A propagation-based ranking semantics in Explainable Bipolar Weighted Argumentation

Original Citation:
A propagation-based ranking semantics in Explainable Bipolar Weighted Argumentation / Fasciano, Corrado; Loseto, Giuseppe; Pinto, Agnese; Ruta, Michele; Scioscia, Floriano. - In: ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE. - ISSN 0952-1976. - STAMPA. - 142:(2025). [10.1016/j.engappai.2024.109767]

Published version
DOI:10.1016/j.engappai.2024.109767

Publisher:

A propagation-based ranking semantics in Explainable Bipolar Weighted Argumentation

Corrado **Fasciano**[a], Giuseppe **Loseto**[b], Agnese **Pinto**[a], Michele **Ruta**[a] and Floriano **Scioscia**[a,*]

[a]*Department of Electrical and Information Engineering, Polytechnic University of Bari, via E. Orabona 4, Bari, I-70125, Italy*
[b]*LUM "Giuseppe Degennaro" University, Strada Statale 100 km 18, Casamassima, I-70010, Italy*

ARTICLE INFO

Keywords:
Bipolar Weighted Argumentation
Semantic matchmaking
Ranking Semantics
Web Ontology Language
Multi-Agent Systems

ABSTRACT

The Semantic Web of Things enables exchanging knowledge-based fragments in pervasive computing through machine-to-machine interactions, sustaining collaborative decision-making in networks of smart objects. Abstract argumentation is widely acknowledged, instead, as a general framework for decision-making in multi-agent systems, able to solve negotiation conflicts and evaluate acceptable options for a given scenario. Nevertheless, abstract argumentation frameworks disregard the problem of characterizing the argument structure and their mutual relations. This paper proposes a novel approach integrating Dung-style abstract argumentation with semantic matchmaking in pervasive computing. Object interactions are seen as an argumentative dialogue, where annotations shared by each agent play the role of arguments. A matchmaking scheme, exploiting non-standard non-monotonic inferences, allows the appraisal of argument relations in a bipolar weighted argumentation approach. A propagation-based gradual semantics provides acceptability ranking of arguments with a formal explanation. A fading mechanism is also exploited to take into account limited computational resources. As the proposed approach enables applications in cyber-physical multi-agent systems engineering, early validation and performance experiments are provided by means of a case study on a real-time strategy game.

1. Introduction

The Semantic Web of Things (SWoT) paradigm [1] aims at integrating Semantic Web and Internet of Things (IoT) technologies in pervasive computing scenarios. SWoT environments support automated reasoning on so-called *ubiquitous knowledge bases*, comprising individual knowledge fragments physically disseminated among heterogeneous smart objects interconnected in mobile ad-hoc networks (MANETs). Knowledge exchange occurs via machine-to-machine interactions based on collaborative protocols. Basically, SWoT leverages ontologies and semantic annotations to ensure that data exchanged between IoT devices is not only syntactically compatible but also semantically meaningful [2, 3, 4]. This allows machines to understand the context and content of the data they receive, enabling more intelligent and autonomous decision-making processes. Through the use of reasoning engines, smart objects can also infer new knowledge from existing data. This means that machines can deduce implicit information based on the explicit data they receive, enabling more sophisticated interactions and the ability to perform complex tasks collaboratively. This distributed approach aims to enable devices to pool their knowledge and coordinate their behaviors, leading to more comprehensive and accurate outcomes in several IoT and pervasive computing domains such as wireless sensor networks [5, 2, 6], healthcare [7, 8], smart transportation [9, 10, 11], and home automation [12]. These dynamic interactions can be considered as an ongoing *dialogue* among independent agents [13, 14]. Of course, agreements or conflicts about object assertions can happen. Particularly, disagreements should be solved in order to achieve decentralized coordination and decision-making in significant portions of a MANET.

In this perspective, the *Abstract Argumentation* (AA) theory could be useful as a general-purpose formalism enabling reasoning over conflicting knowledge. AA translates a collection of arguments in a graph, whose edges

---

[*] Corresponding author

✉ `corrado.fasciano@poliba.it` (C. Fasciano); `loseto@lum.it` (G. Loseto); `agnese.pinto@poliba.it` (A. Pinto); `michele.ruta@poliba.it` (M. Ruta); `floriano.scioscia@poliba.it` (F. Scioscia)

https://sisinflab.poliba.it/fasciano/ (C. Fasciano); https://www.lum.it/docenti/giuseppe-loseto/ (G. Loseto); https://sisinflab.poliba.it/pinto/ (A. Pinto); https://sisinflab.poliba.it/ruta/ (M. Ruta); https://sisinflab.poliba.it/scioscia/ (F. Scioscia)

ORCID(s): 0000-0001-7813-5915 (C. Fasciano); 0000-0002-7995-8494 (G. Loseto); 0000-0002-2502-7467 (A. Pinto); 0000-0003-2125-327X (M. Ruta); 0000-0002-7859-9602 (F. Scioscia)

represent the arguments' relations, so that AA allows evaluating the acceptability of an argument based purely on its relationships and abstracted from its content. As stated by [15], *«an argument is a set of assumptions (*i.e.*, information from which conclusions can be drawn), together with a conclusion that can be obtained by one or more reasoning steps»*. Although there exist approaches considering strict or unattackable arguments such as Defeasible Logic Programming (DeLP) [16] and ASPIC$^+$ [17], given a problem to solve (such as strategic decision-making, reasoning with uncertain information, or event classification), arguments are often different from proofs, inasmuch they are *defeasible*, *i.e.*, the validity of conclusions can be disputed by other arguments if new evidence emerges. The SWoT knowledge interchange fits very well the classical *Dung-style* argumentation paradigms [18], nevertheless proper and comprehensive integration of principled argumentation frameworks in SWoT contexts has not been proposed up to now. AA approaches assume that a set of arguments interacting with each other is provided, thereby disregarding the problem of characterizing the arguments structure [19]: this is, however, a fundamental aspect for a concrete appraisal of relations in data-driven contexts such as pervasive computing. Conversely, *Structured Argumentation* (SA) [20] approaches extend AA by adopting formal models to build arguments and inference procedures to implement relations. Well-known SA approaches in literature include ASPIC$^+$ [17], DeLP [16], Assumption-Based Argumentation (ABA) [21], deductive argumentation [22] and others. Nevertheless, the majority of SA frameworks can explain outcomes regarding the acceptability of arguments in a shallow way, merely based on information about the direction, the type (attack or support) and/or the weight assigned to each relation or votes/preferences associated to arguments [16, 17]. In the light of the increasing attention to eXplainable Artificial Intelligence (XAI) [23], this is a barrier for more meaningful interpretability of models and results. In fact, XAI calls for models, techniques and systems which are both *explainable*, *i.e.*, able to justify their decisions to end users and other stakeholders, and *interpretable*, *i.e.*, where users can understand how inputs are linked to outputs [24]. This makes intelligent systems more transparent and therefore trustworthy, as increasingly required by national and international regulations. [25, 26]

This paper introduces a general-purpose approach for improving argumentation frameworks adopting Description Logics-based Knowledge Representation and Reasoning (KRR) to represent arguments, evaluate their relations and rank their acceptability by means of a novel *propagation-based ranking semantics*. Representing high-level information with machine-understandable languages introduces several benefits dealing with key open challenges for fully automatic argumentation, as discussed *e.g.*, in [27]: *i.e.*, how arguments are constructed (*structural layer*), what are the relationships between arguments (*relational layer*), and how a constellation of interacting arguments can be evaluated and conclusions can be drawn (*assessment layer*).

The main contributions of the paper can be summarized hereafter:

- **SWoT for argumentative agents.** Information generated and shared by an agent is annotated according to reference ontologies in standard Semantic Web languages, grounded on Description Logics (DLs) [28]. Specifically, the paper refers to a *Web Ontology Language* (OWL) 2 [29] subset corresponding to the *Attributive Language with unqualified Number restrictions* ($\mathcal{ALN}$) DL, which allows polynomial-time standard and non-standard inference procedures for acyclic ontologies.

- **OWL-based *Bipolar Weighted Argumentation Framework*.** The approach can be integrated with any classical AA framework, where each semantic annotation takes the role of an argument. However, in order to allow more nuanced analysis, a Bipolar Weighted Argumentation Framework (BWAF) extension is adopted, including both attack (a.k.a. defeat) and support relations in the interpretation introduced by [30], with an assigned weight to represent connection type and strength.

- **Non-monotonic reasoning for explainable evaluation of argument relations.** The type and weight of relations are computed leveraging non-standard, non-monotonic inference services including *Concept Contraction*, *Concept Abduction* and *Concept Bonus* [31]. When comparing two arguments represented as $\mathcal{ALN}$ concept expressions, they return another concept expression upon which a *semantic distance* metric can be calculated. The overall relation weight is computed as a combination of distance scores and the corresponding concept expressions act as explicit logic-based explanations. This enables a more meaningful appraisal of individual relations and clear interpretability of results for end users and stakeholders, by virtue of logic-based explanation capabilities of the adopted deductions.

- **New gradual semantics for argument acceptability appraisal.** The framework includes a novel *propagation-based ranking semantics* to compute an acceptability score for each argument. The method combines *path*

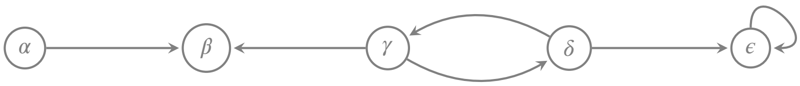


Figure 1: $\mathcal{F}_1$: AF example

strength propagation with *fading* to take into account the resource limitations of pervasive computing devices. Differently from the majority of existing BWAF proposals, it also solves cyclic argumentative graphs, through an iterative procedure with halt conditions, which allow to limit computational complexity and performance challenges.

- **Validation in relevant environment.** The proposal has been applied to a case study regarding tactical decision-making in a real-time strategy game engine, in order to highlight peculiarities of the proposed deductive argumentative reasoning framework in a simulated setting comparable to real-world pervasive SWoT contexts in terms of complexity and unpredictability. An empirical analysis of the effectiveness and explainability benefits of the proposed approach has been conducted, together with experimental tests on computational performance in the reference case study as well as in a new dataset with random graphs and an existing dataset of three well-known graph topologies. Experiments have highlighted the influence of various the argumentation graphs on performance and stability of results.

The remainder of the paper is organized as follows. Section 2 recall useful notions on argumentation and semantic matchmaking in DL, and related work is analyzed in Section 3. Section 4 describes the proposed BWAF framework based on semantic matchmaking to assess argument relations, then Section 5 focuses on the proposed ranking semantics. The case study is in Section 6, while experimental results are reported and discussed in Section 7, before conclusion.

2. Background

In order to make the paper self-contained, this section recalls some concepts and elementary properties concerning Argumentation theory and inference services for semantic matchmaking in Description Logics.

2.1. Argumentation basics

For people, argumentation is pervasive throughout everyday life: it is particularly important for negotiating decisions and managing conflicts in scenarios where the available information is inconsistent or incomplete. *Computational Argumentation* studies models and methods to endow automated agents with the capability to build and assess arguments and counterarguments. Despite its rather long history, it is still gaining relevance within Artificial Intelligence (AI) and XAI [32].

Dung's seminal work [18] defines an *Argumentation Framework* (AF) as a graph-based formalism to reason over conflicting knowledge without considering the internal structure of the arguments, but only their mutual relations of *attack* –denoting the conflicts between pairs of arguments– and the *semantics* for evaluating them, *i.e.*, for determining what arguments can be considered as acceptable.

Definition 1. (Argumentation Framework) *An Argumentation Framework (AF) is a pair* $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$*, where* $\mathcal{A}$ *is a finite set of arguments and* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is the set of relations. The relation* $\alpha\ \mathcal{R}\ \beta$ *means that* α *attacks* β*, or equivalently* β *is attacked by* α*.*

Typically, an AF is depicted as a directed graph, where the nodes represent arguments, and each edge between two nodes indicates the relation from the attacking argument to the argument being attacked. An example of an AF is shown in Figure 1.

Argumentation semantics formally define the method governing the argument evaluation process. The work in [33] provides one of the first systematic analyses of *extension-based* approaches, where semantics determine how to derive a set of extensions from an AF. An *extension* E of an AF $\langle \mathcal{A}, \mathcal{R} \rangle$ is a subset of $\mathcal{A}$, representing intuitively a group of arguments that are "collectively acceptable" or can "survive together." A specific extension-based argumentation semantics provides a mechanism to select reasonable sets of arguments from all possible sets, according to criteria embedded in its definition.

Definition 2 (Extension-based Semantics). *Given an AF* $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$*, an extension-based semantics* $\mathcal{S}$ *associates* $\mathcal{F}$ *with a subset of* $2^{\mathcal{A}}$*, denoted as* $\mathcal{E}_{\mathcal{S}}(\mathcal{F})$*.*

Having only two levels of evaluation –arguments either accepted or rejected– is a significant limitation when using argumentation for decision-making in pervasive contexts, where the collected information is highly heterogeneous and conflicts are frequent. In such cases, even sub-optimal solutions could be considered useful if optimal ones cannot be found, due to unpredictable context changes and computational limitations of devices. In order to overcome these shortcomings, one may want to adopt semantics which distinguish arguments with various degrees *overall strength*, *i.e.*, degree of acceptability. The prevalent approaches rank arguments from the most to the least acceptable. *Ranking-based* semantics [34, 35] (a.k.a. *gradual semantics*) aim at determining such an ordering among a set of arguments. Several ranking-based semantics can be found in literature, with a range of behaviors and logical properties.

