Cerutti, eds, Frontiers in Artificial Intelligence and Applications, Vol. 266, IOS Press, 2014, pp. 221–232. doi:10.3233/978-1-61499-436-7-221. https://doi.org/10.3233/978-1-61499-436-7-221.

[26] M. Vallati, F. Cerutti and M. Giacomin, Argumentation Frameworks Features: an Initial Study, in: *ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, T. Schaub, G. Friedrich and B. O'Sullivan, eds, Frontiers in Artificial Intelligence and Applications, Vol. 263, IOS Press, 2014, pp. 1117–1118. doi:10.3233/978-1-61499-419-0-1117. https://doi.org/10.3233/978-1-61499-419-0-1117.

[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, The WEKA data mining software: an update, *SIGKDD Explorations* **11**(1) (2009), 10–18.

Appendix A. Background in ASPIC and example of ASPIC-like Derived Dung's AF

In the following, we present a minimal variant of the propositional instantiation of ASPIC+ [16]. Note that ASPIC+ is a general framework that can be instantiated using a variety of different logics and is also able to adhere for the inclusion of orderings between rules, but we only stick to a very simple version.[15]

Let $\mathcal{L}$ be a finite set of propositions and let $\hat{\mathcal{L}}$ be the set of literals of $\mathcal{L}$, i. e., $\hat{\mathcal{L}} = \{a, \neg a \mid a \in \mathcal{L}\}$. For $a \in \mathcal{L}$ define $\overline{a} = \neg a$ and $\overline{\neg a} = a$. ASPIC+ differentiates rules into strict rules (rules that are always supposed to hold) and defeasible rules (rules that "usually" hold).

Definition 11. A *knowledge base* $\mathcal{K}$ is a pair $\mathcal{K} = (\mathcal{K}_s, \mathcal{K}_d)$ where

- $\mathcal{K}_s$ is a set of strict rules of the form $\phi_1, \ldots, \phi_n \rightarrow \phi$ with $\phi_1, \ldots, \phi_n, \phi \in \hat{\mathcal{L}}$.
- $\mathcal{K}_d$ is a set of defeasible rules of the form $\phi_1, \ldots, \phi_n \Rightarrow \phi$ with $\phi_1, \ldots, \phi_n, \phi \in \hat{\mathcal{L}}$.

A strict rule $\phi_1, \ldots, \phi_n \rightarrow \phi$ with $n = 0$ is written as $\rightarrow \phi$ and is also called an *axiom*. A defeasible rule $\phi_1, \ldots, \phi_n \Rightarrow \phi$ with $n = 0$ is written as $\Rightarrow \phi$ and is also called an *assumption*. For practical reasons we often identify $\mathcal{K} = (\mathcal{K}_s, \mathcal{K}_d)$ with $\mathcal{K}_s \cup \mathcal{K}_d$, e. g., expressions such as "$r \in \mathcal{K}$" are to be read as "$r \in \mathcal{K}_s$ or $r \in \mathcal{K}_d$".

Arguments can now be constructed by chaining rules. Following [16], for each argument A we denote by $Prem(A)$ the set of axioms and assumptions used to construct A, $Conc(A)$ is the conclusion of A, $Sub(A)$ is the set of sub-arguments of A, $DefRules(A)$ the set of defeasible rules in A, and $TopRule(A)$ is the last rule used in A.

Definition 12. The set of arguments $A_{\mathcal{K}}$ generated by a knowledge base $\mathcal{K} = (\mathcal{K}_s, \mathcal{K}_d)$ is inductively defined as follows:

- If $\Rightarrow \phi \in \mathcal{K}$ then $(\Rightarrow \phi)$ is an argument with $Prem(\Rightarrow \phi) = \{\Rightarrow \phi\}$, $Conc(\Rightarrow \phi) = \phi$, $Sub(\Rightarrow \phi) = \{\Rightarrow \phi\}$, $DefRules(\Rightarrow \phi) = \{\Rightarrow \phi\}$, $TopRule(\Rightarrow \phi) = (\Rightarrow \phi)$.
- If $\rightarrow \phi \in \mathcal{K}$ then $(\rightarrow \phi)$ is an argument with $Prem(\rightarrow \phi) = \{\rightarrow \phi\}$, $Conc(\rightarrow \phi) = \phi$, $Sub(\rightarrow \phi) = \{\rightarrow \phi\}$, $DefRules(\rightarrow \phi) = \emptyset$, $TopRule(\rightarrow \phi) = (\rightarrow \phi)$.
- If $\phi_1, \ldots, \phi_n \Rightarrow \psi \in \mathcal{K}$ and $A_1, \ldots, A_n$ are arguments such that $\phi_1 = Conc(A_1), \ldots, \phi_n = Conc(A_n)$, then $A = (A_1, \ldots, A_n \Rightarrow \psi)$ is an argument such that: $Prem(A) = Prem(A_1) \cup \ldots \cup Prem(A_n)$, $Conc(A) = \psi$, $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$, $DefRules(A) = DefRules(A_1) \cup \ldots \cup DefRules(A_n) \cup \{\phi_1, \ldots, \phi_1 \Rightarrow \psi\}$, $TopRule(A) = \phi_1, \ldots, \phi_n \Rightarrow \psi$.
- If $\phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{K}$ and $A_1, \ldots, A_n$ are arguments such that $\phi_1 = Conc(A_1), \ldots, \phi_n = Conc(A_n)$, then $A = (A_1, \ldots, A_n \rightarrow \psi)$ is an argument such that: $Prem(A) = Prem(A_1) \cup \ldots \cup Prem(A_n)$, $Conc(A) = \psi$, $Sub(A) = Sub(A_1) \cup \ldots \cup Sub(A_n) \cup \{A\}$, $DefRules(A) = DefRules(A_1) \cup \ldots \cup DefRules(A_n)$, $TopRule(A) = \phi_1, \ldots, \phi_n \rightarrow \psi$.

An argument A is called *strict* if $DefRules(A) = \emptyset$, otherwise it is called *defeasible*. In our simplified framework, we only consider *rebuts* [16] as the attack relation between arguments.

Definition 13. Let A and B be two arguments. We say that A attacks B, denoted as $A \rightsquigarrow B$, if $Conc(A) = \overline{a}$ for some $B' \in Sub(A)$ of the form $B''_1, \ldots, B''_n \Rightarrow a$.

Using the previous two definitions an abstract argumentation framework can be derived from a knowledge base $\mathcal{K}$ as follows.

[15] Note also that we depart from ASPIC+ terminology at times.

Table 2

Jaccard's distance—for the aMillion dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.44							
ST-S	0.79	0.64						
PR-C	0.42	0.28	0.59					
PR-S	0.32	0.35	0.55	0.11				
SST-C	0.42	0.27	0.57	0.02	0.11			
SST-S	0.36	0.32	0.52	0.09	0.04	0.07		
ID	0.31	0.35	0.56	0.11	0.01	0.11	0.05	
RAND	0.90	0.85	0.60	0.80	0.83	0.80	0.82	0.83

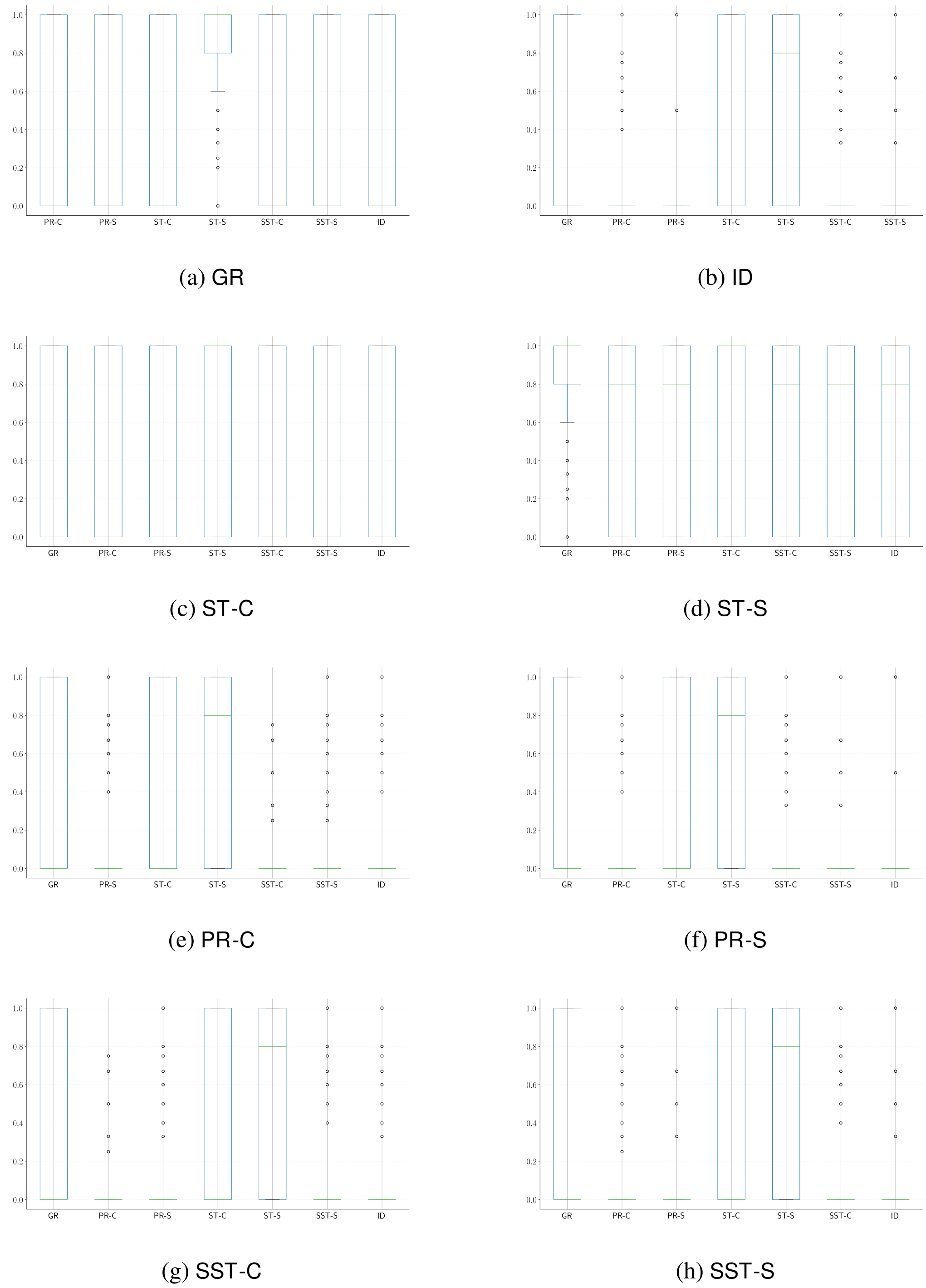
Definition 14. The abstract argumentation framework $AF_{\mathcal{K}}$ corresponding to a knowledge base $\mathcal{K}$ is an argumentation framework $AF_{\mathcal{K}} = (A_{\mathcal{K}}, \rightsquigarrow)$ where $A_{\mathcal{K}}$ is the set of arguments generated by $\mathcal{K}$ as defined by Definition 12 and $\rightsquigarrow$ is the attack relation on $A_{\mathcal{K}}$ as defined by Definition 13.

In the following, an example of a ASPIC-like theory used to derive Dung's AFs of the set sASPIC.

$$
\begin{aligned}
& a_1 \Rightarrow \neg a_1 \\
& a_4 \Rightarrow a_1 \\
& a_5 \Rightarrow a_2 \\
& a_1 \rightarrow \neg a_0 \\
& a_1, \neg a_1 \Rightarrow a_1 \\
& \neg a_3, \neg a_2 \Rightarrow a_1 \\
& \neg a_4, \neg a_3 \Rightarrow \neg a_2 \\
& \neg a_1, a_0 \Rightarrow a_3 \\
& \rightarrow a_2 \\
& \Rightarrow a_3 \\
& a_5, \neg a_5 \rightarrow \neg a_5 \\
& \Rightarrow a_0 \\
& \Rightarrow \neg a_1 \\
& \Rightarrow a_1 \\
& \rightarrow a_1 \\
& \Rightarrow \neg a_2 \\
& \rightarrow \neg a_0 \\
& \neg a_2 \Rightarrow a_4 \\
& \neg a_0 \Rightarrow \neg a_4
\end{aligned}
\tag{1}
$$

Appendix B. Detailed Experimental Results

aMillion. Table 2 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 14 provides a boxplot representation of the distributions.