Definition 3 (Ranking-based Semantics). *Ranking-based semantics* $\mathcal{S}$ *associates to any AF* $\mathcal{F} = \langle \mathcal{A}, \mathcal{R} \rangle$ *a ranking* $\succeq^{\mathcal{S}}_{\mathcal{F}}$ *on* $\mathcal{A}$*, where* $\succeq^{\mathcal{S}}_{\mathcal{F}}$ *is a preorder* (i.e.*, a reflexive and transitive relation) on* $\mathcal{A}$*.* $\alpha \succeq^{\mathcal{S}}_{\mathcal{F}} \beta$ *means that* α *is at least as acceptable as* β*.*

2.2. Semantic matchmaking in Description Logics

In the SWoT, software agents running on ubiquitous devices may extract, process, and share meaningful information fragments in pervasive contexts to enable adaptive context-aware actions in a wide range of applications. Agents must be able to integrate information that has been detected and received, as well as manage potential disagreements or missing perceptions in a cooperative and coordinated manner. Semantic Web languages and technologies grounded on DL [28] allow information modelling based on formal and rigorous interpretation of its meaning. This enables not only greater interoperability across different hardware/software platforms, but also reasoning tasks *i.e.*, inferences, to extract new implicit insight from information explicitly asserted in a Knowledge Base (KB). Given a DL *ontology* (a.k.a. *terminology*) $\mathcal{T}$ and S, R two satisfiable concepts in $\mathcal{T}$, the *satisfiability* and *subsumption* standard inference services provided by DL-based systems can be formalized as follows:

- *Satisfiability*: verifies if the conjunction of S and R is satisfiable w.r.t. the ontology $\mathcal{T}$; in formulas, $\mathcal{T} \not\models S \sqcap R \sqsubseteq \bot$, where $\models$ denotes entailment, $\sqcap$ conjunction, $\sqsubseteq$ the subsumption (*i.e.*, more specific than) relation and $\bot$ the *empty concept* a.k.a. *bottom*.
- *Subsumption*: checks if S is more specific than R w.r.t. the ontology $\mathcal{T}$; in formulas, *i.e.*, $\mathcal{T} \models S \sqsubseteq R$.

Semantic matchmaking is defined as the problem of finding the most relevant element in a set of *resources* w.r.t. a given *request*, where both request and resources are represented as satisfiable concept expressions w.r.t. a common ontology $\mathcal{T}$ in a DL. The output is a list of resources, each with a score representing its semantic relevance w.r.t. the submitted query. In real-world scenarios, frequently featuring detailed descriptions with possible mutual conflicts, standard inference services like *subsumption* and *satisfiability* checks are ineffective, since they return true/false only. Analogously to switching from extension-based to ranking-based semantics in argumentation frameworks, a more fine-grained outcome is required in advanced settings, dealing with heterogeneous information from several independent sources. In those cases, selected non-standard reasoning tasks are more useful, as they provide a measure of semantic distance of the resource w.r.t. the request, along with justifications for (missed) subsumption and (un)satisfiability. In particular, the following three non-standard inference services, originally devised for semantic matchmaking scenarios [31] and optimized for mobile and embedded devices in *Tiny-ME* [36], are suitably combined to evaluate the type and weight of relations between pairs of arguments. Let $\mathcal{L}$ be a DL, $\mathcal{T}$ a set of axioms in $\mathcal{L}$, and R, S two concepts in $\mathcal{L}$ –a request and a resource, respectively– both satisfiable in $\mathcal{T}$. Then:

- *Concept Contraction* is computed if $\mathcal{T} \models R \sqcap S \sqsubseteq \bot$. It consists of finding a pair of concepts $\langle G, K \rangle \in \mathcal{L} \times \mathcal{L}$ such that $\mathcal{T} \models R \equiv G \sqcap K$, and $\mathcal{T} \models K \sqcap S \not\sqsubseteq \bot$. We name K a *contraction* of R according to S and $\mathcal{T}$. In other words, Contraction determines which part of R is conflicting with S. By retracting only conflicting requirements G (for *Give up*) in R, an expression K (for *Keep*) still remains, *i.e.*, a contracted version of the original request. The solution G to Contraction explains "why" the conjunction of R and S is not satisfiable;

- *Concept Abduction* is computed if $\mathcal{T} \models R \sqcap S \not\sqsubseteq \bot$ and $\mathcal{T} \models S \not\sqsubseteq R$. It consists of finding a concept $H \in \mathcal{L}$ such that $\mathcal{T} \models S \sqcap H \sqsubseteq R$ and $S \sqcap H$ is satisfiable in $\mathcal{T}$. The solution H (for *Hypothesis*) can be interpreted as what is requested in R and not specified in S, providing an explanation for missed Subsumption;
- *Concept Bonus* extracts a concept B from S, denoting what the resource provides even though the request did not ask for it. Formally, finding the Bonus B of S with respect to R is equivalent to find H in a Concept Abduction problem where R and S are swapped.

Concept Contraction derives from belief revision based on the classic AGM (Alchourrón - Gärdenfors - Makinson) paradigm [37]. Concept Abduction and Bonus rely on the Open World Assumption, stating that missing information is not equivalent to negation, but simply unspecified *e.g.*, due to it being unknown or irrelevant. For further discussions and examples, the reader is referred to [31].

Adopting a *normal form* for concept expressions allows defining a quantitative metric space with a *norm* operator $\|\cdot\|$. The proposed framework refers to $\mathcal{ALN}$ concept expressions in Conjunctive Normal Form (CNF) [31]. Any $\mathcal{ALN}$ concept expression C can be expressed in CNF as: $C \equiv C_{CN} \sqcap C_{\leq} \sqcap C_{\geq} \sqcap C_{\forall}$, where C_{CN} is the conjunction of all (possibly negated) atomic concepts, $C_{\leq}$ (respectively, $C_{\geq}$) is the conjunction of all maximum (resp. minimum) cardinality restrictions (at most one per role), and $C_{\forall}$ is the conjunction of all universal restrictions (at most one per role), with fillers recursively in CNF. The CNF norm –computed as sum of the number of concept components in the above 4 sub-expressions of each concept expression– of G and H then represents a semantic distance *penalty* for Concept Contraction and Abduction, respectively, used to assign a score to a resource w.r.t. a request; similarly, $\|B\|$ provides a measure of relevance for the Concept Bonus [31]. The computed concept expressions G, H and B act as logical *explanations* of the semantic distance and thus of the matchmaking outcomes.

3. Related work

Adapting the argumentation theory originated by Dung [18] to unpredictable pervasive computing scenarios is a currently relevant research trend.

In Social Internet of Things (SIoT) [38] contexts, various solutions have been proposed to model the information exchange among smart objects for an autonomous argumentation-based decision-making. In [13], the argumentation graph formalizes a natural language dialogue where speaking and hearing objects can argue with each other, so as to find an agreement on the perceived state of the affairs and to share decisions; two case studies on ambient assisted living and traffic management showcase the proposal. Likewise, in [39] social abstract argumentation theory is applied to solve conflicts for lane selection among smart vehicles on a congested road. By maintaining a reference objective, each agent collects environmental data and defines the projected next highway position it aims to occupy as an argument in the argumentation pool. The winning arguments –identified by processing the argumentation graph– output the lanes for every vehicle agent according to the current road configuration. Despite both the above works have shown the feasibility and usefulness of argumentation in IoT, formal modeling of arguments and automatic relation assessment have not been covered there.

A fuzzy argument-based classification scheme named CLeFAR has been proposed in [40] for fall detection in ambient-assisted living. By combining a fuzzy reasoning process and the Extended Social Abstract Argumentation (ESAA) framework, CLeFAR verifies whether the event class predicted by a common classifier is correct or not, thus outperforming basic classification schemes. Unfortunately, obtained ontology-based fuzzy arguments are not exploited to evaluate relations between them.

The argument acceptability in a graph is an essential element. Extension semantics label arguments with the classical accepted/rejected evaluation. [41] introduced two BAF semantics where attacks and supports can cancel each other in the acceptance decision, including one based on a majority evaluation between attacks and supports; however, for applications with a big number of arguments, this could still be too coarse. The approach in this paper exploits weighted relations –as determined via semantic matchmaking– to construct argumentation graphs that are amenable to gradual semantics. They produce finer assessments of the arguments, based on numerical scales [34]. It can be better suited to pervasive contexts, since even sub-optimal solutions can be considered useful due to unpredictable context changes and strict computational constraints of devices. The authors in [42] have systematically explored the application of ranking-based semantics to structured argumentation, highlighting how these semantics yield *culpability measures* and demonstrate robustness across various argument construction methods. In [35] a comparative study on ranking-based semantics has been carried out, underlining their differences in behavior and applicability. Based on a set of 13 well-defined principles that a semantics should satisfy in a bipolar setting, the comparison in [43] has shown extension semantics do not exploit support relations, as they declare all unattacked arguments as equally acceptable. In this paper, instead, depending on the strength propagation of attack and support arguments forming paths toward a target argument, the latter may be more or less acceptable than arguments which are both unattacked and unsupported. This is in agreement with multi-agent systems in pervasive computing, where information claimed by an agent needs confirmation or validation by other agents to be considered trustworthy [11]. In [43] the authors have proposed a novel *exponent-based* semantics for weighted bipolar argumentation, which satisfies all the principles and is not affected by the problem; however, it only deals with acyclic graphs. On the other hand, some ranking semantics that deal with cyclic weighted argumentative graphs (WAF) have been devised. Among them, [44] propose three novel semantics and provide a comprehensive comparison of the semantics that have been defined in the literature for evaluating arguments in weighted graphs. Unfortunately, the WAFs have some limitations and it is not possible to fully compare them with the BWAFs as (i) the attack is the only relation contemplated between pairs of nodes and (ii) the weight (called *basic score*) is preliminarily assigned to each argument and not to relations, as in the proposed framework. The work in [45] introduces a formalization of the notion of *compensation* in classical Dung's AF, whereby an argument receiving one strong attack is as good as an argument receiving several weak attacks. A strong attack is made by a non-attacked argument, whereas an attack is said weak if it comes from an argument that is attacked by non-attacked arguments. A large family of semantics allowing compensation, called α-BBS (for α burden-based semantics), is formalized.

The role of non-attacked arguments in AF has been also investigated in [46], defining six semantics based on the *propagation* of the weight of each argument to its neighbors. The propagation method encompasses two steps: (i) assigning an early weight to each argument so that the weight of unattacked arguments is greater than the attacked ones and (ii) iterative propagation of these weights into the graph changing their polarities in order to comply with the attack relation meaning (attack or defense). Propagation occurs with respect to weights defined randomly *a priori* for each argument. On the contrary, in our approach all arguments are assigned the same base score, then the strength propagation of paths is calculated by multiplying the weights of all relations in the path. Therefore it actually takes into account both the polarity and the intensity of each relation, which are determined via inferences on the DL concept descriptions representing the arguments. This makes the outcome of strength propagation on each path more interpretable and explainable.

In the extension [47] of the aforementioned work, the authors have introduced a new ranking semantics based on propagation, with an additional parameter to gradually decrease the impact of arguments when the length of the path between two arguments increases. This semantics satisfies two persuasion principles: *procatalepsis*, *i.e.*, it is often effective to anticipate counter-arguments from interlocutors, and *fading*, *i.e.*, long paths of argumentation become ineffective. However, even in that case, propagation takes place with respect to the weights associated to arguments in a preliminary step, by discriminating attacked arguments from unattacked ones.

Table 1 summarizes the main features of the proposed approach with respect to the foregoing state-of-the-art efforts. In summary, the semantics proposed in the present paper: (i) is suitable for BWAF, a generalization of classic AF that provides weighted attack and support relations; (ii) exploits the propagation of relation weights, accurately evaluated in an explainable semantic matchmaking process between pairs of annotations in $\mathcal{ALN}$ DL; (iii) promotes a more fine-grained argument ranking process that satisfies the fading property to model spatial and temporal distance of argument chains decreasing the impact on the acceptability of arguments; (iv) has a full prototype implementation, integrated with a complex scenario simulator for validation.

4. Explainable Bipolar Weighted Argumentation

Various extensions to Dung's argumentation framework have been proposed, aiming to enable more fine-grained modeling and evaluation of constellations of arguments with conflicting information. A key insight in some of these extensions is that not all arguments and their relations are equal. In many contexts, it is natural to assign inherent preference values to arguments and/or discriminate relations by their intensity, *e.g.*, by associating attacks with weights. Additionally, while Dung's AF only allows expressing support implicitly by means of defended arguments, in certain scenarios it would be beneficial to introduce a different relation type w.r.t. attacks. To meet these requirements, several generalizations of the basic AF have been introduced.

Among the various approaches aimed at improving the original AF introduced by Dung [18], the Bipolar Weighted AF (BWAF) [48] melds two other important AFs: the Bipolar AF (BAF) [30] and the Weighted AF (WAF) [49]. Unlike other AF approaches attaching a weight only to attack relations [50, 51] or BAFs using probabilities [52], BWAF allows

Table 1
Comparison of argumentation approaches

Reference	Lippi *et al.* [13]	Lovellette *et al.* [39]	Gulati and Kaur [40]	Bonzon *et al.* [47]	Pazienza *et al.* [48]	*Proposed Approach*
Context	Social IoT for smart home and traffic management	Lane selection in smart vehicles	Fall activity recognition in smart home	Fictional sales pitch persuading someone to buy a specific car	Debating system for social Web discussions	Coordination and decision-making in multi-agent systems
Main Goal	Agreement on perceived state and decision sharing	Conflicts resolution and actions assignment	Conflicts resolution and argumentation-based activity classification	Formalize the process of persuasion and evaluate existing semantics	Evaluate the most reliable claims in a debate	Integrate argumentation with explainable semantic matchmaking
Argumentation Framework	Bipolar Weighted AF	Social Abstract Argumentation	Extended Social Abstract Argumentation	Abstract Argumentation	Bipolar Weighted AF	Bipolar Weighted AF
Argumentation Semantics	Extension-based	Extension-based	Gradual semantics with fuzzy arguments	Acyclic propagation-based ranking with fading	Acyclic propagation-based ranking	Acyclic and cyclic propagation-based ranking with fading
Weight Assignment	Dynamic, according to the knowledge of previous situations	Not specified	Based on fuzzy membership functions	Randomly assigned; modified by propagation	Sentence similarity through word embeddings and sentiment polarity	Dynamic, based on semantic evaluation of arguments descriptions
Knowledge Representation	Not supported	Not supported	OWL ontology for event modeling	Not supported	Not supported	OWL ontology for argument modeling and non-standard inferences
Implementation and validation	Not available	NetLogo simulation	R prototype, integration with smart home simulator	Not available	Prototype applied to Reddit thread	C++ prototype, integration with open source StarCraft II bot
Key Advantages	Dialogue-based interaction of SIoT devices	Multi-agent negotiation to achieve individual objectives	Fuzzy argumentation enhancing accuracy of traditional classification models	Parameterized ranking-based semantics based on argument weight propagation	Ranking-based semantics based on relation strength propagation	Explainable graph structure and ranking outcomes, support for cyclic graphs
Key Limitations	Lack of formal modeling and automatic relation assessment	Conflicts only arise from agents with similar objectives	Static rule-based inference system for argument evaluation	Preliminary static weight assignment	Cyclic argumentative graphs not supported	Argumentative graph currently computed in a centralized way

two kinds of interactions between pairs of arguments: (i) weighted attacks and (ii) weighted supports, respectively identified by a negative or positive weight, whose absolute value indicates the relation strength.