(a) GR (b) ID

(c) ST-C (d) ST-S

(e) PR-C (f) PR-S

(g) SST-C (h) SST-S

Fig. 14. Distributions of Jaccard's distance—for the aMillion dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).

Table 3

Jaccard's distance—for the sABA dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.18							
ST-S	0.13	0.16						
PR-C	0.06	0.13	0.13					
PR-S	0.03	0.16	0.10	0.03				
SST-C	0.06	0.13	0.13	0.00	0.03			
SST-S	0.03	0.16	0.10	0.03	0.00	0.03		
ID	0.03	0.17	0.11	0.04	0.00	0.04	0.01	
RAND	0.75	0.76	0.70	0.74	0.74	0.74	0.74	0.75

Table 4

Jaccard's distance—for the sASPIC dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.48							
ST-S	0.00	0.47						
PR-C	0.48	0.00	0.47					
PR-S	0.00	0.47	0.00	0.47				
SST-C	0.48	0.00	0.47	0.00	0.47			
SST-S	0.00	0.47	0.00	0.47	0.00	0.47		
ID	0.00	0.47	0.00	0.47	0.00	0.47	0.00	
RAND	0.86	0.69	0.86	0.69	0.86	0.69	0.86	0.86

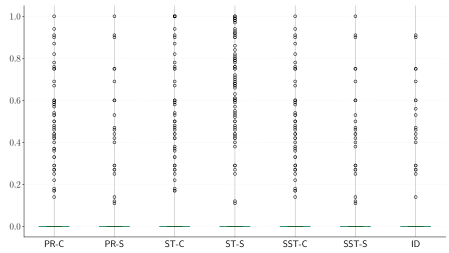
sABA. Table 3 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 15 provides a boxplot representation of the distributions.

sASPIC. Table 4 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 16 provides a boxplot representation of the distributions.

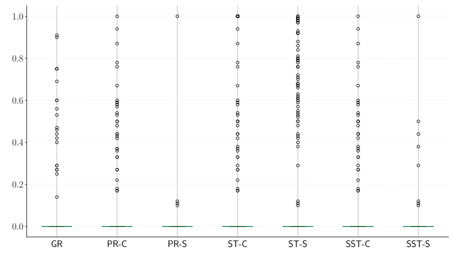
rSCC. Table 5 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 17 provides a boxplot representation of the distributions.

rStable. Table 6 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 18 provides a boxplot representation of the distributions.

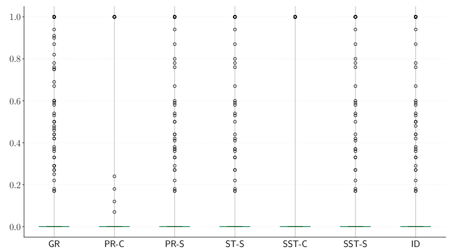
rER. Table 7 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 19 provides a boxplot representation of the distributions.



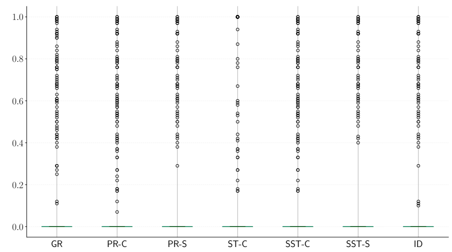
(a) GR



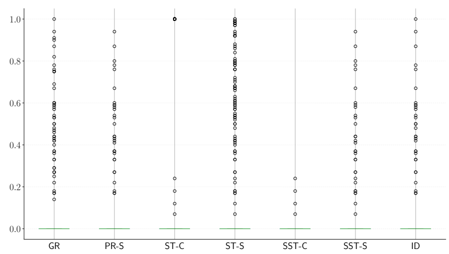
(b) ID



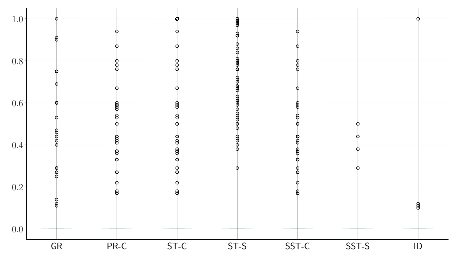
(c) ST-C



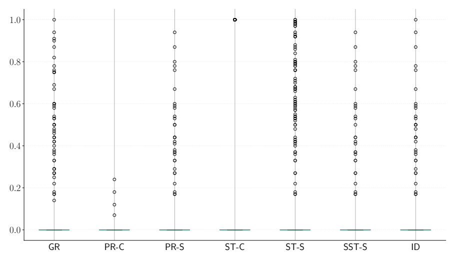
(d) ST-S



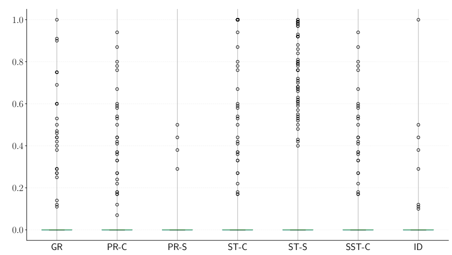
(e) PR-C



(f) PR-S

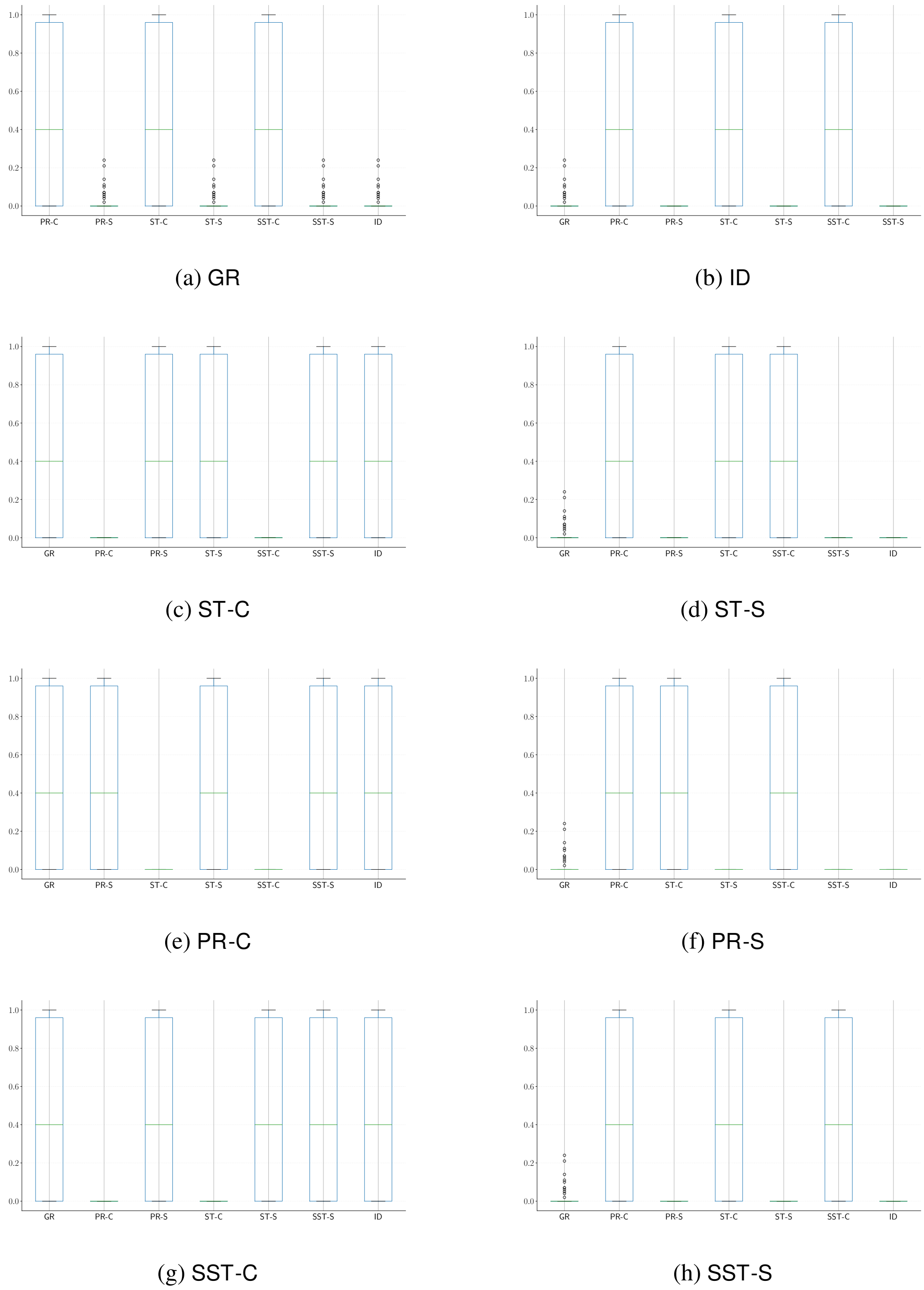


(g) SST-C



(h) SST-S

Fig. 15. Distributions of Jaccard's distance—for the sABA dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).