Definition 4 (Bipolar Weighted Argumentation Framework). *A BWAF is a triple $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$, where $\mathcal{A}$ is a finite set of arguments, $\hat{\mathcal{R}} \subseteq \mathcal{A} \times \mathcal{A}$ is the set of relations and $w_{\hat{\mathcal{R}}} : \hat{\mathcal{R}} \mapsto [-1, 0[\cup]0, 1]$ is the weight function. Attack relations are defined as $\hat{\mathcal{R}}_{att} = \{\langle a, b\rangle \in \hat{\mathcal{R}} \mid w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \in [-1, 0[\ \}$ and support relations as $\hat{\mathcal{R}}_{sup} = \{\langle a, b\rangle \in \hat{\mathcal{R}} \mid w_{\hat{\mathcal{R}}}(\langle a, b\rangle) \in]0, 1]\ \}$.*

As depicted in Figure 2, a BWAF can be represented as a directed graph whose nodes are the arguments, relations define attacks (with solid edges from the attacker to the attacked node) or supports (with dashed edges), and weights label edges denoting the relation strength. In BWAFs, the notion of *defense* in Dung's AF can be generalized by leveraging the *change of sign* property of multiplication [48].

Definition 5 (Path and indirect attack/support). *Given arguments $a, x_1, x_2, \dots, x_n, b \in \mathcal{A}$, a path $p^* = \langle a, x_1, x_2, \dots, x_n, b\rangle$ from a to b exists iff $\langle a, x_1\rangle \in \hat{\mathcal{R}} \wedge \forall i = 2, 3, \dots n$: $\langle x_{i-1}, x_i\rangle \in \hat{\mathcal{R}} \wedge \langle x_n, b\rangle \in \hat{\mathcal{R}}$. Then:*

- *a* bw-attacks *b iff $w_{\hat{\mathcal{R}}}(\langle a, x_1\rangle) \cdot w_{\hat{\mathcal{R}}}(\langle x_1, x_2\rangle) \cdot \dots \cdot w_{\hat{\mathcal{R}}}(\langle x_n, b\rangle) < 0$.*
- *a* bw-supports *b iff $w_{\hat{\mathcal{R}}}(\langle a, x_1\rangle) \cdot w_{\hat{\mathcal{R}}}(\langle x_1, x_2\rangle) \cdot \dots \cdot w_{\hat{\mathcal{R}}}(\langle x_n, b\rangle) > 0$.*

Figure 2: BWAF legend: b_1 attacks a_1, b_2 supports a_2

The multiplication rule: (i) preserves the basic Dung-style property that an even-length path of attacks is a defense (*i.e.*, *the attack of an attack is a defense*); (ii) covers both BAF's notions of *indirect defeat* and *supported defeat* [30] with a single definition.

In the approach proposed here, basically oriented to the Semantic Web of Things, $\mathcal{A}$ consists of a set of knowledge fragments (*i.e.*, semantic annotations) shared by micro-devices about a reference domain. They are expressed as $\mathcal{ALN}$ concept descriptions, unfolded and normalized in Conjunctive Normal Form w.r.t. a scenario-dependent ontology $\mathcal{T}$, and they are shared following the network topology. The set of pairwise machine-to-machine interactions among micro-devices will coincide with the set $\hat{\mathcal{R}}$ of relations between the corresponding arguments. Basically, if a micro-device agent $\mathcal{A}_{\mathcal{S}}$ sends information to another micro-device agent $\mathcal{A}_{\mathcal{R}}$, the argumentative graph will contain two argument nodes for the semantic annotations S and R expressing the knowledge possessed by the two agents, and a relation edge from S to R.

The complete procedure that allows to transform a network of interacting agents into an explainable BWAF is outlined in Algorithm 1. Orientation, type and weight of the relations between each pair of agents are assessed as clarified hereafter. Considering two generic agents A_R and A_S in a (SWoT) scenario, the semantic annotations R and S express their knowledge with respect to a common ontology $\mathcal{T}$ and are treated as arguments. In this work it is assumed that each agent produces its own semantic annotation starting from data gathered from the environment through sensor devices. In order to do that and contrast inherent noise, data mining and semantics-enhanced techniques [10] can be applied; that problem is out of scope for this paper, as it has been investigated widely in the SWoT literature [1, 2, 9, 11]. If A_S interacts with A_R ($A_S \rightsquigarrow A_R$) in the smart agent network, in order to assess the degree of acceptability of R and S, the approach we propose considers R as *request* and S as *resource*, since in typical SWoT S can be used to "respond to needs" expressed by R. This means an argumentative relation can exist with edge orientation from S to R (lines 5–7).

Once the direction of a relation is defined, a *Satisfiability Check* is performed (line 9) to set its type: if $R \sqcap S$ is satisfiable w.r.t. $\mathcal{T}$, the relation that links S to R is a support, otherwise it is an attack; the weight of an attack (respectively, a support) is computed as per Formula 1 (resp. Formula 2). From this characterization, it is clear the proposed approach does not align with the *deductive* or the *necessary* interpretations of support, discussed by [53], as in general the non-acceptance of R (respectively, of S) does *not* imply the non-acceptance of S (resp., of R). Conversely, the approach complies with the *general* interpretation of support in Bipolar AFs, as [53] called the one introduced by [30].

In case of unsatisfiability, the computation of the attack weight must take into account the following informative contributions:

1. Clash (X): the amount of conflicting information between the arguments;
2. Confirmed (C): the amount of information expressed by both arguments;
3. Unconfirmed (U): the amount of information in the attacked argument which is neither confirmed nor rebutted by the attacker;
4. Extra (E): any amount of additional information of the attacker which is not present in the attacked argument.

As detailed in lines 10–16 of Algorithm 1, Concept Abduction, Contraction and Bonus (recalled in Section 2.2) [31] have been integrated in the following formula that includes each contribution as a signed term of an algebraic sum:

$$w_{\hat{\mathcal{R}},att}(\langle S, R\rangle) = -\alpha_X \frac{p_c(R,S)}{\|R\|} + \alpha_C \frac{\|Bonus(Bonus(K,S),S)\|}{\|R\|} + \alpha_U \frac{p_a(K,S)}{\|R\|} + \alpha_E \frac{\|Bonus(R,S)\|}{\|R\| \cdot \|S\|} \tag{1}$$

where p_c and p_a are the penalties computed as the CNF norms of concepts G and H respectively produced by Concept Contraction and Abduction [31]. The first term depends on $p_c(R,S) = ||G||$, where G is the *Give up* computed by Concept Contraction. Taking the *Keep* part K obtained from the same inference task, the second term is determined by means of two nested Concept Bonus inferences: the inner Bonus yields the additional information in S w.r.t. K, then the outer Bonus identifies what is “not additional”, *i.e.*, confirmed by K and S. The third term depends on $||H||$, the norm of the *Hypothesis* produced by Concept Abduction between K and S. The last term is computed with a (single) Concept Bonus between R and S. Weighing coefficients have been determined empirically in preliminary tests: $\alpha_X = 2$, $\alpha_C = 0.1$, $\alpha_U = 0.05$, and $\alpha_E = 0.05$. Since the total weight of an attack must be negative, the largest coefficient is associated to the first term in Formula 1, whilst the remaining three contributions are used to mitigate the attack strength, and therefore they have positive sign. The rationale for the selected values is to ensure the relationship weights fall in the desired range for general-purpose argumentation frameworks. Fine-tuning is possible *e.g.*, via grid or gradient-based search for specific domains and problem instances. The analysis of benefits vs computational costs trade-off of such optimizations is left for future work. Concepts G, K, H, and B –produced by Concept Contraction, Abduction and Bonus– provide a qualitative justification (a.k.a. explanation) of the quantitative evaluation of the relation: specifically, G justifies why (*i.e.*, what part of) S clashes with R; the other concepts mitigate the conflict, as K represents information in S not clashing with R, H represents information in R which is unspecified in S –in an Open World Assumption– and B is additional information in R which does not contradict S.

Conversely, if S supports R, two informative contributions must be considered in the support weight:

1. the amount of information missing in S to reach a full match with R;
2. the amount of information that S has in addition w.r.t. R.

Hence, the following formula is defined:

$$w_{\hat{\mathcal{R}},sup}(\langle S, R\rangle) = 1 - \frac{p_a(R,S)}{\|R\|}\left(1 - \frac{\|Bonus(R,S)\|}{\|S\|}\right) \tag{2}$$

The maximum support of S to R occurs when $\mathcal{T} \models S \sqsubseteq R \Leftrightarrow H \equiv \top \Rightarrow p_a(R,S) = 0$, whereas support is lowest when $H = R \Rightarrow \frac{p_a(R,S)}{\|R\|} = 1$. The last term of Equation 2 quantifies the additional information of S w.r.t. R through the Bonus inference (lines 18–21).

A complete example of argument relation evaluation using the proposed approach is reported in Section 6. At the end of Algorithm 1, the concept expressions of G, H and B can be returned to the end user as explanations of the relation evaluation result.

5. Propagation-based ranking semantics

In highly dynamic and unpredictable scenarios, a gradual information acceptability assessment can support collaborative autonomous decision-making. This section outlines the novel argumentative ranking semantics, based on path strength propagation in an interpretable BWAF.

Definition 6 (Strength Propagation). *Let $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}}\rangle$ be a BWAF and $x_1, x_n \in \mathcal{A}$ two arguments such that there exists a path $p^* = \langle x_1, x_2, \dots, x_n\rangle$. The* strength propagation (sp) *from x_1 to x_n for p^* is given by:*

$$sp\left(x_1, x_n\right)_{p^*} = \prod_{i=2}^{n} w_{\hat{\mathcal{R}}}\left(\langle x_{i-1}, x_i\rangle\right) \tag{3}$$

Basically, $sp\left(x_1, x_n\right)_{p^*}$ computes the strength of a path p^* from x_1 to x_n by multiplying the weight of its constituent relations. By definition, $sp(\cdot)$ is always in $[-1, 1]$ and returns a measure of the role x_1 is playing for x_n: a positive value means a *bw-supports* relation, otherwise a *bw-attacks* one. Each path between any two arguments has its own strength propagation.

In a BWAF, the possibility for an argument to be the target of a relation determines its type. This consideration leads to what follows.

Definition 7 (Connected and free arguments). *Let $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}}\rangle$ be a BWAF. The set of arguments $\mathcal{A}$ is partitioned in two disjoint subsets:*

- $\mathcal{A}^c = \{y \mid y \in \mathcal{A} \wedge \exists x \in \mathcal{A}$ *s.t.* $\langle x, y\rangle \in \hat{\mathcal{R}}\}$ *is the subset of arguments which receive at least one attack or support, denoted as* connected *arguments;*

ALGORITHM 1: Explainable BWAF construction

```
Require: Network of agents, Ontology $\mathcal{T}$ in $\mathcal{ALN}$
Ensure: Explainable Bipolar Weighted Argumentation Framework $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$
$\mathcal{A} \leftarrow$ Set of $\mathcal{ALN}$ concept descriptions generated by network agents
$\hat{\mathcal{R}} \leftarrow \emptyset$ {Initialize the set of relations}
$w_{\hat{\mathcal{R}}} \leftarrow \emptyset$ {Initialize the weight function}
for all pairs of agents $(A_S, A_R)$ in the network do
    {Step 1: Relation direction}
    if $A_S$ interacts with $A_R$ $(A_S \rightsquigarrow A_R)$ then
        Add $(S, R)$ to $\hat{\mathcal{R}}$ {Relation direction $:= (S \rightarrow R)$}
        {Step 2: Relation type and weight}
        if $\mathcal{T} \models S \sqcap R \sqsubseteq \bot$ then
            {Relation type := "Attack"; compute formula (1)}
            $p_c \leftarrow$ ConceptContractionPenalty($R$, $S$)
            $K \leftarrow$ ConceptContractionKeep($R$, $S$)
            BonusTerm1 $\leftarrow$ Bonus(Bonus($K$, $S$), $S$)
            $p_a \leftarrow$ ConceptAbductionPenalty($K$, $S$)
            BonusTerm2 $\leftarrow$ Bonus($R$, $S$)
            Weight $\leftarrow -\alpha_X \frac{p_c}{\|R\|} + \alpha_C \frac{\|BonusTerm1\|}{\|R\|} + \alpha_U \frac{p_a}{\|R\|} + \alpha_E \frac{\|BonusTerm2\|}{\|R\| \cdot \|S\|}$
        else
            {Relation type := "Support"; compute formula (2)}
            $p_a \leftarrow$ ConceptAbductionPenalty($R$, $S$)
            $BonusTerm \leftarrow$ Bonus($R$, $S$)
            Weight $\leftarrow 1 - \frac{p_a}{\|R\|}\left(1 - \frac{\|BonusTerm\|}{\|S\|}\right)$
        end if
        Add $w_{\hat{\mathcal{R}}}(S, R) \leftarrow$ Weight
    end if
end for
return $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$
```

- $\mathcal{A}^f = \mathcal{A} \setminus \mathcal{A}^c$ *is the subset of arguments in* $\mathcal{A}$ *which are not attacked or supported by any other argument, hereinafter called* free *arguments.*

The proposed ranking semantics is based on the *fading* principle [47], stating that the *length* of a chain of arguments should be inversely proportional to the *impact* of that chain on the final argument. The idea of fading derives from the observation that, in common dialogue, longer chains of arguments become less effective, because short-term memory is limited and people easily lose track of far-reaching consequences of arguments and relations. In multi-agent systems, the spatial and temporal distance from an agent decrease the influence of environmental events and of other agents' actions on it [54]. Since in the proposed approach argument chains represent sequences of interactions or information exchanges among agents, fading models this phenomenon in a straightforward way. Furthermore, since the majority of pervasive devices have constrained memory and storage space, the fading effect provides useful benefits, therefore agents should weigh less or "forget" those arguments which are farther in time and/or space. The proposed path strength propagation via multiplication of relation weights achieves this property rather naturally: regardless of polarity, the absolute value of each weight is in $]0, 1]$ and therefore the overall intensity can never increase and will likely decrease for longer paths.