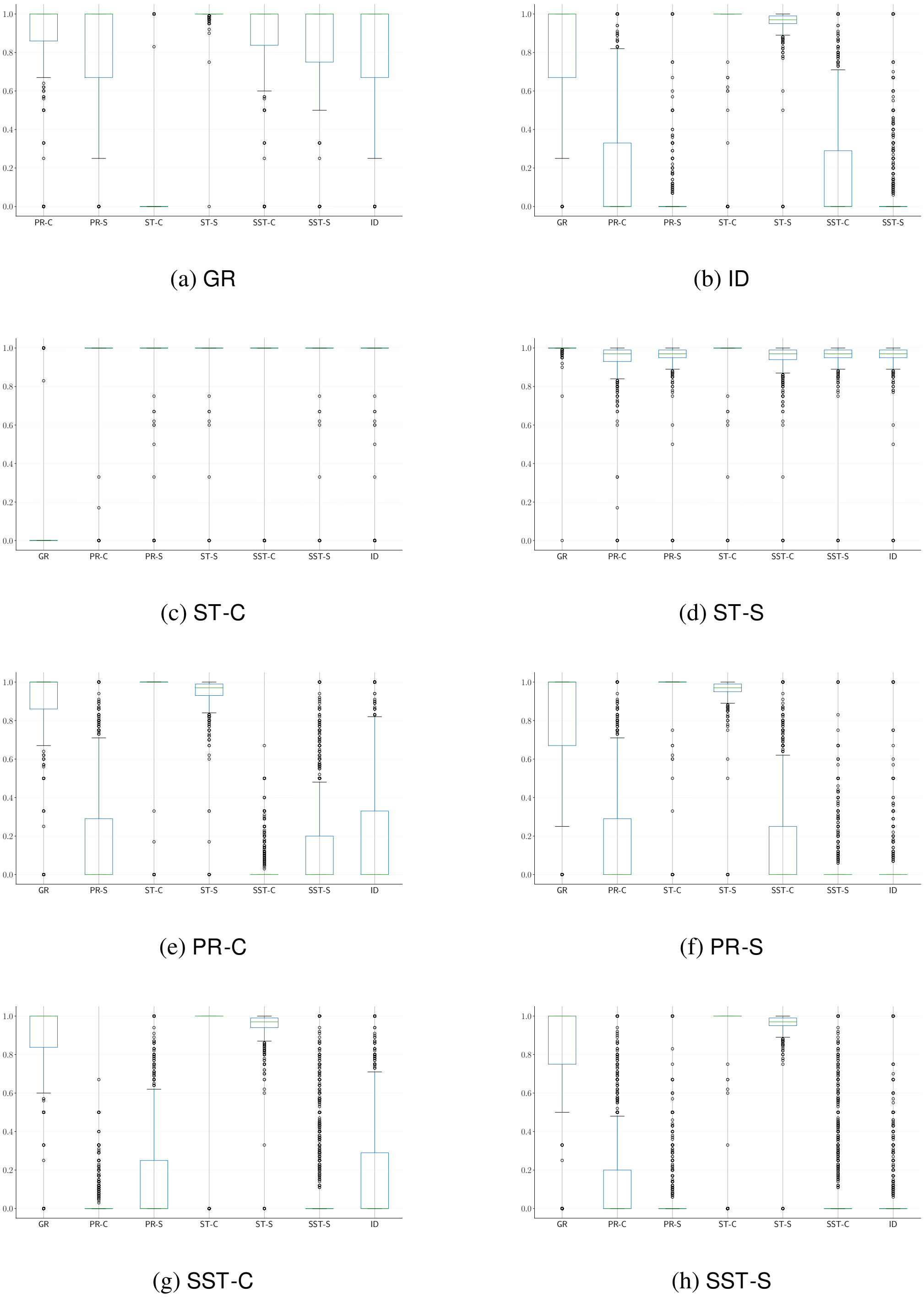
(a) GR (b) ID

(c) ST-C (d) ST-S

(e) PR-C (f) PR-S

(g) SST-C (h) SST-S

Fig. 16. Distributions of Jaccard's distance—for the sASPIC dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).



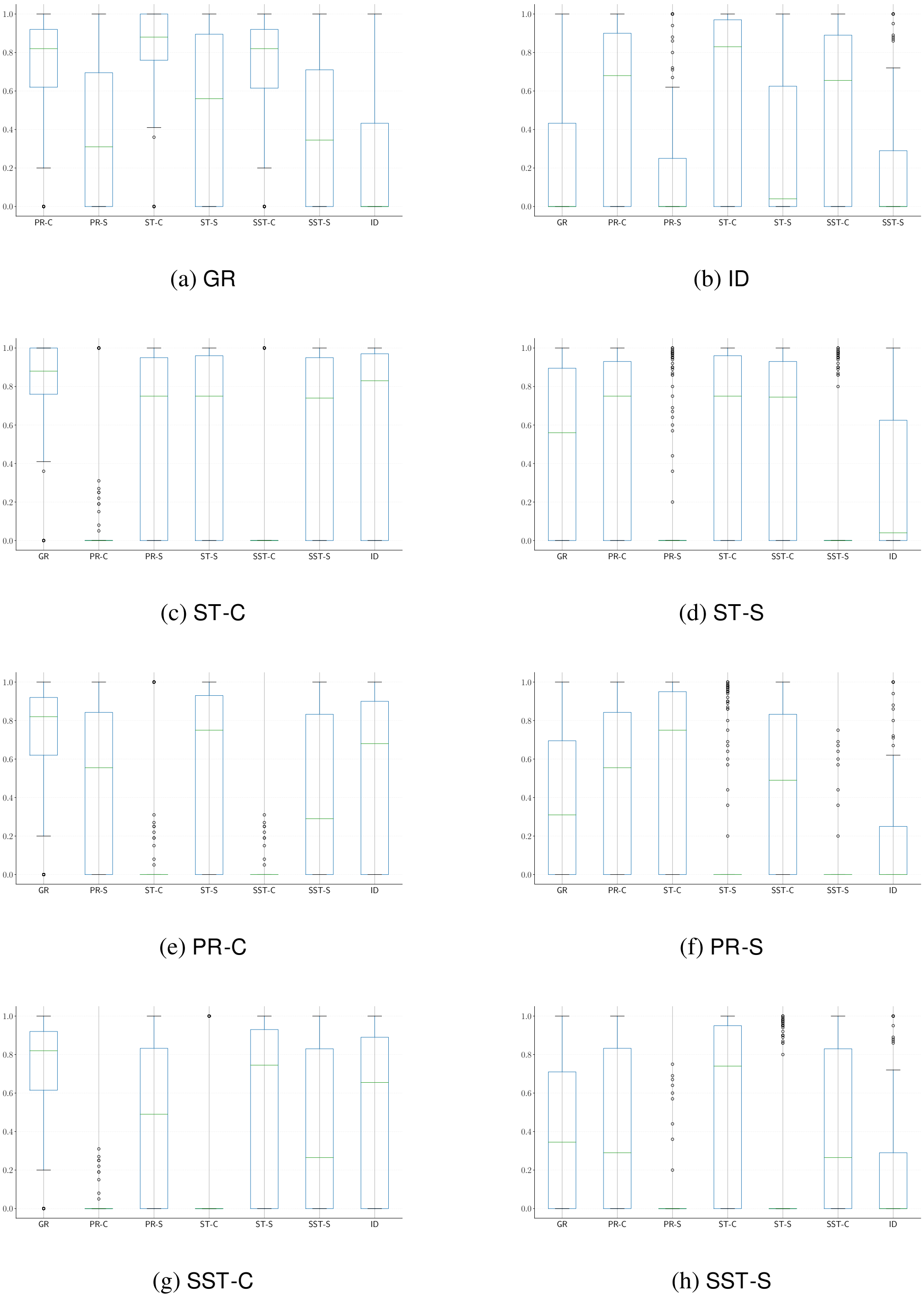
(a) GR (b) ID

(c) ST-C (d) ST-S

(e) PR-C (f) PR-S

(g) SST-C (h) SST-S

Fig. 17. Distributions of Jaccard's distance—for the rSCC dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).



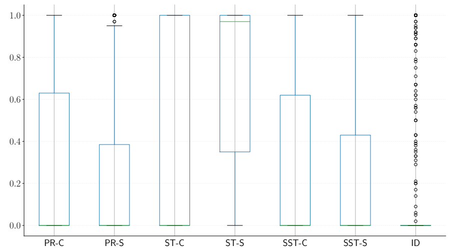
(a) GR (b) ID

(c) ST-C (d) ST-S

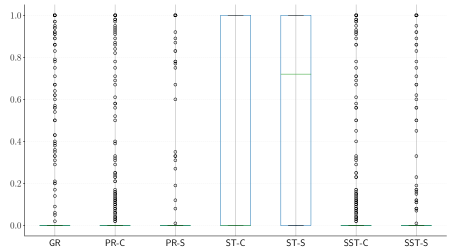
(e) PR-C (f) PR-S

(g) SST-C (h) SST-S

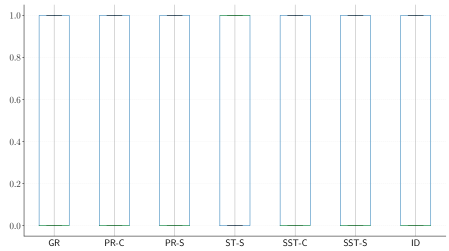
Fig. 18. Distributions of Jaccard's distance—for the rStable dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).



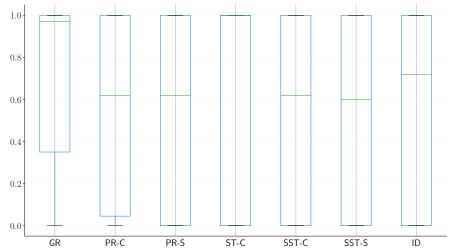
(a) GR



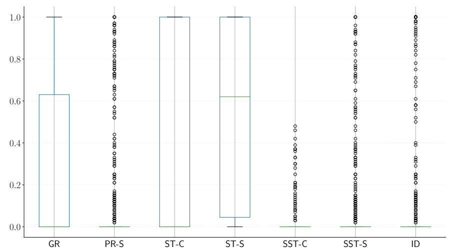
(b) ID



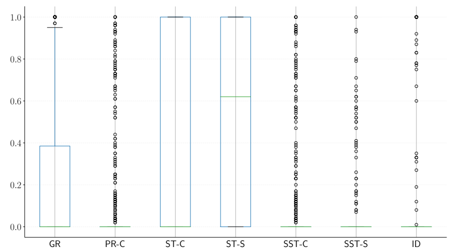
(c) ST-C



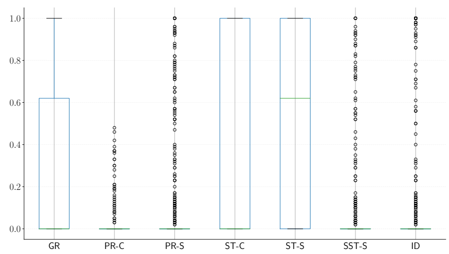
(d) ST-S



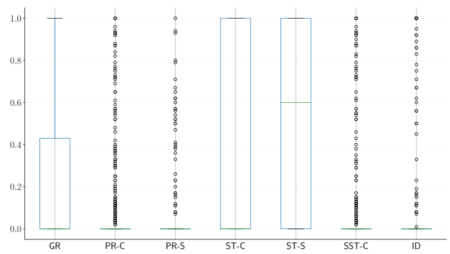
(e) PR-C



(f) PR-S



(g) SST-C



(h) SST-S

Fig. 19. Distributions of Jaccard's distance—for the rER dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).

Table 5

Jaccard's distance—for the rSCC dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.13							
ST-S	1.00	0.99						
PR-C	0.79	0.81	0.94					
PR-S	0.75	0.78	0.96	0.16				
SST-C	0.79	0.81	0.95	0.02	0.15			
SST-S	0.77	0.79	0.95	0.14	0.04	0.12		
ID	0.74	0.77	0.96	0.17	0.02	0.16	0.06	
RAND	1.00	1.00	0.51	0.96	0.97	0.96	0.97	0.97

Table 6

Jaccard's distance—for the rStable dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.80							
ST-S	0.52	0.57						
PR-C	0.68	0.16	0.56					
PR-S	0.36	0.57	0.18	0.43				
SST-C	0.68	0.15	0.56	0.01	0.42			
SST-S	0.38	0.55	0.16	0.41	0.02	0.41		
ID	0.21	0.61	0.32	0.47	0.15	0.47	0.17	
RAND	0.95	0.83	0.83	0.82	0.91	0.82	0.91	0.92

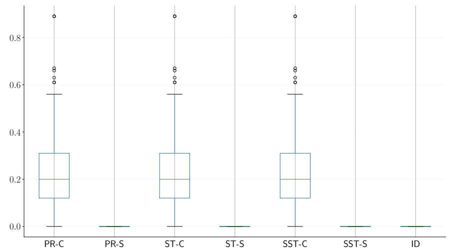
Table 7

Jaccard's distance—for the rER dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

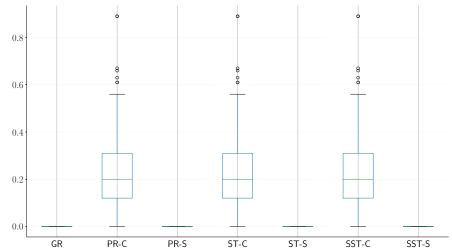
	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.41							
ST-S	0.68	0.68						
PR-C	0.27	0.37	0.57					
PR-S	0.23	0.40	0.56	0.09				
SST-C	0.26	0.36	0.56	0.01	0.08			
SST-S	0.24	0.39	0.53	0.07	0.03	0.06		
ID	0.17	0.40	0.59	0.11	0.07	0.10	0.08	
RAND	0.85	0.89	0.58	0.80	0.82	0.80	0.82	0.83

rBA. Table 8 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 20 provides a boxplot representation of the distributions.

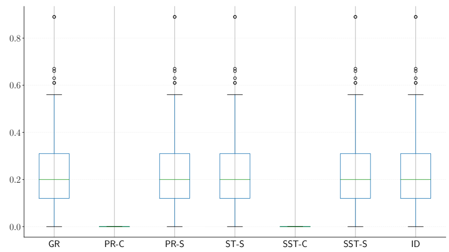
rWS. Table 9 summarises the average Jaccard's distance among the various sets of credulously or skeptically accepted arguments for the semantics identified in Definition 3, and Figure 21 provides a