In the proposed BWAF ranking semantics, the acceptability assessment of arguments is an iterative process, outlined in Algorithm 2. It starts after arguments in the graph have been annotated with DL descriptions and relation edges have been directed and weighed as explained in Section 4. Its outcome consists of a ranking, associating each argument $\theta \in \mathcal{A}$ with an acceptability score (a real number in the]..,1,..[interval), where 1 stands for neutrality, and higher (respectively, lower) values proportionally label acceptable (resp. unacceptable) arguments. An example of the algorithm execution is discussed later in Section 6. At each step $i > 0$, the procedure identifies all paths of length i ending in the argument under evaluation.

ALGORITHM 2: Propagation-based Ranking Semantics

```
Require: BWAF $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$, $\zeta \in ]0, 1[$, $\epsilon \in \mathbb{R}^+$, preliminary computation of $\mathcal{A}^f$, $\mathcal{A}^c$
Ensure: Ranking of arguments $R = \{\langle a_i, s_i \rangle\}, a_i \in \mathcal{A}, s_i \in \mathbb{R}$
 1: $R := \emptyset$
 2: for all $a_i \in \mathcal{A}$ do
 3:     if $a_i \in \mathcal{A}^c$ then
 4:         $flagStop_{counter}[a_i] := 0$
 5:     end if
 6:     $s_i := 1.0$
 7:     add $\langle a_i, s_i \rangle$ to $R$
 8: end for
 9: $fastStop := false$
10: for $j := 1, fastStop \neq true, j++$ do
11:     for all $a_i \in \mathcal{A}^c$ do
12:         $X := Y := 0$
13:         $\langle \mathcal{P}_j^f(a_i), \mathcal{P}_j^c(a_i) \rangle := retrievePaths(a_i, j)$ retrieve all free-born and connected-born paths with lenght $j$, ending in $a_i$, and satisfying constraints (C1) and (C2)
14:         if $(\mathcal{P}_j^f(a_i) \neq \emptyset)$ then
15:             $X = \frac{\zeta^j}{|\mathcal{P}_j^f(a_i)|} \sum_{p \in \mathcal{P}_j^f(a_i)} sp(\alpha, a_i)$
16:         end if
17:         if $(\mathcal{P}_j^c(a_i) \neq \emptyset)$ then
18:             $Y = \frac{(1-\zeta)^j}{|\mathcal{P}_j^c(a_i)|} \sum_{p \in \mathcal{P}_j^c(a_i)} sp(\alpha, a_i)$
19:         end if
20:         $rank_{upd} := s_i + X + Y$
21:         update $\langle a_i, s_i := rank_{upd} \rangle$ in $R$
22:         if $(|rank_{upd} - s_i| < \epsilon$ then
23:             $flagStop_{counter}[a_i]++$
24:             if $(flagStop_{counter}[a_i] == 2)$ then
25:                 remove $a_i$ from $\mathcal{A}^c$
26:             end if
27:         else
28:             $flagStop_{counter}[a_i] := 0$
29:         end if
30:         if $(\mathcal{A}^c == \emptyset)$ then
31:             $fastStop := true$
32:         end if
33:     end for
34: end for
35: return $R$
```

Definition 8 (Path length). *Let $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ be a BWAF and $x_1, x_n \in \mathcal{A}$ two arguments s.t. there exists a path $p^* = \langle x_1, x_2, \dots, x_n \rangle$. The length i of the path p^* is the number $n - 1$ of relations $\langle x_{j-1}, x_j \rangle$, where $j = 2, \dots, n$.*

As per Definition 7, the kind of source node of a path defines its type; consequently, Algorithm 2 evaluates the different contributions separately in computing the acceptability score (lines 14–19).

Definition 9 (Free-born and connected-born paths). *Let $\mathcal{G} = \langle \mathcal{A}, \hat{\mathcal{R}}, w_{\hat{\mathcal{R}}} \rangle$ be a BWAF and $\alpha, \theta \in \mathcal{A}$ two arguments. Any path $p^* = \langle \alpha, \dots, \theta \rangle$ of any length i ending in θ belongs to one of the following two disjoint sets:*

- *Free-born paths $\mathcal{P}_i^f(\theta)$ starting in an argument $\alpha \in \mathcal{A}^f$*
- *Connected-born paths $\mathcal{P}_i^c(\theta)$ starting in an argument $\alpha \in \mathcal{A}^c$*

The iterative ranking procedure initially assumes neutral acceptability for all arguments: $\forall \theta \in \mathcal{A} : s_0(\theta) = 1$ (lines 6–7 of Algorithm 2). This is also the final score for free arguments, while for each connected argument $\theta \in \mathcal{A}^c$

at each step $i > 0$, lines 11–34 of Algorithm 2 compute:

$$s_i(\theta) = s_{i-1}(\theta) + X + Y \tag{4}$$

where:

$$X = \begin{cases} \frac{\zeta^i}{|\mathcal{P}_i^f(\theta)|} \sum_{p \in \mathcal{P}_i^f(\theta)} sp(\alpha_p, \theta) & if\ \mathcal{P}_i^f(\theta) \neq \emptyset \\ 0 & otherwise \end{cases}$$

$$Y = \begin{cases} \frac{(1-\zeta)^i}{|\mathcal{P}_i^c(\theta)|} \sum_{p \in \mathcal{P}_i^c(\theta)} sp(\alpha_p, \theta) & if\ \mathcal{P}_i^c(\theta) \neq \emptyset \\ 0 & otherwise \end{cases}$$

with $sp(\alpha_p, \theta)$ evaluated for the path starting in argument α_p and ending in argument θ, as per Equation 3, and $0 < \zeta < 1$ a coefficient (empirically set to 0.8 after preliminary tests) having two purposes: satisfy the fading principle (by means of the exponential) and give greater weight to free-born than connected-born paths. The latter idea is justified by the *void precedence* property in ranking semantics [34], stating that non-attacked arguments should have a higher acceptability rank than attacked ones, and also bears some similarity to the *evidential* interpretation of support [55], as free-born arguments include *prima facie* arguments as defined in that context. The selected value for ζ appears as adequate for general-purpose argumentation frameworks, while systematic and domain-oriented fine-tuning is left for future work.

In order to ensure termination and make the algorithm less computationally expensive, two constraints are imposed in identifying the paths (with increasing length i), suitable to update the acceptability score of an argument θ. For each step $i \in \mathbb{N}$, $i > 0$, any path $p^* = \langle x_1, x_2, \dots, x_i, x_{i+1} = \theta \rangle$ with length i is admissible for computation only if both the following constraints are satisfied:

(C1) the source argument x_1 and any other argument x_k, with $k = 2, 3, \dots, i$ are different from the argument $x_{i+1} = \theta$ whose acceptability is being evaluated;

(C2) for each argument a repeated $n \geq 3$ times in the path p^*, all the $n - 1$ sequences of arguments included between pairs of occurrences of a are different from each other.

Constraint $(C1)$ is imposed because (i) the sp of any path originating from an argument θ must not impact the computation of the acceptability rank of the same argument θ, and, (ii) if the same θ is in a position $k < i + 1$ of path p^*, it is plausible that the contribution of the path to the rank of θ has already been considered in one of the previous iterations. Constraint $(C2)$ prevents infinite loops, instead, by discarding paths with repeated sub-paths. These constraints on the iterative process for paths on increasing length enable the ranking semantics to cope with graphs containing highly complex and cyclic argument relations.

In addition to path admissibility, in order to prevent non-termination, lines 22–29 of Algorithm 2 further check the acceptability score. In fact, if the absolute value of the difference between $s_{i+1}(\theta)$ and $s_i(\theta)$ is less than a constant ϵ for two consecutive iterations, score convergence is accepted and the procedure stops. Preliminary tests have found the value $\epsilon = 0.02$ consistent with this purpose. In Algorithm 2, following a preliminary initialization in lines 3–5, structure $flagStop_{counter}$ keeps for each connected argument how many times the absolute value of the difference between two consecutive scores is less than ϵ. If for a connected argument a_i the value $flagStop_{counter}[a_i]$ is equal to 2, the argument a_i is removed from the set A^c, since the convergence of the acceptability score has been reached and no iteration is needed anymore (lines 24–26).

It should be noted that, in this approach, the number of required iterations to compute the acceptability score is not the same for all arguments in a BWAF: the early score for free arguments is 1 and no further calculations are needed; conversely, for each connected argument the number of iterations depends on the number, length and sp of the paths ending in it.

Finally, the proposed semantics induces a gradual acceptability ranking of arguments in a BWAF.

Property 1. *The propagation-based semantics $\mathcal{S}(\cdot)$ associates a ranking $\succeq_{\mathcal{G}}$ on $\mathcal{A}$ to any BWAF $\mathcal{G}$, such that $\forall\, a, b \in \mathcal{A}$, $a \succeq_{\mathcal{G}}^{\mathcal{S}} b$ iff $\mathcal{S}(a) \geq \mathcal{S}(b)$.*

Overall, the ranking procedure ensures a gradual assessment of the acceptability of the arguments in a BWAF. Main features of the adoption of this semantics can be summarized as:

- the ability to manage bipolar weighted graphs with cycles and paths of any length;
- completeness of the algorithm thanks to path admissibility constraints and convergence condition;
- persistence of the value 1 as acceptability threshold as well as the score of free arguments;
- compliance with the fading principle.

Furthermore, the proposed ranking semantics satisfies the following postulates defined by [34]:

- *abstraction*: for every pair of arguments, the acceptability ranking is preserved when applying isomorphisms to the argumentative graph;
- *independence*: for every pair of arguments, the acceptability ranking is preserved after the fusion of two graphs;

Other postulates defined for ranking semantics in AA, such as *void precedence* and *defense precedence* [34] (if two arguments A and B receive the same amount of attacks, A receives no defense and B receives at least one defense, then B will have higher acceptability than A) do not always hold in a BWAF, as they depend on the relation strengths. Compared to *Discussion-based semantics* (Dbs) [34], which basically counts the number of linear discussions (LDs) of an argument as the LD length increases, the proposed approach computes the weighted contributions *sp* of the individual paths at each iteration and discriminates free-born from connected-born paths. The *Burden-based semantics* (Bbs) [34] elaborates the ranking by comparing the arguments lexicographically, based on their burden numbers, and the update of the burden numbers occurs considering all the attack relationships between arguments in an AF as equivalent. Unlike the semantics presented here, *«Dbs and Bbs rank the arguments only on the basis of the structure obtained by "unrolling" the cycles. For example, [they] do not distinguish between a loop (e.g. aRa) and a cycle (e.g. aRb, bRa).»*

The above framework is general-purpose, but the devised properties are particularly fit for machine-to-machine interactions in pervasive systems and applications.

6. Illustrative example: argumentative StarCraft II agents

An early validation of the proposed framework has been carried out in a case study concerning the real-time strategy (RTS) game *Starcraft II* (SC2). RTS games are, in fact, MAS simulators with high unpredictability and context dependency: these peculiarities frequently characterize pervasive contexts. The agent (player) is constantly interacting with the environment (game), receiving stimuli and performing actions.

A prototype of argumentative deductive SC2 player agent has been developed in C++ by extending the *CommandCenter*[1] open source SC2 bot. The agent's main loop consists of:

1. real-time data gathering from the game engine using the *StarCraft II (SC2) API*[2];
2. annotation of data concerning the game status with respect to a reference ontology; descriptions are arranged in a BWAF as explained in Section 4 and exemplified hereafter;
3. argument acceptability ranking through the framework in Section 5 to decide the agent's current strategy;
4. strategy actuation issuing a sequence of game commands to the SC2 API.

The SC2 API is a WebSocket-based wrapper over a Protocol Buffers[3] interface for reading and controlling in-game objects.

In SC2 each player commands an army comprising various kinds of units. *Health* is the essential parameter of a unit. Receiving hits decreases health, and when it reaches 0 the unit is destroyed. The goal of the game is to eliminate all the opponent's units, hence it is essential for a player to determine the most useful attack or defense strategy. Figure 3 depicts an example of match between allied units (*Player 1*), controlled by the argumentative deductive reasoning agent, and enemy units (*Player 2*). It is described below.

In a battle scenario where the allied and enemy units are equal in number and formation, 3 *Zergling enemy units are exerting a joint assault on the unit A. All* 3 *Marine units of Player* 1 *have maximum health, whereas unit E of Player* 2 *has low health.*

---

[1] CommandCenter: `https://github.com/cpp-sc2/commandcenter`

[2] StarCraft II API: `https://github.com/Blizzard/s2client-proto`

[3] Protocol Buffers: `https://developers.google.com/protocol-buffers/`

Figure 3: Case study scenario

The case study would show how argumentative deductive reasoning allows selecting the most advantageous strategy in a given moment of the game. Each possible strategy is described through a semantic annotation providing information about both its type and game conditions which make it suitable. The SC2 game state features taken into account include: the types and the number of units in the two groups; differences in assault range, movement speed, and assault strength between the two brigades; the type of assault suffered by the player agent's army; the health level of the allied and enemy units. As an illustrative example, let us consider the following five strategies:

1. *Healthiest Assault* (*HA*): assault the healthiest enemy unit;
2. *Nearest Assault* (NA): assault the nearest enemy unit;
3. *Least Healthy Assault* (*LHA*): assault the enemy unit with the lowest health level;
4. *Least Healthy Defense* (*LHD*): protect the agent's own lowest-health unit, by adopting a compact formation and counterattacking the nearest enemy unit which is assaulting it;
5. *Kiting* (K): defend by distancing enemy units, as the assault range of the agent's units is longer, so they can hit without retaliation.