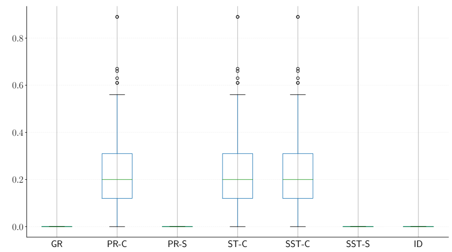
(a) GR



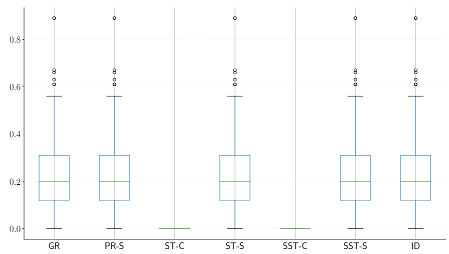
(b) ID



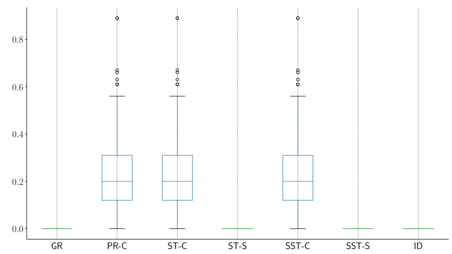
(c) ST-C



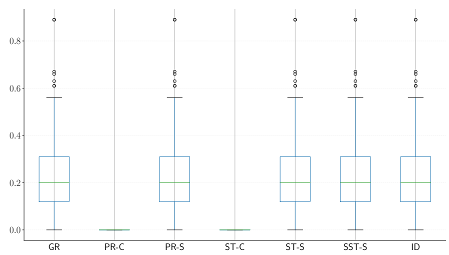
(d) ST-S



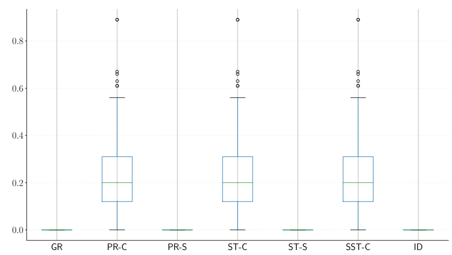
(e) PR-C



(f) PR-S



(g) SST-C



(h) SST-S

Fig. 20. Distributions of Jaccard's distance—for the rBA dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).

Table 8

Jaccard's distance—for the rBA dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.22							
ST-S	0.00	0.22						
PR-C	0.22	0.00	0.22					
PR-S	0.00	0.22	0.00	0.22				
SST-C	0.22	0.00	0.22	0.00	0.22			
SST-S	0.00	0.22	0.00	0.22	0.00	0.22		
ID	0.00	0.22	0.00	0.22	0.00	0.22	0.00	
RAND	0.64	0.57	0.64	0.57	0.64	0.57	0.64	0.64

Table 9

Jaccard's distance—for the rWS dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3.

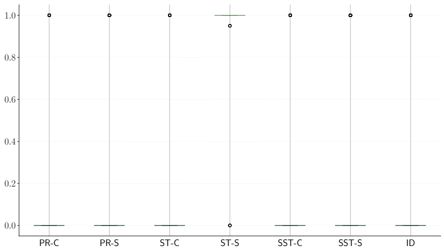
	GR	ST-C	ST-S	PR-C	PR-S	SST-C	SST-S	ID
ST-C	0.08							
ST-S	1.00	0.93						
PR-C	0.08	0.00	0.93					
PR-S	0.07	0.01	0.92	0.01				
SST-C	0.08	0.00	0.93	0.00	0.01			
SST-S	0.07	0.01	0.92	0.01	0.00	0.01		
ID	0.07	0.01	0.93	0.01	0.00	0.01	0.00	
RAND	1.00	0.99	0.53	0.99	0.99	0.99	0.99	0.99

boxplot representation of the distributions.

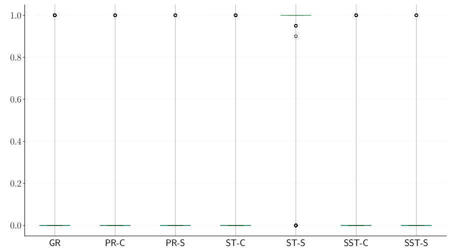
Appendix C. Detailed results of feature selection for relative measure of skepticism

C.1. aMillion

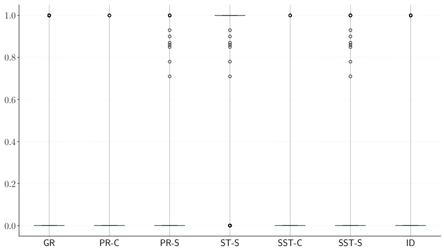
- ST-C: only matrix-related and number of argcs in undirected graph have some relevance.
- ST-S: set of very informative features. In particular, aperiodicity and flow hierarchy.
- PR-C: set of informative features, most of them matrix-related, flow hierarchy, number of scc, aperiodicity. In particular, flow hierarchy seems to be the most important for predictions.
- PR-S: quite hard to deal with, but looks like flow hierarchy, number of SCCs, and aperiodicity are somehow useful.
- SST-C: set of very informative features. In particular, flow hierarchy and a few from matrix representation.
- SST-S: only 3 useful features: flow hierarchy, number of SCCs, and aperiodicity.
- ID: as in SST-S.