The above battle plans are grouped in two families: the first three strategies are *Assaults* and the remaining ones are *Defenses*. This hierarchical strategy model is adopted to enable dividing the decision-making process in two subsequent steps: first determine the most promising family, and subsequently select the most suitable strategy among those in the selected category. This way decreases the number of alternatives to assess in each step, thereby reducing the size of the argumentation graph and possibly leading to computational cost savings.

The case study grounds on a KB modeled in $\mathcal{ALN}$ DL. The Terminological Box (*TBox*) $\mathcal{T}$ contains classes of game state features the player agent can sense or act upon, through the API. Figure 4 reports a significant excerpt of $\mathcal{T}$. For the sake of simplicity, the example focuses on units numerousness, health of ally and enemy units, and assault range. They are information fragments gathered from the SC2 API calls equivalent *e.g.*, to micro-device readings in SWoT scenarios. Each available strategy and relevant game feature is modeled as an individual of the Assertion Box

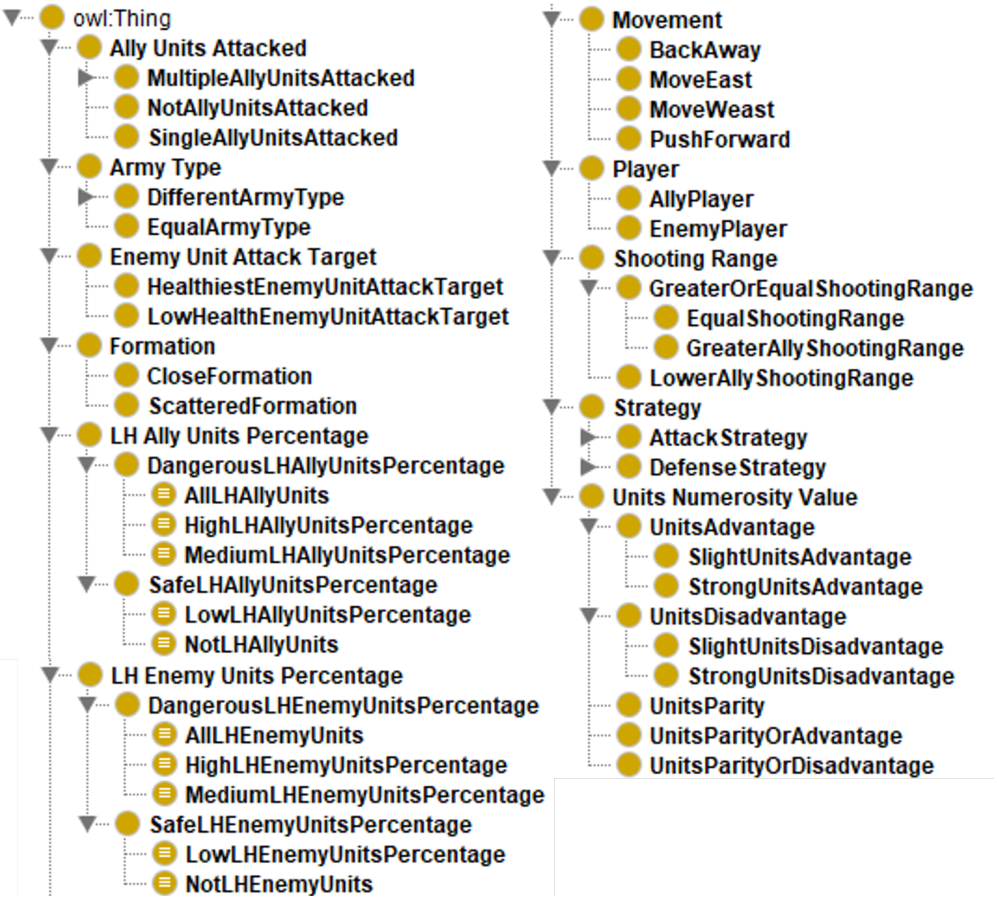


Figure 4: StarCraft II Ontology

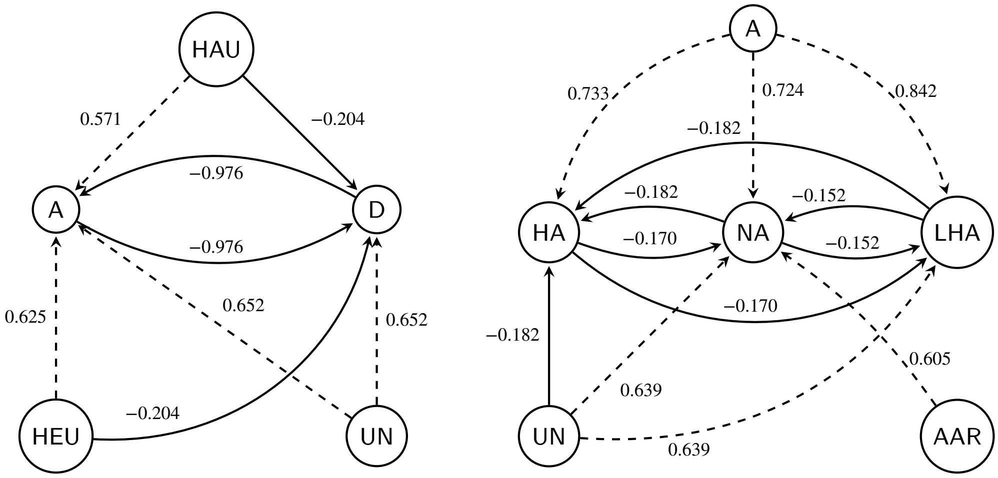
(*ABox*) and is used as a single argument in the BWAF. The Tiny-ME reasoner is leveraged to perform the appraisal of the type and weight of interactions between pairs of arguments as described in Section 4.

The BWAF graph $\mathcal{G}_1$ in Figure 5a reflects the identification of the most appropriate group of strategies. The node describing the *Assault* (A) category is supported by: (i) *Units Numerosity* (UN), as parity between allied and enemy units allows an assault; (ii) *Health Ally Units* (HAU), as no ally units have low health; and (iii) *Health Enemy Units* (HEU), as 33% on average of enemy units have low health. In the same way, the *Defense* (D) strategy argument is attacked by HAU and HEU, as a protective battle plan is not beneficial if allied units are mostly healthy and the enemy army has a medium percentage of low-health units. Conversely, *Defense* is supported by the UN due to unit parity between brigades. The two categories of strategies are in a mutual attack relationship.

To clarify and illustrate how the above graph is built, an example referred to the appraisal of the relation of HAU w.r.t. D is given. Concept expressions representing arguments are reported in OWL 2 *Manchester Syntax* [56].

```
HAU: AllyPlayer and (hasLowHealthAllyUnits some owl:Thing) and
(hasLowHealthAllyUnits only NoLHAllyUnits)
D: DefenseStrategy and (hasLowHealthAllyUnits some owl:Thing) and
(hasLowHealthAllyUnits only DangerousLHAllyUnitsPercentage) and
(hasLowHealthEnemyUnits some owl:Thing) and (hasLowHealthEnemyUnits only SafeLHEnemyUnitsPercentage)
and (hasUnitsNumerousness some owl:Thing) and (hasUnitsNumerousness only UnitsParityOrDisadvantage)
```

D and HAU respectively play the roles of request and resource in the relation appraisal method described in Section



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$

Figure 5: BWAF of first ($\mathcal{G}_1$) and second ($\mathcal{G}_2$) reasoning steps

4. The preliminary consistency check fails, as the conjunction of the two expressions is unsatisfiable due to the clash of disjoint classes `DangerousLHAllyUnitsPercentage` and `NoLHAllyUnits`. The relation is thus recognized as an attack and its weight must be computed with Equation 1. Concept Contraction of D w.r.t. HAU and $\mathcal{T}$ returns:

```
G := (hasLowHealthAllyUnits only DangerousLHAllyUnitsPercentage)
K := DefenseStrategy and (hasLowHealthAllyUnits some owl:Thing) and
(hasLowHealthAllyUnits only LHAllyUnitsPercentage) and (hasLowHealthEnemyUnits some owl:Thing) and
(hasLowHealthEnemyUnits only SafeLHEnemyUnitsPercentage) and (hasUnitsNumerousness some owl:Thing) and
(hasUnitsNumerousness only UnitsParityOrDisadvantage)
```

and $-\frac{p_c(D,HAU)}{||D||} = -\frac{||G||}{||D||} = -0.125$. This value expresses the strength of conflict between the arguments. The remaining terms in Equation 1, taken with positive sign, mitigate the attack. The second term, quantifying the common elements between the two arguments, is computed as:

```
Bonus(K,HAU) := AllyPlayer and (hasLowHealthAllyUnits some owl:Thing) and (hasLowHealthAllyUnits only
SafeLHAllyUnitsPercentage)
BB := Bonus(Bonus(K,HAU),HAU) := (hasLowHealthAllyUnits some owl:Thing)
```

and $\frac{||BB||}{||D||} = 0.063$. The third term of the formula reflects the information in the attacked argument D that is missing in the attacking one HAU, computed as Hypothesis of the Concept Abduction between K and HAU:

```
H := DefenseStrategy and (hasLowHealthEnemyUnits some owl:Thing) and
(hasLowHealthEnemyUnits only SafeLHEnemyUnitsPercentage) and (hasUnitsNumerousness some owl:Thing) and
(hasUnitsNumerousness only UnitsParityOrDisadvantage)
```

hence $\frac{p_a(K,HAU)}{||D||} = \frac{||H||}{||D||} = 0.75$. Finally, the information that the attacking argument HAU has in addition w.r.t. the attacked one D is determined through:

```
B := Bonus(D,HAU) := AllyPlayer and (hasLowHealthAllyUnits only SafeLHAllyUnitsPercentage)
```

with $\frac{||B||}{||D||\cdot||HAU||} = 0.045$. The weighted algebraic sum of the above terms as per Equation 1 is -0.204, which labels the edge from HAU to D in the BWAF $\mathcal{G}_1$ in Figure 5a. Concept expressions G, BB, H, and B provide the end users

Table 2
Acceptability Ranking on $\mathcal{G}_1$

Step i	*s(HAU)*	*s(A)*	*s(D)*	*s(HEU)*	*s(UN)*
0	**1**	**1**	**1**	**1**	**1**
1		1.296	0.870		
2		**1.247**	**0.485**		

Table 3
Acceptability Ranking on $\mathcal{G}_2$

Step i	*s(A)*	*s(HA)*	*s(NA)*	*s(LHA)*	*s(UN)*	*s(AAR)*
0	**1**	**1**	**1**	**1**	**1**	**1**
1		1.184	1.493	1.560		
2		1.104	1.443	1.511		
3		**1.114**	**1.450**	**1.519**		

and stakeholders of the system with a transparent and formal explanation which makes the result interpretable and offer a qualitative assessment of the relation between the pair of arguments. Following the evaluation of the type and weight of each argumentative relation, the acceptability of the arguments is measured by determining the ranking.

Table 2 shows the acceptability scores for each argument in $\mathcal{G}_1$, as computed via Algorithm 2; shaded cells indicate the final values. After the initialization ($i = 0$) for all arguments, no subsequent iteration is needed for *s(HAU)*, *s(HEU)* and *s(UN)*, as they are free arguments. Instead, the processing goes on for A and D. For example, let us focus on the argument A: as per Algorithm 2, for $i = 1$ paths ending in A are: $\langle HAU, A\rangle$, $\langle HEU, A\rangle$, $\langle UN, A\rangle$ and $\langle D, A\rangle$. The first, second and third path belong to $\mathcal{P}_1^f(A)$, while the last one is in $\mathcal{P}_1^c(A)$. The sp values of the four paths (as arranged in Equation 4) amount to $s_1(A) = 1.296$. Subsequently for $i = 2$ the paths ending in A are: $\langle HAU, D, A\rangle$, $\langle HEU, D, A\rangle$ and $\langle UN, D, A\rangle$, which amount to $s_2(A) = 1.247$. The iteration for calculating the acceptability score of A ends since at $i = 3$ all paths violate the two constraints of admissibility.

The execution of the proposed ranking semantics for $\mathcal{G}_1$ produces the following results:
$s(\mathtt{A}) = 1.247 > s(\mathtt{HAU}) = s(\mathtt{HEU}) = s(\mathtt{UN}) = 1.0 > s(\mathtt{D}) = 0.485$.

The *Assault* argument wins the argumentative dispute: in the current state of the game an offensive strategy is to be preferred. This enables the second reasoning stage to identify the most proper battle strategy in this family.

The BWAF graph $\mathcal{G}_2$ in Figure 5b illustrates the second reasoning step. The three strategies *Healthiest Assault* (*HA*), *Nearest Assault* (*NA*) and *Least Healthy Assault* (*LHA*) are all supported by their family *Assault* and are in a mutual attack relation, since only one can be selected. *HA* is attacked by *UN*, as it is advantageous only when the number of allied units is significantly greater than enemy ones. Conversely, *NA* and *LHA* are supported by *UN*, as they are adoptable also in case of parity between allied and enemy units. Furthermore, the feature *Ally Assault Range* (*AAR*), characterized by equal assault range among the units of the players, supports the node *NA*, which requires an allied range greater than or equal to the enemy one.

Table 3 reports on the ranking iteration for arguments in $\mathcal{G}_2$. The final outcome is:
$s(\mathtt{LHA}) = 1.519 > s(\mathtt{NA}) = 1.450 > s(\mathtt{HA}) = 1.114 > s(\mathtt{A}) = s(\mathtt{UN}) = s(\mathtt{AAR}) = 1.0$.

All the above strategies have a rank > 1 and are judged as acceptable in the particular battle situation. However, the proposed ranking semantics allows to arrange the arguments (and particularly the strategies) w.r.t. reliability and effectiveness in a given game context. Selecting the *Least Healthy Assault* among the available *Assault* strategies appears consistent with the scenario, as it allows Player 1 to quickly eliminate the least health opponent unit and outnumber the enemy.

The player agent repeats the foregoing main event loop (data gathering - annotation - argumentative deductive strategy selection - strategy actuation) continuously. It may be expected that in consecutive frames the state of the game will not vary by much, and therefore arguments and their relations could be cached and reused to save computations. Like in real pervasive computing environments, however, information gathered on the SC2 game state can be volatile and noisy, undergoing unpredictable updates, as events (like units moving on the battlefield) are discernible only within the intersection of the game viewport and the *fog of war*, a typical feature of RTS games which limits detectable updates

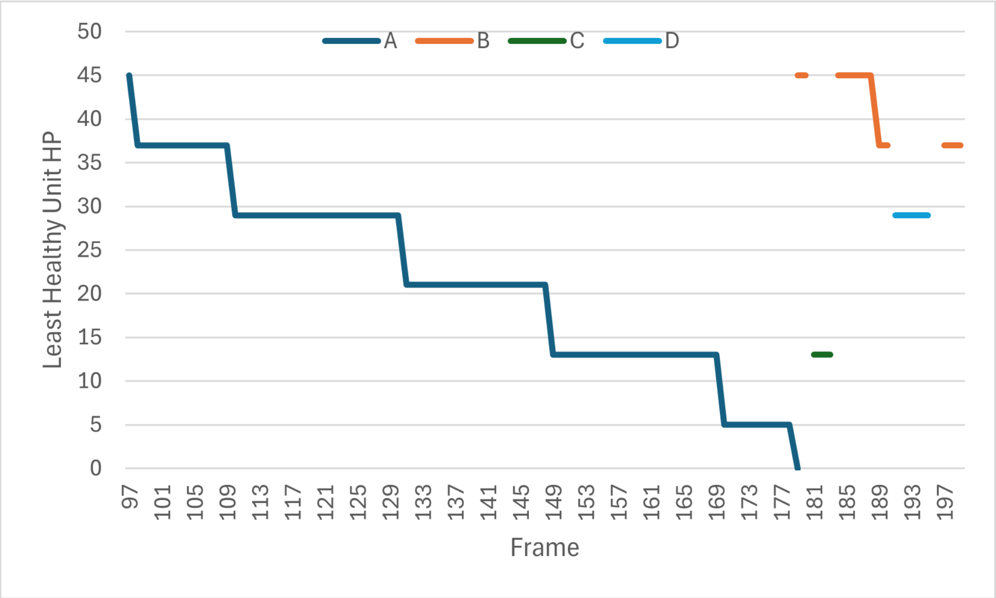


Figure 6: Impact of unit volatility on Least Healthy Assault strategy application

to areas in direct visibility range of allied units. As an example, Figure 6 plots data extracted from a game log saved by the deductive argumentative agent. Starting from frame no. 97, agent adopts the LHA strategy and targets enemy unit A as the one with the minimum value of *hit points* (HP): assaults on this unit reduce its HP amount until frame 179, when it reaches 0 and is destroyed. The agent selects the LHA strategy again with enemy unit B as the new minimum-HP target, but just 2 frames later enemy unit C appears in the visibility area of the agent, which selects it as the new target. Merely 3 frames later, C disappears –it is likely moving near the fog-of-war border– and the agent switches back to B. After a few more frames –and a successful assault reducing B's health– another enemy unit D becomes visible and is selected as new target, having a lower health amount. Also in this case, however, D disappears after a few frames and the agent reverts to targeting B. Despite the data noisiness and event volatility, inspection of the game session log confirms that the LHA strategy selection by means of the proposed argumentation framework is always appropriate.

The current implemented prototype, while useful to validate the proposed argumentative deductive approach in a relevant simulated environment, has still several constraints preventing its exploitation as a complete autonomous SC2 player. First, the data gathering from the game engine and the ontology used for data annotation cover only on the skirmish aspects of a SC2 battle. Both the ontology in Figure 4 and the data gathering code module should be expanded in order to include the economy management aspects of the game, such as resource harvesting and construction of bases and units. At the same time, the KB of strategies should be expanded to include more situations and possible behaviors, and each additional strategy may require new operations in the strategy actuation code module. In all cases, the required extensions can be deemed as rather straightforward development work starting from the current prototype.

7. Experiments

This section presents early experimental results in order to assess both effectiveness and scalability of the proposed approach. Unfortunately, it is not possible to directly compare the proposed ranking semantics with approaches discussed in Section 3, since their code is not available, to the best of our knowledge. Moreover, datasets in literature have only acyclic argumentative graphs, disregarding one of the main benefits of the proposed approach. Anyway, tests have been divided in three groups: the first one has just evaluated the efficacy of the deductive argumentative agent integrated in SC2, by recording performance metrics while executing the case study example. Subsequently, the proposed ranking semantics has been tested on random bipolar weighted graph configurations of increasing size and complexity. Finally, the new propagation-based semantics has been also tested on an available dataset with three well-known random graph models, evaluating execution times and the number of iterations. The first two test campaigns have been run on a desktop PC equipped with Intel Core i7-4790 CPU (3.6 GHz), 16 GB RAM, and Windows 10

Table 4
Performance results on $\mathcal{G}_1$ and $\mathcal{G}_2$

	Vertexes	Edges	T_1 (ms)	T_2 (ms)	Iterations
$\mathcal{G}_1$	5	8	1	1	2
$\mathcal{G}_2$	6	13	2.24	4.25	3

(64 bit) operating system. Instead, the last set of tests has been performed on a desktop PC equipped with Intel Core i9-10920X CPU (3.5 GHz), 32 GB RAM, and Windows 10 (64 bit) operating system.

SC2 tests. A complete loop for the agent involves the preliminary construction of the graph and assessment of argument relations, before the ranking calculation phase. The first performance evaluation of the proposed approach has been carried out on the $\mathcal{G}_1$ and $\mathcal{G}_2$ graphs from Figure 5. Table 4 shows the main graph characteristics and the obtained results. Time results are the average of the last 4 out of 5 consecutive runs. Besides the number of vertices and edges in the graphs, columns T_1 and T_2 represent the average time to construct the explainable BWAF graph by evaluating all argument relations –as described in Section 4– and to execute the propagation-based ranking semantics (Algorithm 2), respectively. Finally, *Iterations* represents the number of iterations of the algorithm needed to compute the acceptability of all arguments. The graph construction (including relation evaluation) took under 2.5 ms in both cases. Hence a fully-comprehensive strategy evaluation round, comprising two hierarchical stages with the above graphs, would take nearly 10 ms on average, which allows for real-time strategy update against fast-changing context conditions in an RTS game or in a typical pervasive scenario. As reported in [57], SC2 updates the simulation 16 or 22.4 times per second, depending on game speed settings, thus leaving 62.5 or 44.6 ms, respectively, for the agent's event loop. Computation time results for the proposed algorithm appear as compatible with the above constraints and therefore they can be deemed as satisfactory. The short duration of tests has not allowed reliable memory consumption measurement.

The time spent for the construction of the graph is lower than the ranking time, since the leveraged semantic matchmaking inference procedures are strongly optimized for a reduced consumption of time and memory [36]. This allows balancing the explainability benefits of the proposed semantics-based approach with performance challenges. Moreover, for larger graphs the latter phase could tend to dominate processing time, since the number of paths to be evaluated in the ranking semantics grows more than linearly w.r.t. the number of graph edges (as further elaborated in Section 7.1).

Random graph tests. In order to provide a more robust performance analysis, the second test group has focused on the propagation ranking semantics evaluation algorithm described in Section 5: for this purpose the *Boost Graph Library* (BGL)[4] has been used to build random bipolar weighted graphs. BGL is a standardized generic interface for composing and browsing graphs that exposes their structure while concealing implementation information. *Small* (S), *Medium* (M), and *Large* (L) scale configurations have been generated to measure turnaround time and memory consumption peak, both sensitive metrics in pervasive computing contexts. The provided configurations depend on three parameters:

- number of nodes in the graph: 10, 50, and 100, respectively for S, M and L;
- number of weighted edges: 15 and 30 for S, 200 and 700 for M, 400 and 1500 for L;
- percentages of mutual attack (Att%) and mutual support (Sup%) relations in the graph for S, M, L: (0%, 0%), (20%, 0%), (40%, 0%), (0%, 20%), (0%, 40%), (20%, 20%).

Each graph configuration test has been run five consecutive times, and the results are the average of the last four runs. The following measurements have been gathered: (i) memory usage peak; (ii) processing time to rank the graph arguments; (iii) iterations required to find a solution. The graphs of every configuration are generated randomly, by only setting up the number of nodes and edges, and the percentages of the latter attacking and supporting each other.

Experimental results are reported in Table 5: short execution time has not allowed to evaluate the memory consumption in Small configurations.

Especially in the M and L configurations, required computational resources seem strongly dependent on the number of edges, if the number of vertexes in a graph remains the same. For example, with 50 vertexes, the configuration with 700 edges has taken about half the iterations but 1 order of magnitude higher memory and time than the configuration

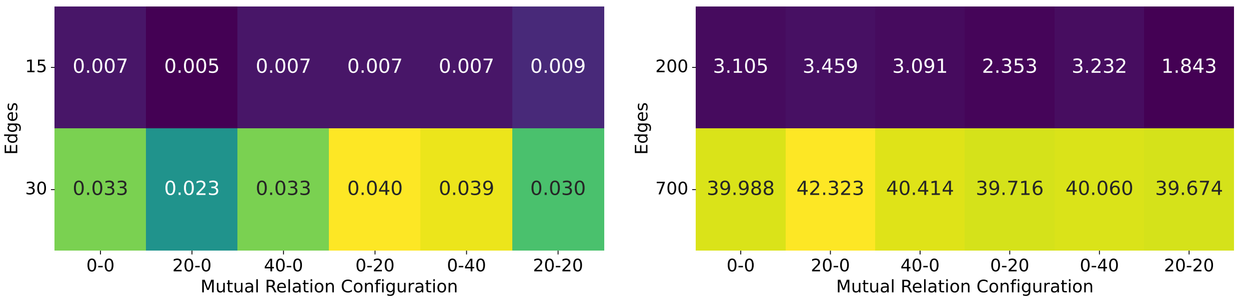
---

[4] Boost Graph Library: `https://www.boost.org/doc/libs/1_77_0/libs/graph/doc/index.html`

Table 5
Argumentative graph performance results

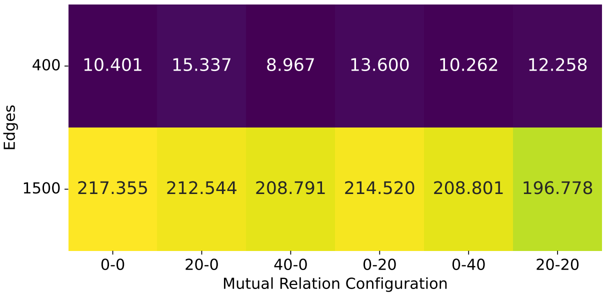
Vertexes	Edges	Mutual attack	Mutual support	Memory (MB)	Time (s)	Iterations
10	15	0%	0%	N.A.	0.007	5
10	15	20%	0%	N.A.	0.005	5.4
10	15	40%	0%	N.A.	0.007	6
10	15	0%	20%	N.A.	0.007	6
10	15	0%	40%	N.A.	0.007	6
10	15	20%	20%	N.A.	0.009	6
10	30	0%	0%	N.A.	0.033	3.8
10	30	20%	0%	N.A.	0.023	5
10	30	40%	0%	N.A.	0.033	5
10	30	0%	20%	N.A.	0.040	5.8
10	30	0%	40%	N.A.	0.039	6
10	30	20%	20%	N.A.	0.030	6
50	200	0%	0%	1.6	3.105	6
50	200	20%	0%	2.4	3.459	6
50	200	40%	0%	1.4	3.091	6.6
50	200	0%	20%	1.4	2.353	7
50	200	0%	40%	1.8	3.232	7
50	200	20%	20%	1.2	1.843	7
50	700	0%	0%	26.8	39.988	3
50	700	20%	0%	27.2	42.323	3
50	700	40%	0%	18.6	40.414	3
50	700	0%	20%	13.4	39.716	3
50	700	0%	40%	13.0	40.060	3
50	700	20%	20%	5.4	39.674	3
100	400	0%	0%	3.8	10.401	6
100	400	20%	0%	5.0	15.337	6
100	400	40%	0%	2.8	8.967	6
100	400	0%	20%	3.6	13.600	6
100	400	0%	40%	3.6	10.262	7
100	400	20%	20%	3.2	12.258	7
100	1500	0%	0%	71.0	217.355	3
100	1500	20%	0%	69.6	212.544	3
100	1500	40%	0%	64.0	208.791	3
100	1500	0%	20%	69.6	214.520	3
100	1500	0%	40%	37.0	208.801	3
100	1500	20%	20%	33.0	196.778	3

with 200 relations. Furthermore, as the pair (Att%, Sup%) increases with the same number of nodes and arcs in a graph, it is observed that the execution time remains almost the same and the use of memory is reduced. Heatmaps in Figure 7 illustrate an ablation of each of the three parameters, in order to understand the impact of individual graph features. The charts show the ranking processing time for graphs with an equal number of nodes but varying numbers of edges and mutual relation percentages (Att%, Sup%) for S, M and L scale configurations. It is evident that the ranking time grows significantly with the number of edges. On the other hand, in graphs with the same number of nodes and edges, the variation of (Att%, Sup%) has almost no effect on the ranking time. In all configurations, of the six hypothesized mutual attack and support couples, the pair (20%, 20%) has the lowest memory usage. This happens because multiple paths quickly satisfy the constraints of Algorithm 2 and it is not necessary to keep them in memory for subsequent iterations. In any case, both medium and large scale configurations appear as unsuitable for real-time applications like SC2 or pervasive computing contexts. These findings limit the amount of knowledge fragments each agent is capable of managing, calling for distributed computation of the argumentative graph.