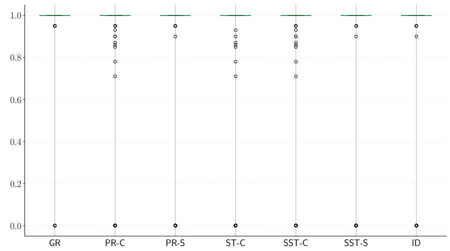
(a) GR



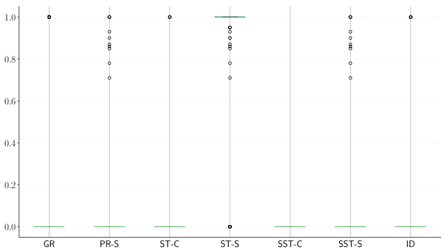
(b) ID



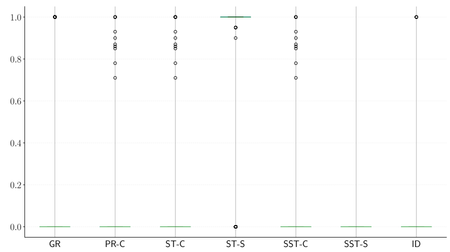
(c) ST-C



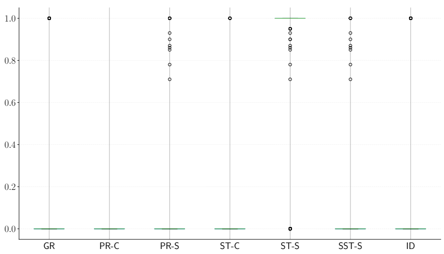
(d) ST-S



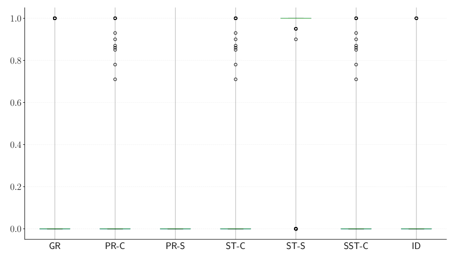
(e) PR-C



(f) PR-S



(g) SST-C



(h) SST-S

Fig. 21. Distributions of Jaccard's distance—for the rWS dataset—between the various sets of credulously or skeptically accepted arguments for the various semantics identified in Definition 3 and: GR (a) (equivalent to μ_S); ID (b); ST-C (c); ST-S (d); PR-C (e); PR-S (f); SST-C (g); SST-S (h).

C.2. sABA

- ST-C: not very informative features, but mostly related with matrix representation, ratio edges/arcs, and flow hierarchy.
- ST-S: not very informative features, but mostly related with matrix representation, ratio edges/arcs, and flow hierarchy.
- PR-C: small set of moderately informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, and matrix-related For predictions, flow hierarchy and aperiodicty are the most relevant.
- PR-S: only 2 informative features: flow hierarchy and aperiodicity
- SST-C: small set of moderately informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, and matrix-related For predictions, flow hierarchy and aperiodicty are the most important.
- SST-S: only aperiodicity and flow hierarchy look to be quite informative.
- ID: extremely hard to predict, we are not able to extract any meaningful piece of information.

C.3. sASPIC

- ST-C: small set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, and number of SCCs. For predictions, flow hierarchy, and aperiodicty are the most important.
- ST-S: extremely hard to predict, we are not able to extract any meaningful piece of information.
- PR-C: small set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, and number of SCCs. For predictions, flow hierarchy and aperiodicty are the most important.
- PR-S: extremely hard to predict, we are not able to extract any meaningful piece of information.
- SST-C: small set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, and number of SCCs. For predictions, flow hierarchy and aperiodicty are the most important.
- SST-S: extremely hard to predict, we are not able to extract any meaningful piece of information..
- ID: extremely hard to predict, we are not able to extract any meaningful piece of information.

C.4. rSCC

- ST-C: huge set of informative features. Aperiodicity is the most informative, while aperiodicity and matrix variance are the most important for predictions.
- ST-S: hard to predict, with just a small set of not-very-informative features. Aperiodicity, ratio edges/arcs, ratio edges/arcs on undirected graph and variance of the matrix diagonal are somewhat used in predictions.
- PR-C: set of quite informative features, mostly related to degree and density, together with some aspects of matrix representation. In predictions, those related to matrix representation seem to be very relevant.
- PR-S: set of quite informative features, mostly related to degree, density, and transitivity, together with some aspects of matrix representation, in particular: transitivity and matrix variance.

- SST-C: large set of quite informative features, mostly related to degree and density, together with some aspects of matrix representation. For predictions, matrix-related average seems to be the most informative feature.
- SST-S: set of quite informative features, mostly related to degree, density, and transitivity, together with some aspects of matrix representation, in particular: transitivity and matrix variance.
- ID: set of quite informative features, mostly related to degree, density, and transitivity, together to some aspects of matrix representation, in particular, transitivity and matrix variance.

C.5. rStable

- ST-C: small set of very informative features: flow hierarchy is the most important (also in prediction), and then ratio edges/arcs, and ratio edges/arcs on undirected graph.
- ST-S: large set of moderately informative features. For predictions, flow hiearchy is the most important.
- PR-C: small set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, matrix-related, and number of SCCs. For predictions, flow hierarchy and aperiodicty are the most important.
- PR-S: very hard to deal with. It looks that only flow hiearchy does provide some sort of information, albeit very limited.
- SST-C: small set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, matrix-related, and number of SCCs. For predictions, flow hierarchy and aperiodicty are the most important.
- SST-S: small set of not very informative features. Again, it looks like flow hierarchy is the most important.
- ID: extremely hard to predict, we are not able to extract any meaningful piece of information.

C.6. rER

- ST-C: very large set of moderately informative features (some 20). Aperioditicy and those from matrix seem to be most informative. For predictions, aperiodicity and matdifdiag (difference of diagonal in matrix representation) are quite informative.
- ST-S: large set of very informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), aperiodicity, matrix-related, and number of SCCs. For predictions, flow hierarchy, aperiodicty and matdifdiag are the most important.
- PR-C: set of moderately informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), number of SCCs, and aperiodicity. In Particular, flow hierarchy seems to be the most important for predictions.
- PR-S: as per before, together with some features from matrix representation. In particular, flow hierarchy seems to be the most important for predictions.
- SST-C: very similar to PR-C.
- SST-S: set of informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), number of SCCs, and aperiodicity. In particular, flow hierarchy seems to be the most important for predictions.
- ID: set of informative features: flow hierarchy, ratio edges/arcs, ratio edges/arcs on undirected graph (multiple arcs collapsed), number of SCCs, and aperiodicity. In particular, flow hierarchy seems to be the most important for predictions.

C.7. rBA

- ST-C: basically, as in PR-C.
- PR-C: very informative small set of features: flow hierarchy, ratio edges/arcs, and number of SCCs. In Particular, flow hierarchy seems to be the most important for predictions.
- SST-C: as in PR-C.

C.8. rWS

In this class we have the same behaviour for all the considered perspectives. There is a huge set of seemingly informative features, too many to list. Combined with the quite poor predictive performance, this may indicate that we do not capture the right aspect for relating the grounded extension with the set of credulously and skeptically accepted arguments w.r.t. other semantics in this set of AFs.