(a) Small scale configuration

(b) Medium scale configuration



(c) Large scale configuration

Figure 7: Variation of processing time [s] to rank graph w.r.t. number of weighted edges and (Att%, Sup%) mutual relations for S, M and L graph scale configurations

Notable topologies dataset tests. In order to further test the performance of the proposed semantics, a benchmark of well-known graph topologies has been identified. The Erdos-Rényi, Barabási-Albert and Kleinberg datasets, generated by Bistarelli *et al.* [58] and freely accessible online[5], have been adopted. The subsequent configurations, identified by the number of vertices in the graph, have been considered for each of the three datasets:

- Erdos-Rényi dataset: 100, 150, 200, 300;
- Barabási-Albert dataset: 30, 40, 50, 60, 250, 300, 500, 600, 1000, 1200, 1500, 1700, 2500;
- Kleinberg dataset: 9, 16, 25, 36, 49, 64, 81, 144, 400, 625, 900.

Each configuration contains 100 graphs with different numbers of edges. All the graphs have unweighted relations, therefore a preprocessing stage has been carried out, by assigning to each edge a weight chosen randomly in the $[-1, 0[\cup]0, 1]$ interval, in order to evaluate the proposed BWAF ranking semantics. This alteration makes a direct performance comparison infeasible even with other approaches, such as [44], which adopted the same datasets to test other gradual semantics.

It should be pointed out that the primary objective of this work is not to create the fastest solver; therefore, this campaign focuses on the analysis of the average number of iterations required to obtain the argument ranking using the proposed semantics (Algorithm 2) and of the general growth trend of execution time with respect to the number of vertexes in a weighted graph. The ranking semantics has been evaluated on each graph of the three datasets by collecting execution times and the number of iterations. Figures 8, 9 and 10 show ranking processing time on Erdos-Rényi, Barabási-Albert, and Kleinberg dataset respectively. The proposed approach has significantly higher execution times than [44], but a comparison of absolute values is not meaningful, as the hardware testbeds are different (a workstation cluster in [44] w.r.t. a stand-alone machine in this paper) and the work in [44] does not take into account bipolar weighted relations or cyclic graphs. However, it can be noticed that for all datasets the execution times increase in

---

[5] `https://conarg.dmi.unipg.it/benchmarks.html`

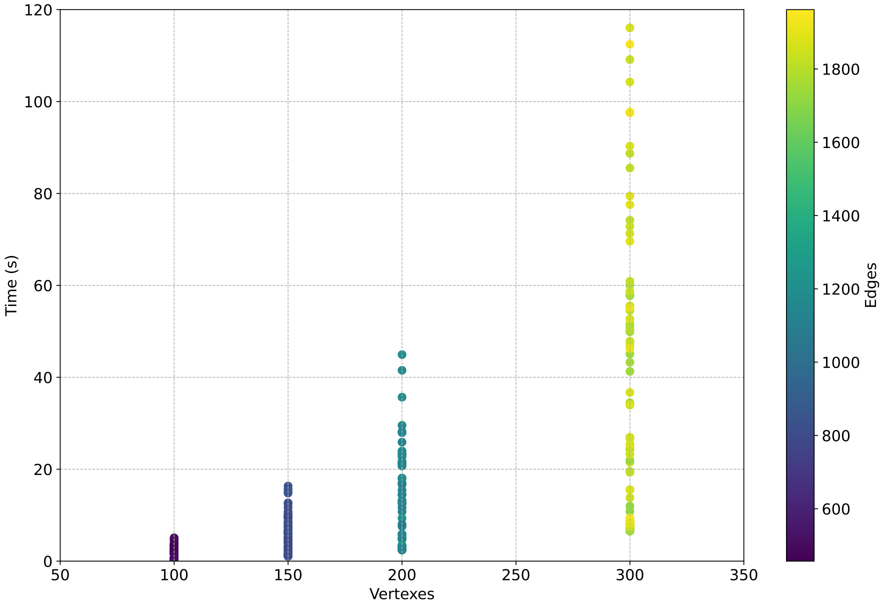


Figure 8: Ranking execution time on Erdos-Rényi dataset

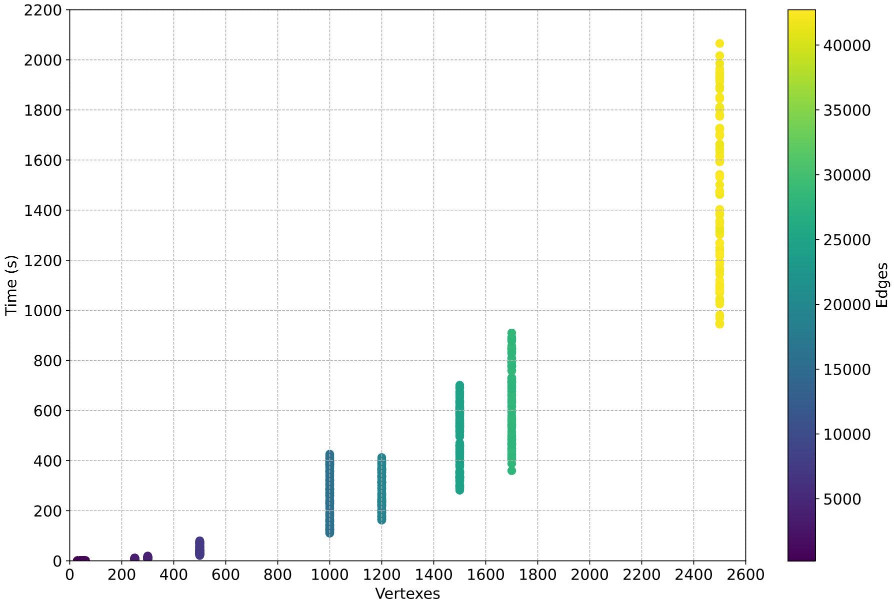


Figure 9: Ranking execution time on Barabási-Albert dataset

Figures 8-10 is substantially comparable to the behavior of [44] as the number of nodes grows. A difference can be appreciated by comparing Figures 9 and 10: the ranking time is reduced by about an order of magnitude for graphs with roughly similar numbers of vertices (*e.g.*, 1000 for the Barabási-Albert dataset and 900 for the Kleinberg dataset) since the number of edges drops dramatically (approximately 16000 edges per 1000 vertices in the Barabási-Albert dataset vs. approximately 4000 edges per 900 vertices in the Kleinberg dataset). This evidence demonstrates once more how the amount of edges in a graph with the same number of vertices influences the ranking execution time. The more

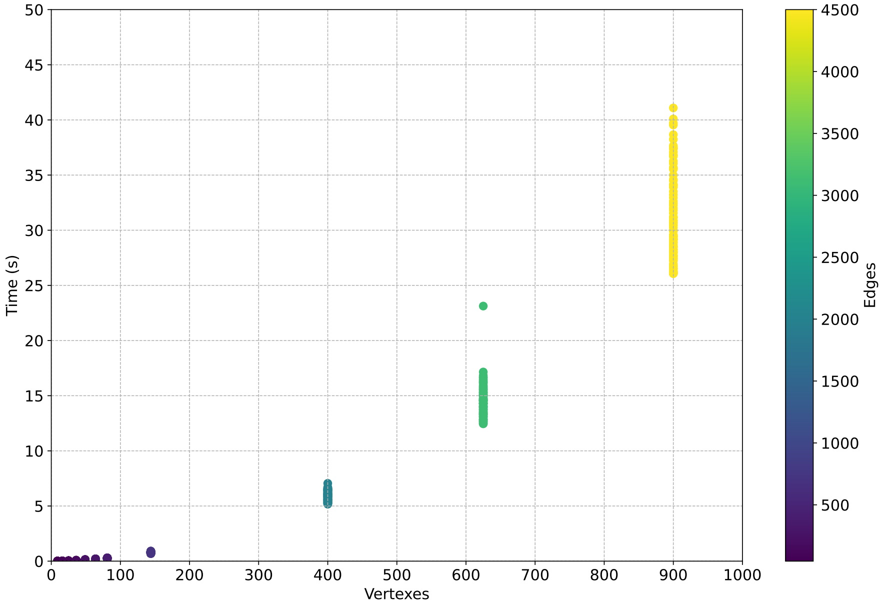


Figure 10: Ranking execution time on Kleinberg dataset

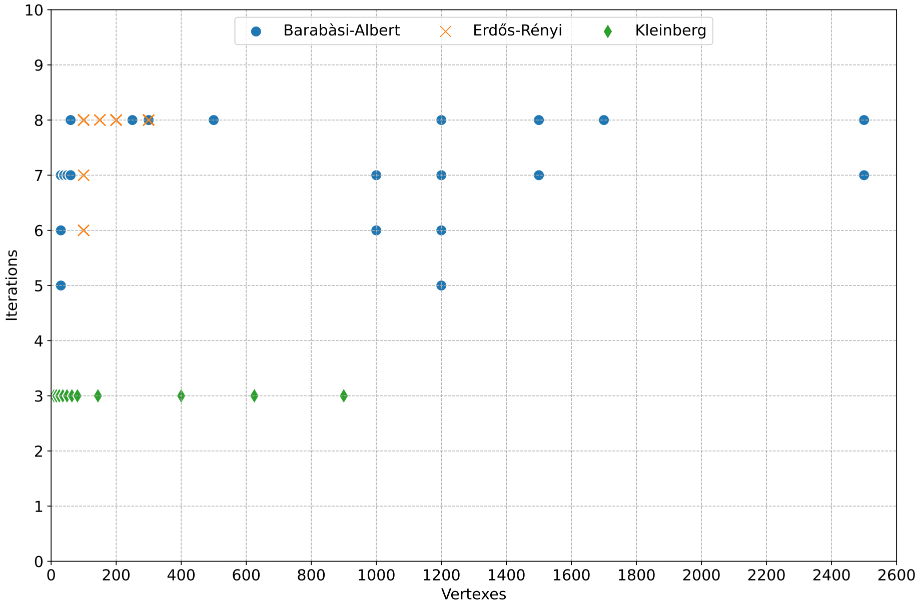


Figure 11: Number of iterations on Erdos-Rényi, Barabási-Albert dataset and Kleinberg datasets

edges there are, the number of paths to be evaluated for each node at each iteration tends to grow, with an increase in time complexity.

Figure 11 shows the number of iterations required to compute the ranking on the graphs of the three datasets. For all the graphs in the Kleinberg dataset, the number of iterations stays steadily at 3. This derives from the fact that the ratio of edges to nodes in the graph of Kleinberg dataset is constant at 5, regardless of the scale configurations of the graph. The iterations on the graphs of the other two datasets do not exhibit the same pattern. They have a rising tendency as the number of nodes increases in the ratio of edges to vertices, which is not constant. However as the graph's size grows, the number of iterations stabilizes at a maximum of 9, as shown for Barabási-Albert graphs. It can therefore be argued that the ranking algorithm scales well, with a similar behavior w.r.t. [44]. Since the maximum number of iterations corresponds to the maximum length of evaluated paths, these results are also useful to appropriately assess the computational and memory requirements to store and process all needed information. At the same time, results can help sizing the decay coefficient ζ of the propagation-based semantics in Section 5: for instance, setting $\zeta = 0.8$ means that for paths of length 9 the coefficient becomes $\zeta^9 = 0.134$ which is the $\frac{\zeta^9}{\zeta^1} \cdot 100 = 16.8\%$ of the strength of paths with length 1, *i.e.*, direct argument relations.

7.1. Discussion

The experimental results provide several significant insights concerning the suggested propagation-based ranking semantics in an explainable BWAF. The lessons learned below highlight the performance challenges counterbalancing the explainability and versatility benefits of the proposed approach, including areas where further improvement is required for practical applications, especially in complex real-world pervasive computing environments.

Edge management. A key observation from the experiments is the significant influence of graph structure – particularly the density of edges– on the performance of the ranking semantics. The computational complexity exhibits an approximately quadratic $O(N^2)$ behavior, leading to significantly longer processing times and larger memory consumption as the number N of edges increases. This is due to the increased number of possible paths that must be evaluated during the ranking process, which grows quadratically with the number of edges in the worst case. This lesson emphasizes the importance of edge management techniques, such as edge pruning or prioritization, to maintain efficiency, especially in dense argumentation graphs.

Parallelization. Although the proposed ranking semantics has satisfactory performance on small-scale graphs, experiments show that scaling becomes more challenging as the graph size increases. The explainable graph construction and the more transparent and fine-grained acceptability ranking of the proposed approach are accompanied by performance challenges in real-time applications, such as the tactical decision-making scenario in the SC2 game engine. The SC2 experimental tests have shown the approach meets the necessary performance requirements for real-time operation, as long as the constructed argumentative graphs remain sufficiently small. Distributing the processing of an argumentative graph by coordinating intermediate results from multiple interconnected sub-graphs could make the approach even more suitable for highly volatile pervasive contexts, but it requires ensuring the correctness of results.

Cyclic graphs. The performance of the suggested ranking semantics is not impacted by the existence of mutual relationships –*i.e.*, loops– and cycles in a bipolar weighted graph. In all tested graph configurations increasing percentages of mutual attacks (Att%) and mutual supports (Sup%) while keeping the same number of nodes and edges leads to almost unchanged calculation times (Figure 7) and to reduced memory usage. This outcome suggests that the proposed semantics efficiently handles cycles by quickly resolving the interactions between mutually related arguments. Cycles seem to stabilize the ranking process, since multiple pathways between arguments satisfy the algorithm's halting conditions faster, requiring fewer iterations. As a result, cycles can speed up computations while avoiding performance degradation. The suggested semantics is therefore particularly suitable for complex argumentation frameworks where cyclic interactions among arguments are common, like mutual communication of smart devices in pervasive computing scenarios.

Robustness. A direct comparison of the proposed ranking semantics with existing related approaches is highly challenging to perform, due to fundamental differences in the underlying argumentation frameworks. Nevertheless, three benchmark datasets containing graphs of varying sizes have been identified and tested. Experimental results have revealed that ranking execution times increase predictably with the number of vertices in the graph, mirroring the trends observed in [44]. Furthermore, ranking semantics iterations exhibits a tendency to stabilize even as the graph configurations scale in size. All this suggests the proposed approach is sufficiently robust and computational efficiency remains stable and predictable, on par with state-of-the-art approaches which are limited to acyclic graphs and disregard fading.

8. Conclusion and future work

This paper has introduced a novel approach for integrating DL-based semantic matchmaking in argumentation frameworks. Non-standard, non-monotonic inferences, endowed with formal motivation of outcomes, allow a structured appraisal of relations between pairs of arguments in a BWAF, while a propagation-based ranking semantics with fading enables argument acceptability score assessment. The proposal is general-purpose, though motivated by collaborative autonomous decision-making and information reliability evaluation in pervasive networks of smart objects. It has been assessed in a case study on a RTS game with comparable complexity and unpredictability challenges. Experiments have shown adequate performance for real-time decision-making agents for reasonably sized argumentation frameworks, as well as robustness and predictability of the increase in computational effort for larger graphs.

Notwithstanding, this work should be considered only as an early result. A significant limitation is that the evaluation algorithm is currently computed in a centralized way, building the whole argumentative graph in advance. Moreover, further performance optimizations are needed: the most important open issue is how to avoid recalculating the whole BWAF ranking every time the graph changes (which may be very frequent in SWoT scenarios). In this perspective, a main future work direction regards how to construct and evaluate the graph according to an incremental approach which evolves in accordance with context changes. Two possible techniques to facilitate that could be: graph modularization, based on some high-level concepts common to all agents in a subset of the network (*e.g.*, in a home automation network, dividing subsystems such as lighting, HVAC, security, etc.); graph simplification, *e.g.*, by pruning weak relations (*i.e.*, low-intensity ones, regardless of polarity) and/or collapsing subgraphs, with careful rules to avoid spurious graph topology changes. The adopted fading mechanism can guide graph simplification by inducing a natural decay of arguments which are too far from a physical or semantic standpoint. In fact, another key investigation area concerns how to perform graph construction and semantics evaluation in a distributed way across a network of cooperating device agents, based on both physical distance (affecting event observation boundaries and possible agent clustering strategies) and semantic distance (enabling to identify portions of the graph related to different knowledge domains). This will increase robustness and scalability of the proposed approach, which will be assessed by means of experiments and case studies in a variety of application scenarios, including home and building automation, smart mobility, digital twins for smart cities and large infrastructures, smart health management, and more. Additionally, systematic optimization techniques will be applied to fine-tune parameters in relation weight formulas and in the ranking semantics algorithm, assessing benefits vs computational costs for diverse applications and types of argumentation graphs. The proposed experimental campaign on notable graph models, evaluating whether topological features of the argumentative graph may have any impact on computational complexity, will be expanded with ablation studies to understand the contribution of each part of the proposal overall efficiency and effectiveness, as well as with the comparison of further existing argumentation semantics. Finally, further ongoing investigation regards embedding the semantic explanations of relations in the argumentation graph itself. This would further increase transparency of the approach, making it immediate to interpret the model and its results.

Acknowledgments

The work was supported by Spoke 9 of the Italian National Center for High-Performance Computing, Big Data and Quantum Computing, funded by the NextGenerationEU program, and by the *UPSIDE* (Urban Playground for masSIve Digital Experiences) project, funded by the Italian Ministry of Enterprises and Made in Italy.

References

[1] F. Scioscia, M. Ruta, Building a Semantic Web of Things: issues and perspectives in information compression, in: Proceedings of the 3rd IEEE International Conference on Semantic Computing, IEEE Computer Society, 2009, pp. 589–594.

[2] Z. Wu, Y. Xu, C. Zhang, Y. Yang, Y. Ji, Towards Semantic Web of Things: from manual to semi-automatic semantic annotation on Web of Things, in: Big Data Computing and Communications: Second International Conference, BigCom 2016, Shenyang, China, July 29-31, 2016. Proceedings 2, Springer, 2016, pp. 295–308.

[3] A. Rhayem, M. B. A. Mhiri, F. Gargouri, Semantic Web Technologies for the Internet of Things: Systematic Literature Review, Internet of Things 11 (2020) 100206. `doi:https://doi.org/10.1016/j.iot.2020.100206`.

[4] F. Z. Amara, M. Hemam, M. Djezzar, M. Maimor, Semantic Web and Internet of Things: Challenges, Applications and Perspectives, Journal of ICT Standardization 10 (2) (2022) 261–291. `doi:10.13052/jicts2245-800X.1029`.

[5] D. Pfisterer, K. Romer, D. Bimschas, O. Kleine, R. Mietz, C. Truong, H. Hasemann, A. Kroller, M. Pagel, M. Hauswirth, et al., SPITFIRE: Toward a Semantic Web of Things, IEEE Communications Magazine 49 (11) (2011) 40–48.

[6] M. Ruta, F. Scioscia, A. Pinto, F. Gramegna, S. Ieva, G. Loseto, E. Di Sciascio, CoAP-based collaborative sensor networks in the Semantic Web of Things, Journal of Ambient Intelligence and Humanized Computing 10 (2019) 2545–2562.

[7] A. J. Jara, A. C. Olivieri, Y. Bocchi, M. Jung, W. Kastner, A. F. Skarmeta, Semantic Web of Things: an analysis of the application semantics for the IoT moving towards the IoT convergence, International Journal of Web and Grid Services 10 (2-3) (2014) 244–272.

[8] G. Loseto, F. Scioscia, M. Ruta, F. Gramegna, S. Ieva, A. Pinto, C. Scioscia, Knowledge-based Decision Support in Healthcare via Near Field Communication, Sensors 20 (17) (2020) 4923.

[9] A. Gyrard, P. Patel, S. K. Datta, M. I. Ali, Semantic Web Meets Internet of Things and Web of Things, in: Proceedings of the 26th International Conference on World Wide Web Companion, WWW '17 Companion, International World Wide Web Conferences Steering Committee, 2017, p. 917–920. `doi:10.1145/3041021.3051100`.

[10] M. Ruta, F. Scioscia, G. Loseto, A. Pinto, E. Di Sciascio, Machine Learning in the Internet of Things: a Semantic-enhanced Approach, Semantic Web Journal 10 (1) (2019) 183–204.

[11] M. Ruta, F. Scioscia, F. Gramegna, S. Ieva, E. Di Sciascio, R. P. De Vera, A knowledge fusion approach for context awareness in vehicular networks, IEEE Internet of Things Journal 5 (4) (2018) 2407–2419.

[12] M. Ruta, F. Scioscia, G. Loseto, F. Gramegna, S. Ieva, A. Pinto, E. Di Sciascio, Social Internet of Things for Domotics: a Knowledge-based Approach over LDP-CoAP, Semantic Web Journal 9 (6) (2018) 781–802.

[13] M. Lippi, M. Mamei, S. Mariani, F. Zambonelli, An argumentation-based perspective over the social IoT, IEEE Internet of Things Journal 5 (4) (2018) 2537–2547.

[14] M. Ruta, F. Scioscia, G. Loseto, A. Pinto, C. Fasciano, G. Capurso, E. Di Sciascio, Internet of Conscious Things: Ontology-Based Social Capabilities for Smart Objects, Future Internet 16 (9) (2024) 327.

[15] P. Besnard, A. Hunter, Elements of Argumentation, MIT Press, 2008.

[16] A. J. García, G. R. Simari, Defeasible logic programming: An argumentative approach, Theory and practice of logic programming 4 (1-2) (2004) 95–138.

[17] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artificial Intelligence 195 (2013) 361–397.

[18] 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.

[19] P. Besnard, C. Cayrol, M.-C. Lagasquie-Schiex, Logical theories and abstract argumentation: A survey of existing works, Argument & Computation 11 (1-2) (2020) 41–102.

[20] P. Besnard, A. Garcia, A. Hunter, S. Modgil, H. Prakken, G. Simari, F. Toni, Introduction to structured argumentation, Argument & Computation 5 (1) (2014) 1–4.

[21] R. Craven, F. Toni, Argument graphs and assumption-based argumentation, Artificial Intelligence 233 (2016) 1–59.

[22] P. Besnard, A. Hunter, Constructing argument graphs with deductive arguments: a tutorial, Argument & Computation 5 (1) (2014) 5–30.

[23] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion 58 (2020) 82–115.

[24] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), IEEE access 6 (2018) 52138–52160.

[25] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, Whitepaper, European Commission (Apr. 2019).

[26] N. A. Smuha, The EU approach to ethics guidelines for trustworthy artificial intelligence, Computer Law Review International 20 (4) (2019) 97–106.

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

[28] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, P. Patel-Schneider, The Description Logic Handbook, Cambridge University Press, 2002.

[29] B. Parsia, S. Rudolph, M. Krötzsch, P. Patel-Schneider, P. Hitzler, OWL 2 Web Ontology Language Primer (Second Edition), W3C Recommendation, W3C, http://www.w3.org/TR/owl2-primer (Dec. 2012).

[30] C. Cayrol, M.-C. Lagasquie-Schiex, On the acceptability of arguments in bipolar argumentation frameworks, in: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU, Vol. 3571 of Lecture Notes in Computer Science, Springer, 2005, pp. 378–389.

[31] M. Ruta, E. Di Sciascio, F. Scioscia, Concept Abduction and Contraction in Semantic-based P2P Environments, Web Intelligence and Agent Systems 9 (3) (2011) 179–207.

[32] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, The Knowledge Engineering Review 36 (2021) e5.

[33] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, Artificial Intelligence 171 (10) (2007) 675–700.

[34] L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, in: International Conference on Scalable Uncertainty Management, Springer, 2013, pp. 134–147.

[35] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation., in: AAAI, 2016, pp. 914–920.

[36] M. Ruta, F. Scioscia, I. Bilenchi, F. Gramegna, G. Loseto, S. Ieva, A. Pinto, A multiplatform reasoning engine for the Semantic Web of Everything, Journal of Web Semantics 73 (2022) 100709.

[37] C. E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: Partial meet contraction and revision functions, The Journal of Symbolic Logic 50 (2) (1985) 510–530.

[38] L. Atzori, A. Iera, G. Morabito, M. Nitti, The social internet of things (siot)–when social networks meet the internet of things: Concept, architecture and network characterization, Computer networks 56 (16) (2012) 3594–3608.

[39] E. Lovellette, H. Hexmoor, K. Rodriguez, Automated argumentation for collaboration among cyber-physical system actors at the edge of the Internet of Things, Internet of Things 5 (2019) 84–96.

[40] N. Gulati, P. D. Kaur, An argumentation enabled decision making approach for Fall Activity Recognition in Social IoT based Ambient Assisted Living systems, Future Generation Computer Systems 122 (2021) 82–97.

[41] N. Potyka, Bipolar Abstract Argumentation with Dual Attacks and Supports, in: Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, 2020, pp. 677–686.

[42] J. Heyninck, B. Raddaoui, C. Straßer, Ranking-based argumentation semantics applied to logical argumentation., in: IJCAI, 2023, pp. 3268–3276.

[43] L. Amgoud, J. Ben-Naim, Evaluation of arguments in weighted bipolar graphs, International Journal of Approximate Reasoning 99 (2018) 39–55.

[44] L. Amgoud, D. Doder, S. Vesic, Evaluation of argument strength in attack graphs: Foundations and semantics, Artificial Intelligence 302 (2022) 103607.

[45] L. Amgoud, J. Ben-Naim, D. Doder, S. Vesic, Ranking arguments with compensation-based semantics, in: Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning, 2016, pp. 12–21.

[46] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, Argumentation ranking semantics based on propagation., COMMA 16 (2016) 139–150.

[47] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A parametrized ranking-based semantics compatible with persuasion principles, Argument & Computation 12 (1) (2021) 49–85.

[48] A. Pazienza, S. Ferilli, F. Esposito, Constructing and evaluating bipolar weighted argumentation frameworks for online debating systems, in: 1st Workshop on Advances In Argumentation In Artificial Intelligence, XVI International Conference of the Italian Association for Artificial Intelligence (AI^3@AI*IA), 2017, pp. 111–125.

[49] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, M. Wooldridge, Weighted argument systems: Basic definitions, algorithms, and complexity results, Artificial Intelligence 175 (2) (2011) 457 – 486.

[50] S. Kaci, C. Labreuche, Valued preference-based instantiation of argumentation frameworks with varied strength defeats, International Journal of Approximate Reasoning 55 (9) (2014) 2004–2027.

[51] C. Cayrol, M.-C. Lagasquie-Schiex, From preferences over arguments to preferences over attacks in abstract argumentation: A comparative study, in: 2013 IEEE 25th International Conference on Tools with Artificial Intelligence, IEEE, 2013, pp. 588–595.

[52] B. Fazzinga, S. Flesca, F. Furfaro, Probabilistic bipolar abstract argumentation frameworks: complexity results, in: International Joint Conference on Artificial Intelligence, 2018, pp. 1803–1809.

[53] A. Cohen, S. Gottifredi, A. J. García, G. R. Simari, A survey of different approaches to support in argumentation systems, The Knowledge Engineering Review 29 (5) (2014) 513–550.

[54] S. Sarkar, K. Mukherjee, A. Ray, Distributed decision propagation in mobile-agent proximity networks, International Journal of Control 86 (6) (2013) 1118–1130.

[55] N. Oren, T. J. Norman, Semantics for evidence-based argumentation, in: Computational Models of Argument: Proceedings of COMMA 2008, IOS Press, 2008, pp. 276–284.

[56] M. Horridge, P. Patel-Schneider, OWL 2 Web Ontology Language Manchester Syntax (Second Edition), Note, W3C, `http://www.w3.org/TR/owl2-manchester-syntax/` (Dec. 2012).

[57] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, et al., Starcraft II: A new challenge for reinforcement learning, arXiv preprint arXiv:1708.04782 (2017).

[58] S. Bistarelli, F. Rossi, F. Santini, A first comparison of abstract argumentation reasoning-tools, in: ECAI 2014, IOS Press, 2014, pp. 969–970.